SIXTH EDITION

Molecular Biotechnology

Principles and Applications of Recombinant DNA

Bernard R. Glick and Cheryl L. Patten

SIXTH EDITION

Molecular Biotechnology

Principles and Applications of Recombinant DNA

SIXTH EDITION

Molecular Biotechnology

Principles and Applications of Recombinant DNA

Bernard R. Glick
Department of Biology, University of Waterloo
Waterloo, Ontario, Canada

Cheryl L. Patten
Biology Department, University of New Brunswick
Fredericton, New Brunswick, Canada

ASM PRESS
WASHINGTON, DC

WILEY

Copyright © 2022 American Society for Microbiology. All rights reserved.
Copublication by the American Society for Microbiology and
John Wiley & Sons, Inc.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted by law. Advice on how to reuse material from this title is available at http://wiley.com/go/permissions.

The right of Bernard R. Glick and Cheryl L. Patten to be identified as the authors of the editorial material in this work has been asserted in accordance with law.

Limit of Liability/Disclaimer of Warranty
While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy of completeness of the contents of this book and specifically disclaim any implied warranties or merchantability of fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The publisher is not providing legal, medical, or other professional services. Any reference herein to any specific commercial products, procedures, or services by trade name, trademark, manufacturer, or otherwise does not constitute or imply endorsement, recommendation, or favored status by the American Society for Microbiology (ASM). The views and opinions of the author(s) expressed in this publication do not necessarily state or reflect those of ASM, and they shall not be used to advertise or endorse any product.

Editorial Correspondence:
ASM Press, 1752 N Street, NW, Washington, DC 20036-2904, USA

Registered Offices:
John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA

For details of our global editorial offices, customer services, and more information about Wiley products, visit us at www.wiley.com.

Wiley also publishes its books in a variety of electronic formats and by print-on-demand. Some content that appears in standard print versions of this book may not be available in other formats.

Library of Congress Cataloging-in-Publication Data

Names: Glick, Bernard R., author. | Patten, Cheryl L., author. | John Wiley & Sons, Inc., publisher.
Title: Molecular biotechnology : principles and applications of recombinant DNA / Bernard R. Glick and Cheryl L. Patten.
Description: Sixth edition. | [Hoboken, NJ] : Wiley-ASM Press, [2022] | Includes bibliographical references and index.
Identifiers: LCCN 2021033752 (print) | LCCN 2021033753 (ebook) | ISBN 9781683673644 (cloth) | ISBN 9781683673651 (adobe pdf) | ISBN 9781683673668 (epub)
Subjects: LCSH: Biotechnology. | Genetic engineering. | Molecular biology.
Classification: LCC TP248.2 .G58 2022 (print) | LCC TP248.2 (ebook) | DDC 660.6/5–dc23
LC record available at https://lccn.loc.gov/2021033752
LC ebook record available at https://lccn.loc.gov/2021033753

Cover illustration: Terese Winslow, CMI
Cover design by: Susan Brown Schmidler
Illustrations: Patrick Lane, ScEYEnce Studios

Set in 10/12pt PalatinoLTStd by Straive, Chennai, India

10 9 8 7 6 5 4 3 2 1

The two of us thank our life partners, Marcia Glick and Patrick Patten, for the enormous support and encouragement that they have provided throughout this endeavor.

B. R. Glick
C. L. Patten

About the Companion Website

This book is accompanied by a companion website for instructors.

www.wiley.com/go/glick/molbiotech6

This website includes:

- Powerpoints of Figures and Tables from the book

Contents

Preface to the Sixth Edition xvii

1 **The Development of Molecular Biotechnology 1**

Emergence of Molecular Biotechnology 1

Recombinant DNA Technology 3

Commercialization of Molecular Biotechnology 6

Concerns and Consequences 8

SUMMARY 10

REFERENCES 10

REVIEW QUESTIONS 11

2 **Fundamental Technologies 13**

Molecular Cloning 13

Preparation of DNA for Cloning 13

Insertion of Target DNA into a Plasmid Vector 18

Transformation and Selection of Cloned DNA in a Bacterial Host 21

Cloning Eukaryotic Genes 26

Recombinational Cloning 30

Genomic Libraries 32

Genome Engineering Using CRISPR Technology 36

Polymerase Chain Reaction 39

Amplification of DNA by PCR 39

Cloning PCR Products 42

Quantitative PCR 44

Chemical Synthesis of Genes 46

Assembling Oligonucleotides into Genes 46

Assembling PCR Products into Genes 46

DNA Sequencing Technologies 48
Dideoxynucleotide Sequencing 50
Sequencing Using Reversible Chain Terminators 53
Single-Molecule Real-Time Sequencing 55
Nanopore Sequencing 56
Sequencing Whole Genomes 56
Preparation of Genomic DNA Sequencing Libraries 57
High-Throughput Next-Generation Sequencing 59
Genome Sequence Assembly 60
Sequencing Metagenomes 61
Genomics 62
Transcriptomics 65
Proteomics 70
Metabolomics 84
SUMMARY 86
REFERENCES 87
REVIEW QUESTIONS 89

3 Production of Recombinant Proteins 91
Protein Production in Prokaryotic Hosts 91
Regulating Transcription 92
Increasing Translation Efficiency 96
Increasing Protein Stability 100
Increasing Protein Secretion 105
Facilitating Protein Purification 111
Integrating DNA into the Host Chromosome 114
Heterologous Protein Production in Eukaryotic Cells 119
Posttranslational Modification of Eukaryotic Proteins 120
General Features of Eukaryotic Expression Systems 123
Yeast Expression Systems 123
Baculovirus–Insect Cell Expression Systems 137
Mammalian Cell Expression Systems 145
Protein Engineering 156
Directed Mutagenesis 156
Random Mutagenesis 161
Examples of Protein Engineering 164
SUMMARY 173
REFERENCES 174
REVIEW QUESTIONS 176

4 Molecular Diagnostics 179
Immunological Approaches To Detect Protein Biomarkers 180
Antibodies 180
Agglutination 185
Enzyme-Linked Immunosorbent Assays 186
Protein Arrays To Detect Polygenic Diseases 194
Immunoassays for Protein Conformation-Specific Disorders 197

DNA-Based Diagnostic Approaches 199
Hybridization Probes 199
PCR-Based Detection Methods 207
CRISPR-Cas-Based Diagnostic Assays 218
DNA Microarrays 219
Whole-Genome Sequencing To Assess Genetic Disease Risk 225
Detecting RNA Signatures of Disease 226
Detection of Disease-Associated Changes in Gene Expression 227
Detection of RNA Signatures of Antibiotic Resistance in Bacteria 228
Detection of miRNA Signatures of Disease 230
Biofluorescent and Bioluminescent Systems 233
Fluorescent Proteins 233
Luciferase 234
Microbial Biosensors 235
SUMMARY 238
REFERENCES 239
REVIEW QUESTIONS 241

5 Protein Therapeutics 243
Pharmaceuticals 244
Human Interferons 244
Human Growth Hormone 248
Tumor Necrosis Factor Alpha 251
Extending Protein Half-Life 252
Enzymes 253
DNase I 253
Alginate Lyase 254
Phenylalanine Ammonia Lyase 258
α_1-Antitrypsin 259
Glycosidases 261
Masking Nonhuman Epitopes 263
Toxin-Intein Fusions 264
Targeting Mitochondria 265
Bacteria and Therapeutics 267
Interleukin-10 270
Leptin 272
An HIV Inhibitor 274
Insulin 276
Parkinson's Disease 279
Cancer and Bacteria 279
Recombinant Antibodies 280
Hybrid Human–Mouse Monoclonal Antibodies 284
Human Monoclonal Antibodies 287
Antibody Fragments 289
Combinatorial Libraries of Antibody Fragments 294
A Combinatorial Library of Full-Length Antibodies 297
Shuffling CDR Sequences 298
Dual-Variable-Domain Antibodies 298

Bispecific Antibodies against Hemophilia 300
Anti-HIV Antibodies 300
Anticancer Antibodies 302
Antibodies against Various Diseases 309
Antiobesity Antibodies 313
Enhanced Antibody Half-Life 315

Affibody Molecules 315

SUMMARY 318

REFERENCES 318

REVIEW QUESTIONS 322

6 Nucleic Acids as Therapeutic Agents 325

Targeting Specific mRNA and DNA Sequences 327
Antisense RNA 327
Aptamers 331
Ribozymes and DNAzymes 338
Interfering RNA 341
Zinc Finger Nucleases 348
CRISPR-Cas System 349
Nanozymes 351
Nanoparticles 352
Engineering Bacteriophages 352

Viral Delivery Systems 357

Nonviral Delivery Systems 365
Direct Injection 365
Lipids 367
Bacteria 369
Dendrimers 372
Antibodies 373
Aptamers 373
Transposons 374

Gene Therapy 376
Mitochondrial Diseases 378
Prodrug Activation Therapy 378
Promoterless Gene Targeting 379

SUMMARY 382

REFERENCES 382

REVIEW QUESTIONS 386

7 Vaccines 387

Vaccination 387

Current and Future Vaccines 389

Subunit and Peptide Vaccines 392
Herpes Simplex Virus 393
Bovine Herpes Virus-1 394
Cholera 396
Influenza 396
SARS 397
COVID-19 399

Staphylococcus aureus 401
Human Papillomavirus 402
Foot-and-Mouth Virus 404
Streptococcus 405
Peptides 407
Malaria 408
Delivery 411

Genetic Immunization: DNA Vaccines 414
Delivery 414
Cancer 422
Zika Virus 422
Dental Caries 423

Engineered Attenuated Vaccines 424
Herpes Simplex Virus 425
Cholera 426
Salmonella Species 428
Leishmania Species 430

Vector Vaccines 430
Vaccines Directed against Viruses 430
Vaccines Directed against Bacteria 441
Bacteria as Antigen Delivery Systems 444

Monoclonal Antibody Passive Immunity 449
Influenza Virus 450

SUMMARY 452

REFERENCES 452

REVIEW QUESTIONS 456

8 Industrial and Environmental Uses of Recombinant Microorganisms 459

Restriction Endonucleases 459

Small Biological Molecules 461
L-Ascorbic Acid 463
Indigo 467
Amino Acids 468
Lycopene 473
Antibiotics 474
Biopolymers 487
Solvent Tolerance 493
Systems Metabolic Engineering To Optimize Product Yield 494

Microbial Degradation of Xenobiotics 496
Genetic Engineering of Biodegradative Pathways 497
Plastics 507

Utilization of Starch and Sugars 508
Commercial Production of Fructose and Alcohol 508
Increasing Alcohol Production 510
Improving Fructose Production 517

Utilization of Cellulose and Hemicellulose 518
Lignocellulosics 519
Cellulase Genes 522

Direct Conversion of Biomass to Ethanol 530
Alcohol Production by *Zymomonas mobilis* 531

Lipids from Cyanobacteria 534

Hydrogen Production 535

SUMMARY 538

REFERENCES 539

REVIEW QUESTIONS 542

9 Large-Scale Production of Proteins and Nucleic Acids from Recombinant Microorganisms 545

Principles of Microbial Growth 547
Batch Fermentation 548
Fed-Batch Fermentation 549
Continuous Fermentation 550

Maximizing the Efficiency of the Fermentation Process 551
High-Density Cell Cultures 552
Increasing Plasmid Stability 555
Quiescent *E. coli* Cells 555
Protein Secretion 558
Reducing Acetate 558
Improving Antibody Production in *E. coli* 561

Bioreactors 561

Typical Large-Scale Fermentation Systems 565
Two-Stage Fermentation in Tandem Airlift Reactors 566
Two-Stage Fermentation in a Single Stirred-Tank Reactor 568
Batch versus Fed-Batch Fermentation 569

Harvesting Microbial Cells 574

Disrupting Microbial Cells 576

Downstream Processing 578
Inclusion Bodies 579
Utilizing an Immobilized Enzyme 582
Magnetic Separation of Proteins 582

Large-Scale Production of DNA and RNA 583
Plasmid DNA 583
mRNA 586

SUMMARY 587

REFERENCES 587

REVIEW QUESTIONS 590

10 Genetic Engineering of Plants: Methodology 591

Plant Transformation with the Ti Plasmid of *A. tumefaciens* 595

Ti Plasmid-Derived Vector Systems 597
Increasing Transformation Efficiency 601

Microprojectile Bombardment 603

Chloroplast Engineering 604
Very-High-Level Protein Expression 607

Use of Reporter Genes in Transformed Plant Cells 610

Manipulation of Gene Expression in Plants 611
Transient Gene Expression 611
Plant Promoters 616
Manipulation of Genes in Plants 617
Facilitating Protein Purification 621
Protein Glycosylation 623
Gene Stacking 624
CRISPR-Based Directed Evolution 625
Polycistronic Gene Expression 626

Production of Marker-Free Transgenic Plants 626
Removing Marker Genes from Nuclear DNA 627
Removing Marker Genes from Chloroplast DNA 632

SUMMARY 633

REFERENCES 634

REVIEW QUESTIONS 636

11 Transgenic Plants 637

Insect Resistance 637
Bacillus thuringiensis Insecticidal Toxin 637
Increasing Expression of the *B. thuringiensis* Protoxin 642
Other Strategies for Protecting Plants against Insects 645
Preventing the Development of *B. thuringiensis*-Resistant Insects 652
Targeting Aphids 657

Virus Resistance 658
Viral Coat Protein-Mediated Protection 658
Protection by Expression of Other Genes 663

Herbicide Resistance 668
Glyphosate 669
Dicamba 672
Other Herbicides 673

Fungus and Bacterium Resistance 674
Transgenic Plants 675
RNAi and CRISPR/Cas 681

Salt and Drought Stress 682
Increasing Trehalose Production 683
Sequestering Sodium Ions 684
Delaying Drought-Induced Senescence 685

Phytoremediation 686

Fruits and Flowers 688
Flavr Savr Tomato 688
Lowering Ethylene Levels 688
CRISPR Mutants 690

Modification of Plant Nutritional Content 690
Amino Acids 690
Lipids 692
Vitamins 695
Iron 698
Gluten 700

Modification of Food Plant Taste and Appearance 701
Preventing Discoloration 701
Starch 703

Plants as Bioreactors 706
Antibodies 706
Pharmaceuticals and Vaccines 709
Poly(3-Hydroxybutyric Acid) 710

Edible Vaccines 711
Edible Cholera Vaccines 712
Edible *E. coli* Vaccines 714

Plant Yield 716
Increasing Grain Yield 716
Increasing Harvest Index 716
Decreasing Lignin Content 717
Decreasing Pectin Content 720
Increasing Oxygen Content 722

SUMMARY 723

REFERENCES 724

REVIEW QUESTIONS 729

12 Transgenic Animals 731

Transgenic Animal Methodologies 733
DNA Microinjection Method 733
Retroviral Vector Method 736
Engineered Embryonic Stem Cell Method 737
Somatic Cell Nuclear Transfer for Transgenic Livestock 743
Genome Editing with the CRISPR-Cas System 744
Conditional Gene Modification with the Cre–*loxP* Recombination System 747
Control of Transgene Expression with the Tetracycline-Inducible System 749
Gene Knockdown by RNA Interference 754

Transgenic Animal Models of Human Diseases 756
Mouse Models of Alzheimer's Disease 756
Mouse Model of Duchenne Muscular Dystrophy 759
Rabbit Models of Cardiovascular Disease 761
Zebrafish Melanoma Model 763
Nonhuman Primate Models of Neurodevelopmental Disorders 766

Animal Bioreactors for Production of Recombinant Therapeutic Proteins 767
- Production of Recombinant Antithrombin in Goat Milk 768
- Production of a Human Protease Inhibitor in Rabbits 770
- Production of Therapeutic Proteins in Chicken Eggs 771
- Production of Donor Organs in Pigs 773

Enhancing Production Traits of Food Animals 774
- Disease-Resistant Livestock 774
- Improving Milk Quality 781
- Increasing Muscle Mass in Cattle 782
- Enhancing Growth of Salmon 786

Gene Drives To Eradicate Vector-Transmitted Diseases 787
- Malaria Vector Population Suppression 789
- Dengue Fever Virus-Resistant Mosquitoes 791
- Reversal Drives 792

SUMMARY 795

REFERENCES 796

REVIEW QUESTIONS 797

13 **Molecular Biotechnology and Society 799**

Development of Guidelines for Recombinant DNA Research 800

Deliberate Release of Genetically Modified Microorganisms 802
- Environmental Concerns 802
- Regulations 803

Regulation of Genetically Modified Foods 804
- Food Ingredients Produced by Genetically Engineered Microorganisms 804
- Genetically Modified Crops 807
- Genetically Engineered Livestock 810

Societal Concerns about Genetically Modified Foods 812
- Alteration of Nutritional Content of Food 812
- Potential for Introducing Toxins or Allergens into Food 816
- Potential for Transferring Transgenes from Food to Humans or Intestinal Microorganisms 819
- Controversy about the Labeling of Genetically Modified Foods 820
- Impact of Genetically Engineered Crops on Biodiversity 822
- Who Benefits from the Production of Genetically Modified Foods? 824
- Environmental Benefits of Genetically Modified Crops 825
- How Do Views about Genetically Engineered Organisms Impact Trade? 827

Regulation and Safety of Medical Products of Biotechnology 827
- New Biological Drugs 828
- Genetic and Genomic Testing 832
- Economic Issues 835

Patenting Biotechnology 837
Patenting 838
Patenting in Different Countries 839
Patenting Nucleic Acid Sequences 841
Patenting Living Organisms 842
Patenting and Fundamental Research 844
SUMMARY 845
REFERENCES 846
REVIEW QUESTIONS 848

Amino Acids of Proteins and Their Designations 851

Index 853

Preface to the Sixth Edition

When the first edition of *Molecular Biotechnology: Principles and Applications of Recombinant DNA* was published in 1994, nearly all of the transgenic organisms that were produced included only a single foreign gene or cDNA. Now, nearly 30 years later it is common for scientific researchers to genetically engineer organisms by modifying both the activity and regulation of one or more existing genes or by introducing entire new pathways. Genetic engineering is no longer limited to only a small number of common bacterial strains using recombinant DNA technology. Scientists now routinely modify many different microorganisms, animals, and plants using the techniques of PCR, chemical DNA synthesis, directed mutagenesis, and genome editing. This progress has been greatly facilitated by advances in DNA and RNA sequencing, monoclonal antibody production, genomics, proteomics, and metabolomics. Scientists worldwide can understand and purposefully manipulate the biological world as never before.

In 1994, only a very few products produced by recombinant DNA technology had been commercialized. Today, as a consequence of molecular biotechnology hundreds of new therapeutic agents, diagnostics tests, and vaccines are available in the marketplace with many more in the pipeline. At the time of this writing, RNA vaccines have been produced in record time and approved for the first time by regulatory authorities worldwide to control the devastating COVID-19 global pandemic, which has also spurred advances in molecular diagnostics. Scientists have also genetically modified hundreds of different plants to improve crop yields and traits, with dozens of these transgenic plants already commercialized, and many more in the works. DNA technologies have become a cornerstone of modern forensics, paternity testing, and ancestry determination. The list goes on and on. Molecular biotechnology has clearly lived up to its promise and all of the original hype that has existed since the late 1970s. Worldwide there are several thousand biotechnology companies and research institutes, operating in virtually every corner of the globe, employing hundreds of thousands of highly skilled scientists. In addition to all of the scientific studies with commercial potential that are being conducted, there is also an enormous amount of exciting and innovative

fundamental biological research, using the many techniques of molecular biotechnology, being done at universities and government labs around the world. Never before in recorded history has the world witnessed such a vast amount of discovery and fundamental change in the biological sciences, and there is every indication that there is still much more to come. This sixth edition of *Molecular Biotechnology* builds upon the fundamentals that were established in the previous five editions, and endeavors to provide readers with a window on many of the major developments in this ever growing and ever important field. Given the enormity of the field of molecular biotechnology, we have had to be highly selective in choosing the material that we have included in this edition. Moreover, the window that we are looking through is continually moving. With the many changes in this field that have occurred since the first edition of *Molecular Biotechnology*, we both expect and look forward to a considerable amount of additional change in the coming years, including the implementation and commercialization of many of the discoveries that are discussed here.

We have throughout endeavored to make the text as reader-friendly as possible by minimizing the use of technical jargon and unnecessary abbreviations. When an important term appears for the first time in the text, it is generally followed in parentheses with a synonym or brief explanation. The figures and tables are not just restatements of the data from the original literature; rather, we have made an effort to conceptualize the experimenters' thinking and rationale. We have endeavored to be as up-to-date as possible, expanding on some previous discussions and providing large numbers of practical examples.

Each chapter opens with an outline of topics and concludes with a summary and list of review questions to sharpen students' critical thinking skills. All of the key ideas in the book are illustrated by the more than 585 full-color figures and elaborated in nearly 100 tables. After introducing molecular biotechnology as a scientific and economic venture in Chapter 1, the next two chapters explain the detailed methodologies of molecular biotechnology. These chapters provide a solid scientific base for the remainder of the book. Chapters 4 to 8 present examples of applications for microbial molecular biotechnology covering such topics as diagnostic techniques, both protein and nucleic acid therapeutic agents, vaccines, bioremediation of pollutants, the production of metabolites, and biomass utilization by industry. Chapter 9 describes some of the key components of large-scale fermentation processes using recombinant microorganisms. Chapters 10 to 12 describe the molecular manipulation of plants and animals addressing both fundamental approaches and a wide range of applications, with a particular emphasis on agricultural improvements. The book concludes in Chapter 13 with a discussion of the interaction of molecular biotechnology with society including some discussion of controversies that have occurred as a consequence of this technology, coverage of the regulation of molecular biotechnology, and patents.

Throughout the text we have relied extensively upon the recent published work of many researchers. In all cases, although not cited directly in the body of a chapter, the original published articles are cited in the references section of the appropriate chapter. In some cases, we have taken "pedagogic license" and either extracted or reformulated data from the original publications. Clearly, we are responsible for any distortions or

misrepresentations from these simplifications, although we hope that none have occurred. The references section also contains other sources that we used in a general way, which might, if consulted, bring the readers closer to a particular subject.

Bernard R. Glick
Cheryl L. Patten

1 The Development of Molecular Biotechnology

Emergence of Molecular Biotechnology
Recombinant DNA Technology
Commercialization of Molecular Biotechnology
Concerns and Consequences
SUMMARY
REFERENCES
REVIEW QUESTIONS

Emergence of Molecular Biotechnology

Long before we knew that microorganisms existed or that genes were the units of inheritance, humans looked to the natural world to develop methods to increase food production, preserve food, and heal the sick. Our ancestors discovered that grains could be preserved through fermentation into beer, that storing horse saddles in a warm, damp corner of the stable resulted in the growth of a saddle mold that could heal infected saddle sores, that intentional exposure to a "contagion" could somehow provide protection from an infectious disease on subsequent exposures, and that plants and animals with enhanced production traits could be developed through crossbreeding. Following the discovery of the microscopic world in the 17th century, microorganisms have been employed in the development of numerous useful processes and products. Many of these are found in our households and backyards. Lactic acid bacteria are used to prepare yogurts and probiotics, insecticide-producing bacteria are sprayed on many of the plants from which the vegetables in our refrigerator are harvested, nitrogen-fixing bacteria are added in the soil used for cultivation of legumes, the enzymatic stain removers in laundry detergent come from a microorganism, and antibiotics that are derived from common soil microbes are used to treat infectious diseases. These are just a few examples of traditional biotechnologies that have improved our lives. Up to the early 1970s, however, biotechnology was not a well-recognized scientific discipline, and research in this area was centered in departments of chemical engineering and occasionally in specialized microbiology programs.

Molecular Biotechnology: Principles and Applications of Recombinant DNA, Sixth Edition. Bernard R. Glick and Cheryl L. Patten.

© 2022 American Society for Microbiology.

In a broad sense, biotechnology is concerned with the manipulation of organisms to develop and manufacture useful products. The term "biotechnology" was first used in 1917 by a Hungarian engineer, Karl Ereky, to describe an integrated process for the large-scale production of pigs by using sugar beets as the source of food. According to Ereky, biotechnology was "all lines of work by which products are produced from raw materials with the aid of living things." This fairly precise definition was more or less ignored. For a number of years, biotechnology was used to describe two very different engineering disciplines. On one hand, it referred to industrial fermentation. On the other, it was used for the study of efficiency in the workplace—what is now called ergonomics. This ambiguity ended in 1961 when the Swedish microbiologist Carl Göran Hedén recommended that the title of a scientific journal dedicated to publishing research in the fields of applied microbiology and industrial fermentation be changed from the *Journal of Microbiological and Biochemical Engineering and Technology* to *Biotechnology and Bioengineering*. From that time on, biotechnology has been defined as the application of scientific and engineering principles to the processing of material by biological agents to provide goods and services. It is grounded on expertise in microbiology, genetics, biochemistry, immunology, cell biology, and chemical engineering.

Large-scale production of commodities from natural organisms is often considerably less than optimal. Initial efforts to enhance yields of microbial products focused on creating variants (mutants) using chemical mutagens or radiation to induce changes in the genetic constitution of existing strains. The level of improvement that could be achieved in this way was usually limited biologically. If, for example, a bacterium was mutated to produce high levels of a compound, other metabolic functions often were impaired, thereby causing the bacterium's growth during large-scale fermentation to be less than desired. Despite this constraint, the traditional "induced mutagenesis and selection" strategies of strain improvement were extremely successful for a number of processes, such as the production of increased levels of antibiotics.

The traditional genetic improvement regimens were tedious, time-consuming, and costly because of the large numbers of microbial cells that had to be screened and tested. The best result that could be expected with this approach was the improvement of an existing inherited property of a microorganism rather than the expansion of its genetic capabilities. Despite these limitations, by the late 1970s, effective processes for the mass production of a wide range of commercial products from microorganisms had been perfected.

Today we have acquired sufficient knowledge of the biochemistry, genetics, and molecular biology of microbes and other organisms to significantly accelerate the development of useful and improved biological products and processes and to create new products that would not otherwise occur. Distinct from traditional biotechnology, the modern methods require knowledge of and manipulation of genes, the functional units of inheritance, and the discipline that is concerned with the manipulation of genes for the purpose of producing useful goods and services using living organisms is known as molecular biotechnology. The pivotal developments that enabled this technology were the establishment of techniques to isolate genes and to transfer them from one organism to another. The joining of DNA molecules from different sources was first demonstrated in 1971 by biochemist Paul Berg at Stanford University, who inserted genes from the bacterial

virus (bacteriophage) lambda into simian virus 40 DNA. This technology is known as recombinant DNA technology, and it was further developed by two scientists working in different fields who met at a scientific conference in 1972. In his laboratory at Stanford University in California, Stanley Cohen had been developing methods to transfer plasmids, small circular DNA molecules that replicate independently of chromosomal DNA, into bacterial cells. Meanwhile, Herbert Boyer at the University of California at San Francisco was working with enzymes that cut DNA at specific nucleotide sequences. Over lunch at a scientific meeting in Hawaii, they reasoned that Boyer's enzyme could be used to splice a specific segment of DNA into a plasmid and then the modified (recombinant) plasmid could be introduced into a host bacterium using Cohen's method.

Recombinant DNA Technology

It was clear to Cohen and Boyer, and others, that recombinant DNA technology had far-reaching possibilities. As Cohen noted at the time, "It may be possible to introduce in *E. coli*, genes specifying metabolic or synthetic functions such as photosynthesis or antibiotic production indigenous to other biological classes." The first commercial product produced using recombinant DNA technology was human insulin, which is used in the treatment of diabetes. The DNA sequence that encodes human insulin was synthesized, a remarkable feat in itself at the time, and was inserted into a plasmid that could be maintained in a nonpathogenic strain of the bacterium *Escherichia coli*. The bacterial host cells acted as biological factories for the production of the two peptide chains of human insulin that could be purified and combined and used to treat diabetics who were allergic to the commercially available porcine (pig) insulin. Today, this type of genetic engineering is commonplace.

The nature of biotechnology was changed forever by the development of recombinant DNA technology. Genetic engineering provided the means to create, rather than merely isolate, highly productive microbes and other organisms. Not long after the production of the first commercial preparation of recombinant human insulin in 1982, bacteria and then eukaryotic cells were used for the production of other therapeutic proteins, such as interferon, growth hormone, and viral antigens. Recombinant DNA technology also facilitated the biological production of large amounts of useful low-molecular-weight compounds and macromolecules that occur naturally in minuscule quantities. Plants and animals became natural bioreactors for producing new or altered gene products that could never have been created either by mutagenesis and selection or by crossbreeding. From its modest beginnings, around 50 years ago, molecular biotechnology has become the standard method for developing living systems with novel functions and capabilities for the synthesis of thousands of important commercial products.

Most new scientific disciplines do not arise solely on their own. They are often formed by the synthesis of knowledge from different areas of research. For molecular biotechnology, the biotechnology component was perfected by industrial microbiologists and chemical engineers, whereas the recombinant DNA technology portion owes much to discoveries in molecular biology, bacterial genetics, and nucleic acid enzymology (Table 1.1). In a broad sense, molecular biotechnology draws on knowledge from a diverse set of fundamental scientific disciplines to create products that are useful in a wide range of applications (Fig. 1.1).

Table 1.1 Selected developments in the history of molecular biotechnology[a]

Date	Event
1917	Hungarian agricultural engineer Karl Ereky coins the term "biotechnology"
1940	Danish microbiologist A. Jost coins the term "genetic engineering" in a lecture on sexual reproduction in yeast
1943	Penicillin is produced on an industrial scale
1944	Avery, MacLeod, and McCarty demonstrate that DNA is the genetic material
1953	Watson and Crick determine the structure of DNA
1961–1966	Entire genetic code deciphered
1970	First restriction endonuclease isolated
1972	Khorana and coworkers synthesize an entire tRNA gene
1973	Boyer and Cohen establish recombinant DNA technology
1975	Kohler and Milstein describe the production of monoclonal antibodies
1976	First guidelines for the conduct of recombinant DNA research are issued
1976	Techniques are developed to determine the sequence of DNA
1978	Genentech produces human insulin in *E. coli*
1980	U.S. Supreme Court rules in the case of Diamond vs Chakrabarty that genetically manipulated microorganisms can be patented
1981	First commercial, automated DNA synthesizers are sold
1981	First monoclonal antibody-based diagnostic kit is approved for use in the United States
1982	First animal vaccine produced by recombinant DNA methodologies is approved for use in Europe
1983	Engineered Ti plasmids are used to transform plants
1988	U.S. patent is granted for a genetically engineered mouse susceptible to cancer
1988	PCR method is published
1990	Approval is granted in the United States for a trial of human somatic cell gene therapy
1990	Recombinant chymosin is used for cheese making in the United States
1994–1995	Detailed genetic and physical maps of human chromosomes are published
1994	FDA announces that genetically engineered tomatoes are as safe as conventionally bred tomatoes
1995	First genome of a cellular organism, the bacterium *Haemophilus influenzae*, is sequenced
1996	Recombinant protein erythropoietin exceeds $1 billion in annual sales
1996	Complete DNA sequence of all the chromosomes of a eukaryotic organism, the yeast *Saccharomyces cerevisiae*, is determined
1996	Commercial planting of genetically modified crops
1997	A sheep is cloned by somatic cell nuclear transfer
1998	FDA approves first antisense drug
1999	FDA approves recombinant fusion protein (diphtheria toxin–interleukin-2) to treat cutaneous T-cell lymphoma
2000	*Arabidopsis* genome is sequenced
2000	Monoclonal antibodies exceed $2 billion in annual sales
2000	Development of "Golden Rice" (provitamin-A-producing rice) is announced
2001	Human genome sequence is published
2002	Complete human gene microarrays are commercially available
2002	FDA approves first nucleic acid test to screen whole blood from donors for HIV and HCV
2004	Large-scale sequencing of the Sargasso Sea metagenome
2005	NCBI announces 100 gigabases of nucleotides in GenBank sequence database
2006	Recombinant cancer vaccine (Gardasil) is available to protect against cervical cancer
2009	FDA approves first drug produced in a genetically engineered animal (goat)

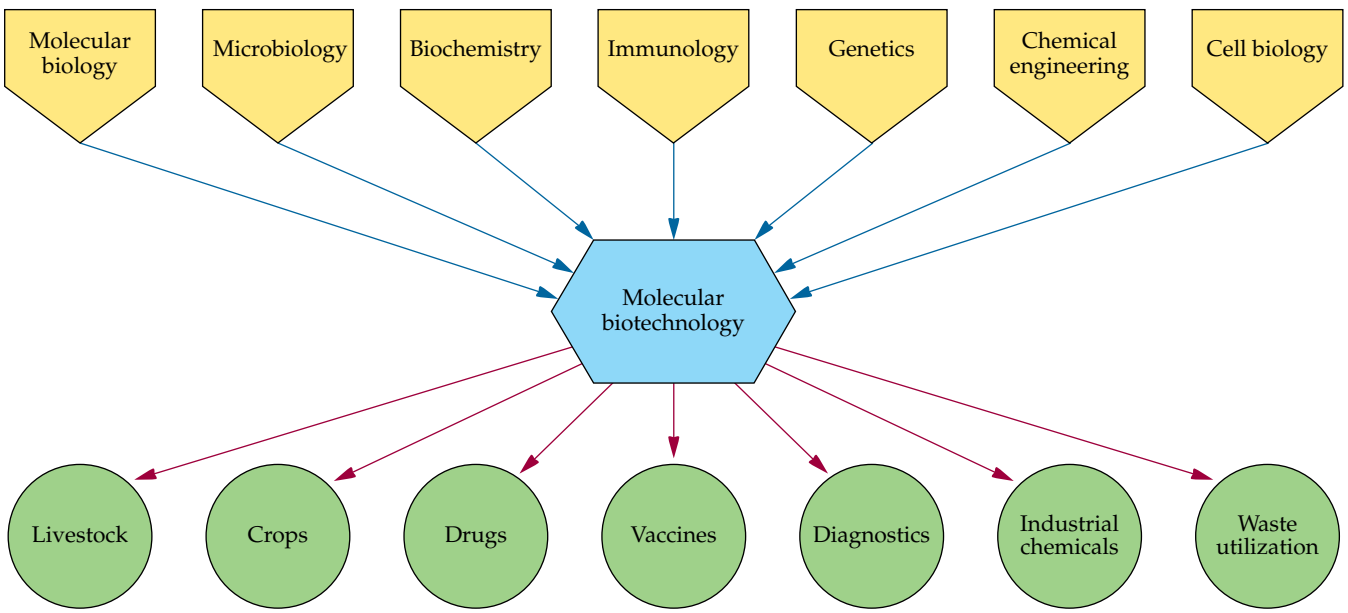
(continued)

Table 1.1 *(continued)*

Date	Event
2009	First clinical trial using embryonic stem cells
2010	Researchers create the first synthetic cell
2012	Doudna and Charpentier demonstrate targeted cleavage of DNA by CRISPR-Cas system
2013	U.S. Supreme Court rules that isolated genes are not eligible for patenting
2014	Patent granted for CRISPR-Cas systems and methods for altering expression of gene products
2015	FDA approves first transgenic animal (salmon) for human consumption
2016	National Bioengineered Food Disclosure Law is passed by the U.S. Congress requiring labeling of genetically engineered food
2017	FDA approves *in vivo* gene therapy for treatment of an inherited form of retinal dystrophy
2018	First RNA interference therapy (to treat hereditary transthyretin amyloidosis) approved in the United States and Europe
2018	Birth of genome-edited babies in China triggers widespread criticism
2019	Genetically engineered crops are grown in 29 countries on 190 million hectares
2019	Clinical trials begin to test the safety and efficacy of CRISPR-Cas genome editing to treat two blood disorders, β-thalassemia and sickle cell anemia
2019	Philippines is first Asian country to receive safety approval for genetically modified Golden Rice
2020	Results from the first CRISPR clinical trial indicate that the technique is safe
2020	First report of genome-edited mitochondria
2020	COVID-19 pandemic impels development of innovative molecular diagnostic tests, treatments, and vaccines
2021	Trial release of genetically modified mosquitoes in the United States

[a]FDA, Food and Drug Administration; HCV, hepatitis C virus; HIV, human immunodeficiency virus; NCBI, National Center for Biotechnology Information; NIH, National Institutes of Health; PCR, polymerase chain reaction; tRNA, transfer ribonucleic acid; CRISPR, clustered regularly interspaced short palindromic repeats.

Figure 1.1 Many scientific disciplines contribute to molecular biotechnology, which generates a wide range of commercial products.



Construction of Biologically Functional Bacterial Plasmids *In Vitro*

S. N. Cohen, A. C. Y. Chang, H. W. Boyer, and R. B. Helling
Proc. Natl. Acad. Sci. USA **70:**3240–3244, 1973

milestone

The landmark study of Cohen et al. established the scientific foundation for recombinant DNA technology by showing how genetic information from different sources could be joined to create a novel, replicable genetic structure. In this case, the new genetic entities were derived from bacterial autonomously replicating extrachromosomal DNA structures called plasmids. In a previous study, Cohen and Chang (*Proc. Natl. Acad. Sci. USA* **70:**1293–1297, 1973) produced a small plasmid from a large naturally occurring plasmid by shearing the larger plasmid into smaller random pieces and introducing the mixture of pieces into a host cell, the bacterium *E. coli*. One of the fragments that was about 1/10 the size of the original plasmid was perpetuated as a functional plasmid. To overcome the randomness of this approach and to make the genetic manipulation of plasmids more manageable, Cohen and his coworkers used an enzyme (restriction endonuclease) that cuts a DNA molecule at a specific nucleotide sequence and produces a short single-stranded extension at each end. The extensions of the cut ends of a restriction endonuclease-treated DNA molecule can combine (due to complementarity of the nitrogenous bases) with the extensions of another DNA molecule that has been cleaved with the same restriction endonuclease. Consequently, when DNA molecules from different sources are treated with the same restriction endonuclease and mixed together, new DNA combinations (recombinant DNA) that never existed before can be formed. In this way, Cohen et al. not only introduced a gene from one plasmid into another plasmid but also demonstrated that the introduced gene was biologically active. To their credit, these authors fully appreciated that their strategy was "potentially useful for insertion of specific sequences from prokaryotic or eukaryotic chromosomes or extrachromosomal DNA into independently replicating bacterial plasmids." In other words, any gene from any organism could theoretically be cloned into a plasmid which, after introduction into a host cell, would be maintained indefinitely and, perhaps, produce the protein encoded by the cloned gene. By demonstrating the feasibility of gene cloning, Cohen et al. provided the experimental basis for recombinant DNA technology and established that plasmids could act as vehicles (vectors) for maintaining cloned genes. This motivated others to pursue research in this area that rapidly led to the development of more sophisticated vectors and gene cloning strategies. It also engendered concerns about the safety and ethics of this kind of research that, in turn, were responsible for the establishment of official guidelines and governmental agencies for conducting and regulating recombinant DNA research, respectively, and contributed to the formation of the molecular biotechnology industry.

The Cohen and Boyer strategy for gene cloning was an experiment "heard round the world" (see Milestone). Once their concept was made public, many other researchers immediately appreciated the power of its potential. Consequently, scientists created a large variety of experimental protocols that made identifying, isolating, characterizing, and utilizing genes more efficient and relatively easy. These technological developments have had an enormous impact on generating new knowledge in practically all biological disciplines. Indeed, the emergence of the field of genomics was dependent on the ability to clone large fragments of DNA into plasmids in preparation for sequence determination.

Commercialization of Molecular Biotechnology

The potential of recombinant DNA technology reached the public with a frenzy of excitement, and many people became rich on its promise. Indeed, within 20 minutes of the start of trading on the New York Stock Exchange on 14 October 1980, the price of shares in Genentech, the company founded

by Boyer with chemist and entrepreneur Robert Swanson that produced recombinant human insulin, went from $35 to $89. This was the fastest increase in the value of any initial public offering in the history of the market up to that time. It was predicted that some genetically engineered microorganisms would replace chemical fertilizers and others would eat up oil spills; plants with inherited resistance to a variety of pests and exceptional nutritional content would be created; and livestock would have faster growing times, more efficient feed utilization, and meat with low fat content. Many were convinced that as long as a biological characteristic was genetically determined by one or a few genes, organisms with novel genetic constitutions could be readily created. Today, in many cases, the promise of recombinant DNA technology has become a reality.

Since the commercial production of recombinant human insulin, hundreds of drugs produced by recombinant DNA technology have been developed to treat diseases such as cancer, multiple sclerosis, rheumatoid arthritis, cystic fibrosis, and strokes, and to provide protection against numerous infectious diseases. The majority of these are therapeutic monoclonal antibodies, hormones, and growth factors, many of which are more effective and have fewer side effects than other therapies. Very recently, a small number of nucleic acid therapies have been approved to target diseases caused by specific mutations, including gene therapies to replace defective genes that cause blindness and spinal muscular atrophy, and RNA therapies to treat an inherited form of amyloidosis and Duchenne muscular dystrophy. Hundreds more new biological drugs and therapies are in the process of being tested in human clinical trials, most to treat various cancers and also genetic, neurological, autoimmune, and infectious diseases. Beyond medical applications, many molecular biotechnology products are available to enhance crop and livestock yields, decrease pesticide use, and improve industrial processes such as the manufacture of pulp and paper, food, energy, and textiles.

The impact on agriculture has been tremendous. While the global population is expanding rapidly, yield increases of all major crops have decreased due to poor agricultural management practices, decreased acreage of arable land, and increased reliance on fertilizers and pesticides that diminish soil quality. To produce more food on less land, in 2019, 17 million farmers in 29 countries planted genetically engineered crops on 190 million hectares of land. These crops are predominantly soybeans, corn, cotton, and canola that are resistant to herbicides and insects, although many others such as drought-resistant sugarcane and nonbrowning potatoes and apples are produced. The global market value of genetically modified crops is currently around $20 billion. Small resource-poor farmers benefit substantially from agricultural biotechnology. In a comparative study of small cotton farms in South Africa, it was found, over three seasons, that the yield of cotton from plants that were genetically engineered to produce a bacterial insecticide was on average about 70% greater than those from non-genetically modified plants. Higher yields and reduced pesticide and labor costs translated into doubled revenues despite the slightly higher costs of the transgenic seeds. In a recent 2-year study in Bangladesh, the net returns per hectare for insect-resistant eggplant were six times higher than that for conventional eggplant.

The ultimate objective of all biotechnology research is the development of commercial products. Consequently, molecular biotechnology is driven

to a great extent by the prospect of financial gain. By nightfall on 14 October 1980, the principal shareholders of Genentech stock were worth millions of dollars. The unprecedented enthusiastic public response to Genentech encouraged others to follow. Between 1980 and 1983, about 200 small biotechnology companies were founded in the United States with the help of tax incentives and funding from both stock market speculation and private investment. Like Herbert Boyer, who was first a research scientist at the University of California at San Francisco and then a vice president of Genentech, university professors started many of the early companies.

Today, there are about 8,000 biotechnology companies worldwide, most in the United States and Europe, with annual earnings in the hundreds of billions of dollars. The biotechnology industry in these regions employs more than 200,000 people. Large multinational chemical and pharmaceutical companies, such as Bayer, DuPont, Pfizer, GlaxoSmithKline, Merck, Novartis, Hoffmann-LaRoche, Gilead Sciences, and Amgen, to name but a few, have made significant research commitments to molecular biotechnology. The roster of biotechnology companies is extensive and includes those focused on vaccines, protein and nucleic acid therapeutics, drug delivery, molecular diagnostics, genomics, industrial processing, and agricultural biotechnology.

Concerns and Consequences

While many people appreciate the potential of molecular biotechnology to solve important problems in agriculture, medicine, and industry, they recognize the need to be cautious about its widespread application. Indeed, one of the first scientific responses to recombinant DNA technology was a voluntary moratorium on certain experiments that were thought to be potentially hazardous. This research ban was self-imposed by a group of molecular biologists, including Cohen and Boyer. They were concerned that combining genes from two different organisms might unintentionally create a novel organism with undesirable or even dangerous properties. Within a few years, however, these apprehensions were allayed as scientists gained laboratory experience with the technology and safety guidelines were formulated for recombinant DNA research. The temporary cessation of some recombinant DNA research projects did not dampen the enthusiasm for genetic engineering. In fact, the new technology continued to receive unprecedented attention from both the public and the scientific community.

Molecular biotechnology can benefit humanity by

- Providing opportunities to accurately diagnose, prevent, treat, or cure a wide range of infectious and genetic diseases
- Increasing crop yields by creating plants that are resistant to insect predation, fungal and viral diseases, and environmental stresses such as short-term drought and excessive heat and at the same time reducing applications of hazardous agrichemicals
- Creating microorganisms that will produce metabolites (products of metabolism), polymers, amino acids, enzymes, and additives that are important for food production and other industries

- Developing livestock and other animals that have genetically enhanced attributes
- Facilitating the removal of pollutants and waste materials from the environment

Although it is exciting and important to emphasize the positive aspects of new advances, there are also social concerns and consequences that must be addressed. For example,

- Will some genetically engineered organisms, or their products, be harmful to humans or other organisms, or to the environment?
- Will the development and use of genetically engineered organisms reduce natural biological diversity?
- Should humans be genetically manipulated?
- Will new diagnostic procedures, especially those based on genome sequencing, undermine individual privacy?
- Will the emphasis on commercial success mean that the benefits of molecular biotechnology will be available only to wealthy individuals or nations?
- Will agricultural molecular biotechnology undermine traditional farming practices and limit food choices?
- Will medical therapies based on molecular biotechnology supersede equally effective traditional treatments?
- Will the quest for patents inhibit the free exchange of ideas among research scientists?

These and many other issues have been considered by government commissions, discussed extensively at conferences, and thoughtfully debated and analyzed by individuals in both popular and academic publications. On this basis, regulations have been formulated, guidelines have been established, and policies have been created. There has been active and extensive participation by both scientists and the general public in deciding how molecular biotechnology should proceed, although some controversies still remain.

Molecular biotechnology, with much fuss and fanfare, became a comprehensive scientific and commercial venture in a remarkably short period of time. Many scientific and business publications are now devoted to molecular biotechnology, and graduate and undergraduate programs and courses are available at universities throughout the world to teach molecular biotechnology. It could be debated whether the early promise of biotechnology has been fulfilled as it was predicted to in a 1987 document published by the U.S. Office of Technology Assessment, which declared that molecular biotechnology is "a new scientific revolution that could change the lives and futures of . . . citizens as dramatically as did the Industrial Revolution two centuries ago and the computer revolution today. The ability to manipulate genetic material to achieve specified outcomes in living organisms . . . promises major changes in many aspects of modern life." It does offer solutions to some serious global problems, including the spread of infectious diseases, the burden of waste accumulation, and food shortages that may become increasingly dire as the climate changes. The potential of molecular biotechnology to solve some of these imminent problems is the subject of this book.

summary

In 1973, Stanley Cohen, Herbert Boyer, and their coworkers devised a method for transferring genetic information (genes) from one organism to another. This procedure, which became known as recombinant DNA technology, enabled researchers to isolate specific genes and perpetuate them in host organisms. Recombinant DNA technology has been beneficial to many different areas of study. However, its impact on biotechnology has been extraordinary.

Biotechnology uses organisms, often on a large scale, for the production of commercially important products. Before the advent of recombinant DNA technology, the most effective way of increasing the productivity of an organism was to induce mutations and then use selection procedures to identify individuals with the desired traits. This process was not foolproof; it was time-consuming, labor-intensive, and costly, and only a small set of traits could be enhanced in this way. Recombinant DNA technology, however, provided a rapid, efficient, and powerful means for creating organisms with specific and expanded genetic attributes. The tools of recombinant DNA technology enable microorganisms, plants, and animals to be genetically engineered. Combining recombinant DNA technology with biotechnology created a dynamic and exciting discipline called molecular biotechnology.

A large number of molecular biotechnology products are currently available. Many vaccines, and protein and nucleic acid therapeutics are produced in genetically engineered microorganisms, mammalian cells, or transgenic plants and animals and then used to treat a variety of diseases in humans and animals. Molecular diagnostic tests produced through recombinant DNA technology detect specific proteins or nucleic acid sequences that indicate susceptibility to or progress of a disease, or response to a treatment. Proteins or small molecules of industrial importance are often produced on a large scale by genetically engineered microorganisms. Increased agricultural output has been possible due to the development of transgenic plants that are resistant to pests and pathogens and to abiotic stresses such as drought and salt. A significant milestone was the approval of the first genetically engineered animal (a salmon) for human consumption. The salmon, which has an enhanced growth rate, was available to U.S. consumers in 2020. Many more exciting solutions to medical, environmental, industrial, and agricultural problems are under development.

Because of their broad impact, molecular biotechnology products are scrutinized carefully for their potential effects on society. Some of the concerns are the impact on the health of humans and other organisms, possible negative effects on the environment, and the private or public ownership of genetically engineered organisms.

REFERENCES

Anonymous. 1987. *New Developments in Biotechnology—Background Paper: Public Perceptions of Biotechnology*. Office of Technology Assessment, US Congress, US Government Printing Office, Washington, DC.

Baeshen NA, Baeshen MN, Sheikh A, Bora RS, Ahmed MMM, Ramadan HAI, Saini KS, Redwan EM. 2014. Cell factories for insulin production. *Microb Cell Fact* **13:**141–149.

Cohen SN, Chang AC. 1973. Recircularization and autonomous replication of a sheared R-factor DNA segment in *Escherichia coli* transformants. *Proc Natl Acad Sci USA* **70:**1293–1297.

Cohen SN, Chang ACY, Boyer HW, Helling RB. 1973. Construction of biologically functional bacterial plasmids *in vitro*. *Proc Natl Acad Sci USA* **70:**3240–3244.

Jackson DA, Symons RH, Berg P. 1972. Biochemical method for inserting new genetic information into DNA of simian virus 40: circular SV40 DNA molecules containing lambda phage genes and the galactose operon of *Escherichia coli*. *Proc Natl Acad Sci USA* **69:**2904–2909.

James C. 2018. *Global status of commercialized biotech/GM crops in 2018. International Service for the Acquisition of Agri-Biotech Applications (ISAAA) Brief No. 54*. ISAAA, Ithaca, N.Y.

Morse S, Bennett R, Ismael Y. 2004. Why *Bt* cotton pays for small-scale producers in South Africa. *Nat Biotechnol* **22:**379–380.

OECD Internal Co-ordination Group for Biotechnology. 2019. *Biotechnology Update*. OECD Publishing, Paris, France.

Shelton AM, Hossain MJ, Paranjape V, Azad AK, Rahman ML, Khan ASMMR, Prodhan MZH, Rashid MA, Majumder R, Hossain MA, Hussain SS, Huesing JE, McCandless L. 2018. Bt eggplant project in Bangladesh: history, present status, and future direction. *Front Bioeng Biotechnol* **6:**106.

review questions

1. What is biotechnology?
2. Distinguish between traditional biotechnology and molecular biotechnology.
3. What are the shortcomings of the "mutation and selection" method for developing enhanced organisms for commercial purposes?
4. Why was the work reported by Cohen and Boyer and their coworkers in 1973 considered important?
5. How did recombinant DNA technology enable the production of human insulin?
6. What are some of the problems that molecular biotechnology has the potential to solve?
7. Discuss the statement "Molecular biotechnology is a diverse science."
8. Discuss some of the social concerns that have been raised about molecular biotechnology.
9. Go to your favorite news website and conduct a search with the word "biotechnology." Describe and discuss three recent biotechnology news stories.

Molecular Cloning
Preparation of DNA for Cloning
Insertion of Target DNA into a Plasmid Vector
Transformation and Selection of Cloned DNA in a Bacterial Host
Cloning Eukaryotic Genes
Recombinational Cloning
Genomic Libraries
Genome Engineering Using CRISPR Technology
Polymerase Chain Reaction
Amplification of DNA by PCR
Cloning PCR Products
Quantitative PCR
Chemical Synthesis of Genes
Assembling Oligonucleotides into Genes
Assembling PCR Products into Genes
DNA Sequencing Technologies
Dideoxynucleotide Sequencing
Sequencing Using Reversible Chain Terminators
Single-Molecule Real-Time Sequencing
Nanopore Sequencing
Sequencing Whole Genomes
Preparation of Genomic DNA Sequencing Libraries
High-Throughput Next-Generation Sequencing
Genome Sequence Assembly
Sequencing Metagenomes
Genomics
Transcriptomics
Proteomics
Metabolomics
SUMMARY
REFERENCES
REVIEW QUESTIONS

2 Fundamental Technologies

Molecular Cloning

Molecular biotechnology uses a variety of techniques for isolating genes and transferring them from one organism to another. At the root of these technologies is the ability to join a sequence of deoxyribonucleic acid (DNA) of interest to a vector that can then be introduced into a suitable host. This process is known as recombinant DNA technology or molecular cloning. A vast number of variations on this basic process have been devised to produce products that are important for medicine, food production, environmental remediation, and industrial processes.

Preparation of DNA for Cloning

In theory, DNA from any organism can be cloned. The target DNA may be obtained directly from genomic DNA, derived from messenger ribonucleic acid (mRNA), subcloned from previously cloned DNA, or synthesized *in vitro*. The target DNA may contain the complete coding sequence for a protein, a part of the protein coding sequence, a random fragment of genomic DNA, or a segment of DNA that contains regulatory elements that control expression of a gene. Prior to cloning, both the source DNA that contains the target sequence and the cloning vector must be cut into discrete fragments, predictably and reproducibly, so that they can be joined (ligated) together to form a stable molecule. Bacterial enzymes known as type II restriction endonucleases, or (more commonly) restriction enzymes, are used for this purpose. These enzymes recognize and cut

Molecular Biotechnology: Principles and Applications of Recombinant DNA, Sixth Edition. Bernard R. Glick and Cheryl L. Patten.
© 2022 American Society for Microbiology.

Cleavage of DNA by a Restriction Endonuclease Generates Cohesive Ends

milestone

Recombinant DNA technology requires a vector to carry cloned DNA, the specific joining of vector and cloned (insert) DNA molecules to form a vector–insert DNA construct, the introduction of the vector–insert DNA construct into a host cell, and the identification of host cells that acquire the cloned DNA. The discovery of type II restriction endonucleases facilitated cloning genes into vectors. In 1968, M. Meselson and R. Yuan (*Nature* **217:**1110–1114) showed that the capability of a strain of *E. coli* to prevent (restrict) the development of a bacterial virus (bacteriophage) was due to a host cell enzyme that cleaved the DNA of the infecting bacteriophage. The study done by Mertz and Davis (*Proc. Natl. Acad. Sci. USA* **69:**3370–3374, 1972) established that the RI restriction endonuclease from *E. coli*, which is now called EcoRI, cut DNA at a specific site and produced complementary extensions. Briefly, they showed that after circular DNA was linearized by treatment with EcoRI, some of the molecules formed hydrogen-bonded circular DNA molecules which were converted to covalently closed circular DNA molecules by treating the sample with a DNA ligase. The extensions of all of the cut DNA molecules were the same and were estimated to be 4 to 6 nucleotides long, with the recognition site being 6 nucleotide pairs. Mertz and Davis concluded that ". . . any two DNA molecules with RI sites can be 'recombined' at their restriction sites by the sequential action of RI endonuclease and DNA ligase to generate hybrid DNA molecules." The discovery that EcoRI created cohesive ends was one of the most important contributions to the development of recombinant DNA technology because it provided, according to Mertz and Davis, a "simple way . . . to generate specifically oriented recombinant DNA molecules *in vitro*."

double-stranded DNA molecules at specific base pair sequences and are produced naturally by bacteria to cleave foreign DNA, such as that of infecting bacterial double-stranded DNA viruses (bacteriophages). A bacterium that produces a specific restriction endonuclease also has a corresponding system to modify the sequence recognized by the restriction endonuclease in its own DNA to protect it from being degraded. Often, methylation of a cytosine within a restriction endonuclease-binding site prevents the enzyme from cutting at these sites.

Table 2.1 Recognition sequences of some restriction endonucleases

Enzyme	Recognition site[a]	Type of cut end
EcoRI	G↓A–A–T–T–C C–T–T–A–A↑G	5′ phosphate extension
BamHI	G↓G–A–T–C–C C–C–T–A–G↑G	5′ phosphate extension
PstI	C–T–G–C–A↓G G↑A–C–G–T–C	3′ hydroxyl extension
Sau3AI	↓G–A–T–C C–T–A–G↑	5′ phosphate extension
PvuII	C–A–G↓C–T–G G–T–C↑G–A–C	Blunt end
HpaI	G–T–T↓A–A–C C–A–A↑T–T–G	Blunt end
HaeIII	G–G↓C–C C–C↑G–G	Blunt end
NotI	G↓C–G–G–C–C–G–C C–G–C–C–G–G–C↑G	5′ phosphate extension

[a]Arrows denote cleavage sites.

A large number of restriction endonucleases from different bacteria are available to facilitate cloning. The sequence and length of the recognition site in the DNA vary among the different enzymes and can be four or more nucleotide pairs (Table 2.1). One of the first restriction endonucleases to be characterized was from the bacterium *Escherichia coli*, designated EcoRI. The name of a restriction endonuclease indicates the genus (capitalized letter), species (first two letters in lower case), and occasionally the strain or serotype (e.g., R in EcoRI indicates that it is from the RY13 strain) of the source bacterium, as well as the order of characterization of different restriction endonucleases from the same bacterium (Roman numerals). As for most restriction endonucleases, EcoRI is a homodimeric protein (made up of two identical polypeptides) that recognizes and binds to a specific, palindromic DNA sequence (Fig. 2.1A). In DNA, a palindrome is a sequence of nucleotides in each of the two strands that is identical when either is read in the same direction, i.e., 5′ to 3′. The EcoRI recognition sequence consists of 6 base pairs (bp) and is cut between the guanine and adenine residues on each strand (Fig. 2.1A). Specifically, it cleaves the bond between the oxygen attached to the 3′ carbon of the sugar of one nucleotide and the phosphate group attached to the 5′ carbon of the sugar of the adjacent nucleotide. The symmetrical staggered cleavage of DNA by EcoRI produces two single-stranded, complementary ends, each with extensions of four nucleotides, often referred to as sticky ends. Each single-stranded extension terminates with a 5′ phosphate group, and the 3′ hydroxyl group of the opposite strand is recessed (Fig. 2.1A). Some other restriction endonucleases, such as PstI, leave 3′ hydroxyl extensions with recessed 5′ phosphate ends (Fig. 2.1B), while others, such as SmaI, cut the backbone of both strands within a recognition site to produce so-called blunt-ended DNA molecules (Fig. 2.1C).

Some restriction endonucleases isolated from different bacteria recognize and cut DNA at the same site (Fig. 2.2A). These enzymes are known as isoschizomers. Some recognize and bind to the same sequence of DNA but cleave at different positions (neoschizomers), producing different single-stranded extensions (Fig. 2.2B). Other restriction endonucleases (isocaudomers) produce the same nucleotide extensions but have different recognition sites (Fig. 2.2C). In some cases, a restriction endonuclease will cleave a sequence only if one of the nucleotides in the recognition site is methylated, while in other cases a restriction endonuclease (type IIS) binds to a specific recognition site but cuts outside of that site, a fixed number of nucleotides away from one or both ends. In the latter case, any particular sequence of nucleotides may be present between the binding sequence and the cut sites. These characteristics of restriction endonucleases are considered when designing a cloning experiment.

Many other enzymes may be used to prepare DNA for cloning. In addition to restriction endonucleases, nucleases that degrade single-stranded extensions, such as S1 nuclease and mung bean nuclease, are used to generate blunt ends for cloning (Fig. 2.3A). This is useful when the recognition sequences for restriction enzymes that produce complementary sticky ends are not available on both the vector and target DNA molecules. Blunt ends can also be produced by extending 3′ recessed ends using a DNA polymerase such as Klenow polymerase derived from *E. coli* DNA polymerase I (Fig. 2.3B). Phosphatases such as calf intestinal alkaline phosphatase cleave the 5′ phosphate groups from restriction enzyme-digested DNA (Fig. 2.3C). A 5′ phosphate group is required for formation

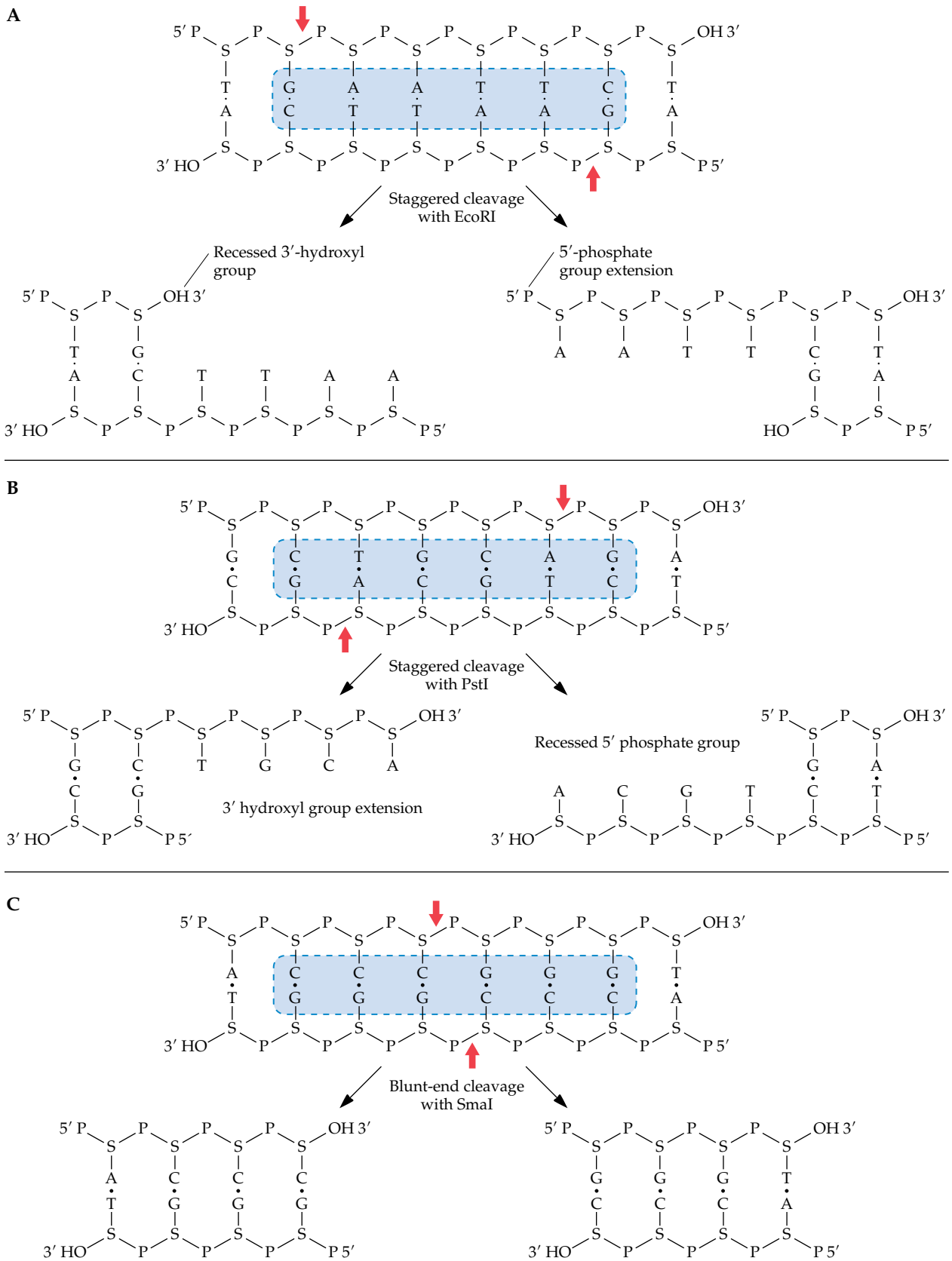


Figure 2.1 Type II restriction endonucleases bind to and cut within a specific DNA sequence. **(A)** EcoRI makes a staggered cut in the DNA strands, producing single-stranded, complementary ends (sticky ends) with a 5′ phosphate group extension. **(B)** PstI also makes a staggered cut in both strands but produces sticky ends with a 3′ hydroxyl group extension. **(C)** Cleavage of DNA with SmaI produces blunt ends. Red arrows show the sites of cleavage in the DNA backbone. S, deoxyribose sugar; P, phosphate group; OH, hydroxyl group.

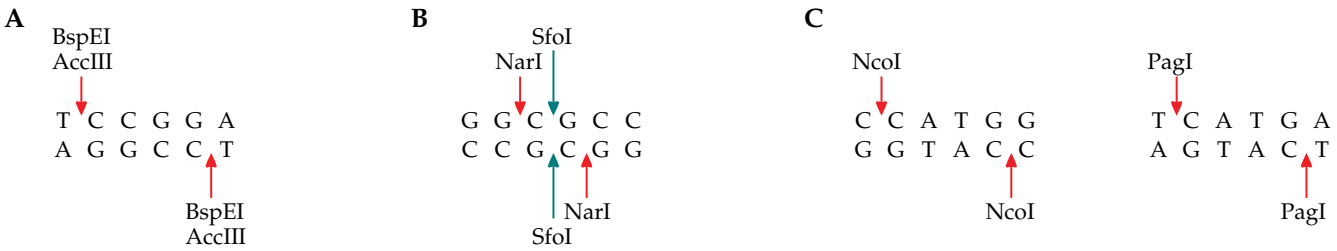
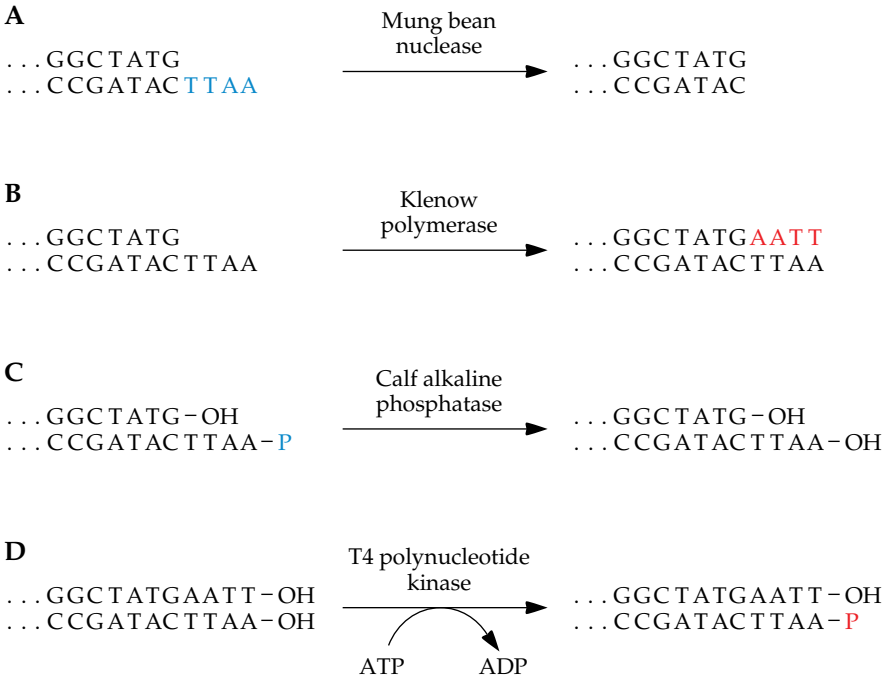


Figure 2.2 Restriction endonucleases have been isolated from many different bacteria. **(A)** Isoschizomers such as BspEI from *Bacillus* sp. and AccIII from *Acinetobacter calcoaceticus* bind the same DNA sequence and cut at the same sites. **(B)** Neoschizomers such as NarI from *Nocardia argentinensis* and SfoI from *Serratia fonticola* bind the same DNA sequence but cut at different sites. **(C)** Isocaudomers such as NcoI from *Nocardia corallina* and PagI from *Pseudomonas alcaligenes* bind different DNA sequences but produce the same sticky ends. Bases in the restriction enzyme recognition sequence are shown. Arrows show the sites of cleavage in the DNA backbone.

of a phosphodiester bond between nucleotides, and therefore, its removal prevents recircularization (self-ligation) of vector DNA. On the other hand, kinases add phosphate groups to the ends of DNA molecules. Among other activities, T4 polynucleotide kinase catalyzes the transfer of the terminal (γ) phosphate from a nucleoside triphosphate to the 5′ hydroxyl group of a polynucleotide (Fig. 2.3D). This enzyme is employed to prepare chemically

Figure 2.3 Some other enzymes used to prepare DNA for cloning. **(A)** Mung bean nuclease degrades single-stranded 5′ and 3′ extensions to generate blunt ends. **(B)** Klenow polymerase extends 3′ recessed ends to generate blunt ends. **(C)** Calf alkaline phosphatase removes the 5′ phosphate group from the ends of linear DNA molecules. **(D)** T4 polynucleotide kinase catalyzes the addition of a 5′ phosphate group to the ends of linear DNA fragments. Dotted lines indicate that only one end of the linear DNA molecule is shown.



synthesized DNA for cloning, as such DNA molecules are often missing a 5′ phosphate group required for ligation to vector DNA.

Insertion of Target DNA into a Plasmid Vector

When two different DNA molecules are digested with the same restriction endonuclease, the same sticky ends are produced in both molecules. After the two molecules are mixed together, new DNA combinations can be formed as a result of complementary base-pairing between the extended regions (Fig. 2.4). The enzyme DNA ligase, usually from the *E. coli* bacteriophage T4, is used to re-form the phosphodiester bond between the 3′ hydroxyl group and the 5′ phosphate group at the ends of DNA strands that are already held together by the hydrogen bonds between the complementary bases of the extensions (Fig. 2.4). DNA ligase also joins blunt ends, although this is generally much less efficient and typically requires a much greater amount of DNA ligase.

Ligation of restriction enzyme-digested DNA provides a means to stably insert target DNA into a vector for introduction and propagation in a suitable host cell. Many different vectors have been developed to act as carriers for target DNA. Most are derived from natural gene carriers, such as genomes of viruses that infect eukaryotic or prokaryotic cells and integrate into the host genome, or plasmids that are found in bacterial or fungal cells. Others are synthetically constructed artificial chromosomes designed for delivery of large pieces of target DNA (>100 kilobase pairs [kb]) into bacterial, yeast, or mammalian host cells. Many different vectors that carry sequences required for specific functions, such as expression of foreign DNA in a host cell, are described throughout this book. Here, vectors based on bacterial plasmids are used to illustrate the basic features of a cloning vector.

Plasmids are small, usually circular, double-stranded DNA molecules that are found naturally in many bacteria. They can range in size from less than 1 kb to more than 500 kb and are maintained as extrachromosomal entities that replicate independently of the bacterial chromosome. While they are not usually essential for bacterial cell survival under laboratory conditions, plasmids often carry genes that are advantageous under particular conditions. For example, they may carry genes that encode resistance to antibiotics or heavy metals, genes for the degradation of unusual organic compounds, or genes required for toxin production. Each plasmid has a sequence that functions as an origin (initiation site) of DNA replication which is required for it to replicate in a host cell. Some plasmids carry information for their own transfer from one cell to another.

The number of copies of a plasmid that are present in a host cell is controlled by factors that regulate plasmid replication and are characteristic of that plasmid. High-copy-number plasmids are present in 10 to more than 100 copies per cell. Low-copy-number plasmids are maintained in 1 to 4 copies per cell. When two or more different plasmids cannot coexist in the same host cell because they use the same mechanism of replication, they are said to belong to the same plasmid incompatibility group. However, plasmids from different incompatibility groups can be maintained together in the same cell. This coexistence is independent of the copy numbers of the individual plasmids. Some microorganisms have been found to contain as many as 8 to 10 different plasmids. In these instances, each plasmid can carry different genes and have its own unique copy number, and

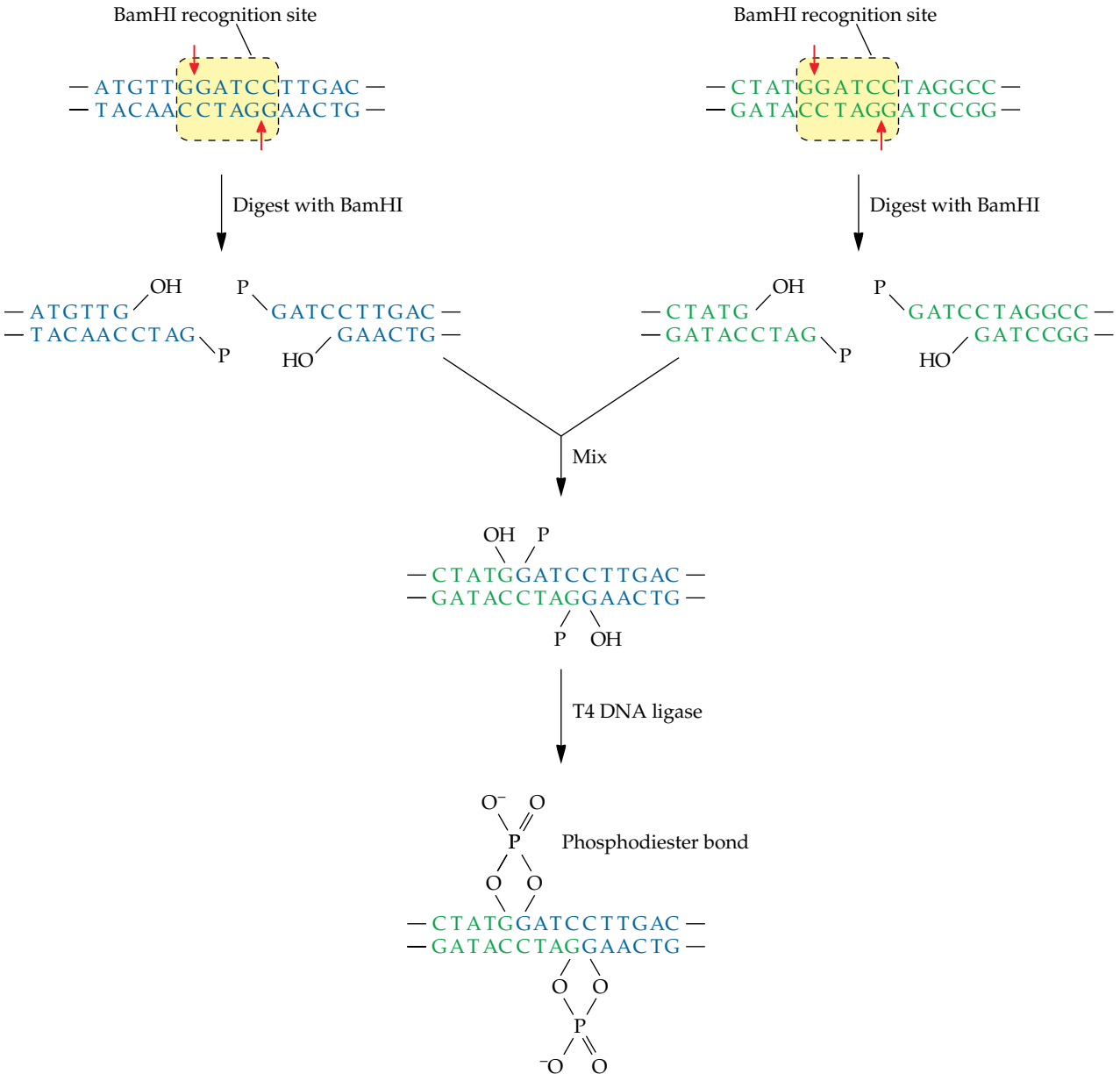


Figure 2.4 Ligation of two different DNA fragments after digestion of both with restriction endonuclease BamHI. Complementary nucleotides in the single-stranded extensions form hydrogen bonds. T4 DNA ligase catalyzes the formation of phosphodiester bonds by joining 5′ phosphate and 3′ hydroxyl groups at nicks in the backbone of the double-stranded DNA.

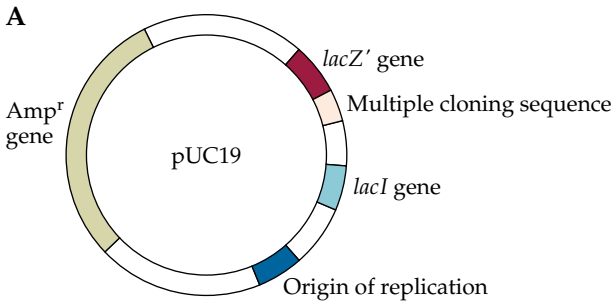
each belongs to a different incompatibility group. Some plasmids can replicate in only one (or very few) bacterial host species because they require very specific proteins for their replication as determined by their origin of replication (often denoted as *oriV* or origin of vegetative replication to distinguish it from the *oriC* or origin of chromosomal replication). These are generally referred to as narrow-host-range plasmids. On the other hand, broad-host-range plasmids have less-specific origins of replication and

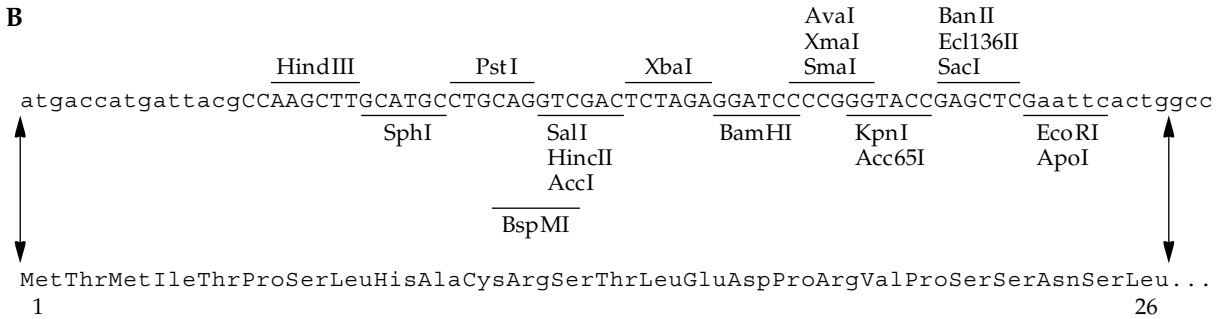
can replicate in a number of different bacterial species. The copy number, incompatibility group, and host range of a plasmid are considered when choosing a suitable vector for a molecular cloning experiment.

As autonomous, self-replicating genetic elements, plasmids are useful vectors for carrying cloned DNA. However, naturally occurring plasmids often lack several important features that are required for a good cloning vector. These include a choice of unique (single) restriction endonuclease recognition sites into which the target DNA can be inserted and one or more selectable genetic markers for identifying recipient cells that carry the cloning vector–insert DNA construct. Most of the plasmids that are currently used as cloning vectors have been genetically modified to include these features.

An example of a commonly used plasmid cloning vector is pUC19 (the lowercase p denotes a plasmid), which is derived from a natural *E. coli* plasmid. The plasmid pUC19 is 2,686 bp long, contains an origin of replication that enables it to replicate in *E. coli*, and has a high copy number, which is useful when a large number of copies of the target DNA or its encoded protein are required (Fig. 2.5A). It has been genetically engineered to possess

Figure 2.5 Plasmid cloning vector pUC19. **(A)** The plasmid contains an origin of replication for propagation in *E. coli*, an ampicillin resistance gene (Amp^r) for selection of cells carrying the plasmid, and a multiple-cloning site for insertion of cloned DNA. **(B)** The multiple-cloning site (nucleotides in uppercase letters) containing several unique restriction endonuclease recognition sites (indicated by horizontal lines) was inserted into the *lacZ′* gene (nucleotides in lowercase letters) in a manner that does not disrupt the production of a functional β-galactosidase (LacZα fragment). The first 26 amino acids of the protein are shown. Expression of the *lacZ′* gene is controlled by the LacI repressor encoded by the *lacI* gene on the plasmid. The size of the plasmid is 2,686 bp.





a short (54-bp) DNA sequence that contains many unique restriction endonuclease sites, which is called a multiple-cloning site (also known as a polylinker) (Fig. 2.5B). A DNA sequence from the lactose operon of *E. coli* has also been added that includes a segment of the β-galactosidase gene (*lacZ′*) under the control of the *lac* promoter and a *lacI* gene that produces a repressor protein that regulates the expression of the *lacZ′* gene from the *lac* promoter (Fig. 2.5A). The multiple-cloning site has been inserted within the β-galactosidase gene in a manner that does not disrupt the function of the β-galactosidase enzyme when it is expressed (Fig. 2.5B). In addition, pUC19 carries the *bla* gene (Amp^r gene) encoding β-lactamase that renders the cell resistant to ampicillin and can therefore be used as a selectable marker to identify cells that carry the vector.

To clone a gene of interest into pUC19, the vector is cut with a restriction endonuclease that has a unique recognition site within the multiple-cloning site (Fig. 2.6). The source DNA carrying the target gene is digested with the same restriction enzyme, which cuts at sites flanking the sequence of interest. The resulting linear molecules, which have the same sticky ends, are mixed together and then treated with T4 DNA ligase. A number of different ligated combinations are produced by this reaction, including the original circular plasmid DNA (Fig. 2.6). To reduce the amount of this unwanted ligation product, prior to ligation, the cleaved plasmid DNA is treated with the enzyme alkaline phosphatase to remove the 5′ phosphate groups from the linearized plasmid DNA (Fig. 2.3C). T4 DNA ligase cannot join the ends of the dephosphorylated linear plasmid DNA. However, the target DNA is not treated with alkaline phosphatase and therefore provides phosphate groups to form two phosphodiester bonds with the alkaline phosphatase-treated vector DNA (Fig. 2.6). Following treatment with T4 DNA ligase, the two phosphodiester bonds are sufficient to hold the circularized molecules together, despite the presence of two nicks (Fig. 2.6). After introduction into a host bacterium, these nicks are sealed by the host cell DNA ligase system.

Transformation and Selection of Cloned DNA in a Bacterial Host

After ligation, the next step in a cloning experiment is to introduce the vector–target DNA construct into a suitable host cell. A wide range of prokaryotic and eukaryotic cells can be used as cloning hosts; however, routine cloning procedures are often carried out using a well-studied bacterial host, usually *E. coli*. The process of taking up DNA into a bacterial cell is called transformation, and a cell that is capable of taking up DNA is said to be competent. Competence occurs naturally in many bacteria, usually when cells are stressed in high-density populations or in nutrient-poor environments, and enables bacteria to acquire new sequences that may enhance survival. Although competence and transformation are not intrinsic properties of *E. coli*, competence can be induced by various treatments.

One method to induce the uptake of plasmid DNA by a bacterial host such as *E. coli* is by treating mid-log-phase cells with ice-cold calcium chloride ($CaCl_2$), which facilitates binding of DNA to the cell membrane, and then exposing them for two minutes to a high temperature (42°C). The heat treatment increases membrane permeability that enables the bound DNA molecules to enter the cytoplasm. Alternatively, uptake of plasmid DNA can be induced by subjecting bacteria to a high-voltage

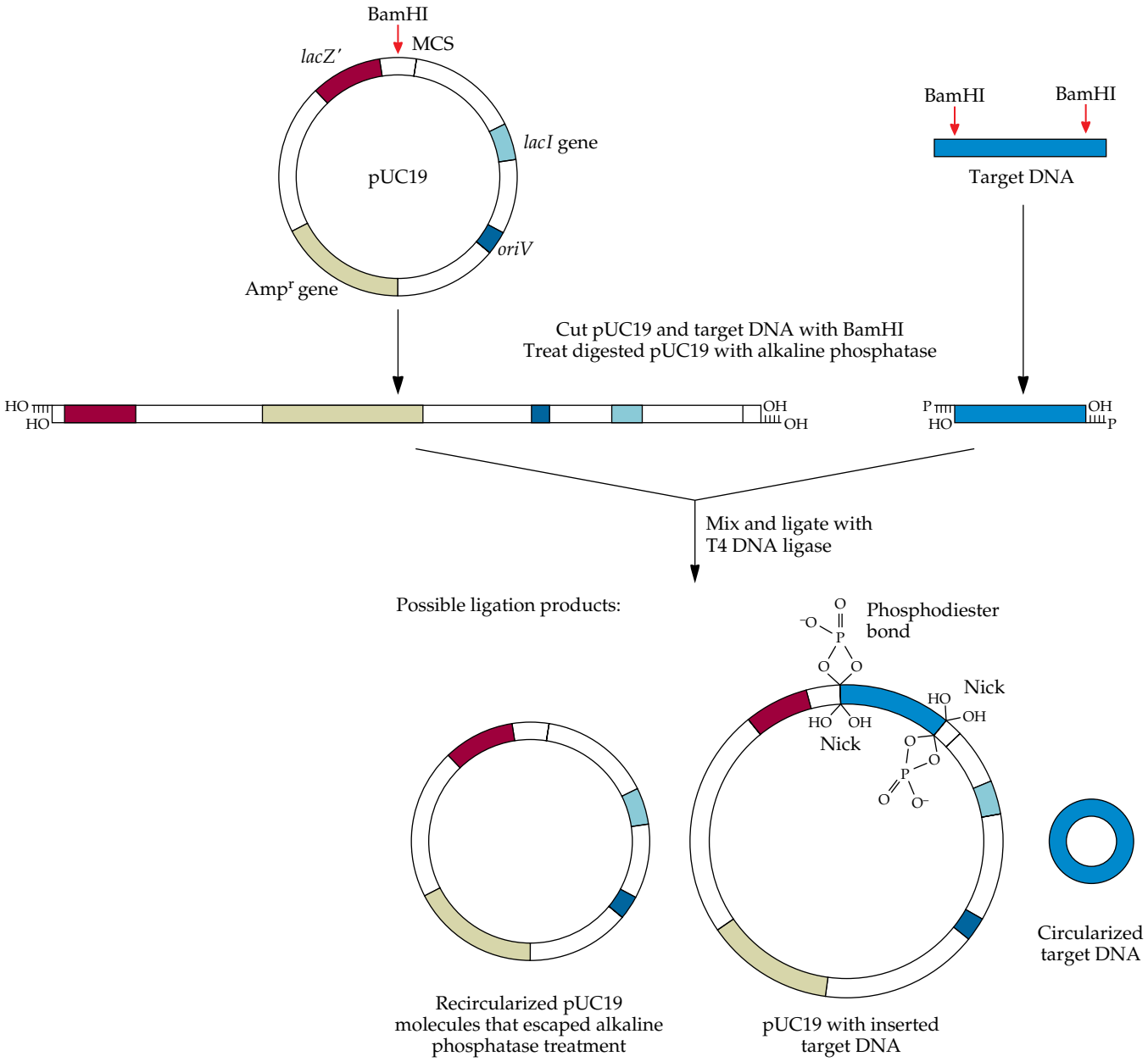


Figure 2.6 Cloning target DNA into pUC19. The restriction endonuclease BamHI cleaves pUC19 at a unique sequence in the multiple-cloning site (MCS) and at sequences flanking the target DNA. The cleaved vector is treated with alkaline phosphatase to remove 5′ phosphate groups to prevent vector recircularization. Digested target DNA and pUC19 are mixed to join the two molecules via complementary single-stranded extensions and treated with T4 DNA ligase to form a phosphodiester bond between the joined molecules. Several ligation products are possible. In addition to pUC19 inserted with target DNA, undesirable circularized target DNA molecules and recircularized pUC19 that escaped treatment with alkaline phosphatase are produced.

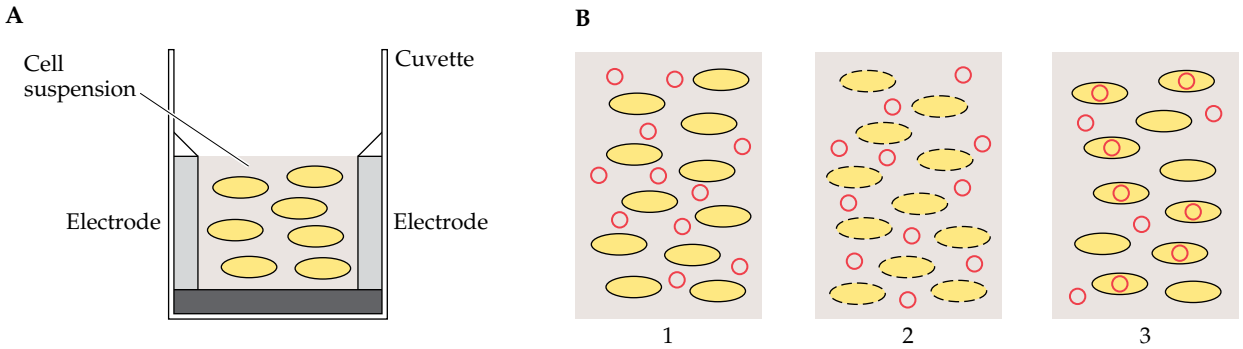


Figure 2.7 Electroporation. **(A)** Electroporation cuvette with a cell suspension between two electrodes. **(B)** (1) Cells (yellow) and DNA (red) in suspension in an electroporation cuvette prior to the administration of high-voltage electric field (HVEF) pulses. (2) HVEF pulses induce transient openings in the cells (dashed lines) that allow entry of DNA into the cells. (3) After HVEF pulsing, some cells acquire exogenous DNA.

electric field in a procedure known as electroporation. Bacterial cells and DNA are placed in a chamber fitted with electrodes (Fig. 2.7A), and a single electrical pulse of approximately 2.5 kilovolts is administered for about 5 milliseconds (ms). The electric shock induces transient pores in the cell membrane, and the DNA is taken into the cell (Fig. 2.7B). Generally, transformation is an inefficient process, and therefore, most of the cells will not have acquired a plasmid; at best, about 1 cell in 1,000 *E. coli* host cells is transformed. The integrity of the introduced plasmid DNA is also more likely to be maintained in host cells that are unable to carry out exchanges between DNA molecules because the gene encoding recombination enzyme RecA has been deleted from the host chromosome.

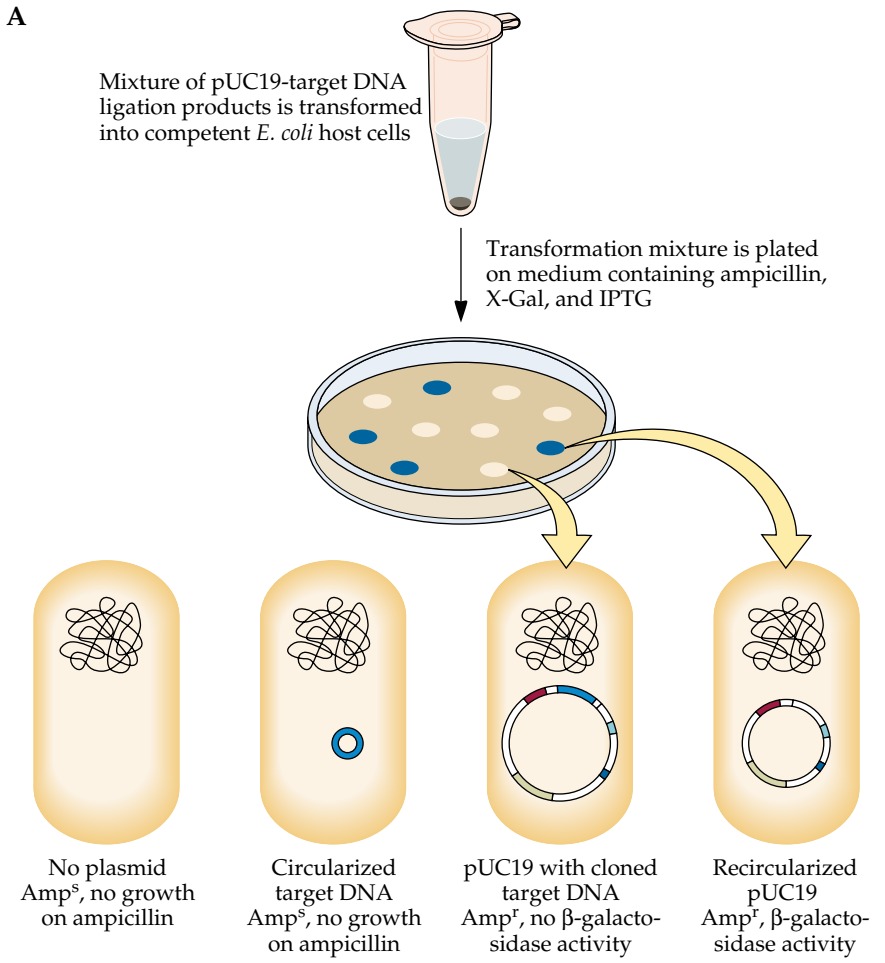
Cells transformed with vectors that carry a gene encoding resistance to an antibiotic can be selected by plating on medium containing the antibiotic. For example, cells carrying the plasmid vector pUC19, which contains the *bla* gene encoding β-lactamase, can be selected on medium containing ampicillin (Fig. 2.8A). Nontransformed cells or cells transformed with circularized target DNA cannot grow in the presence of ampicillin. However, cells transformed with the pUC19–target DNA construct and cells transformed with recircularized pUC19 that escaped dephosphorylation by alkaline phosphatase are both resistant to ampicillin. To differentiate cells carrying the desired vector–target DNA construct from those carrying the recircularized plasmid, loss of β-galactosidase activity that results from insertion of target DNA into the *lacZ′* gene is determined. Recall that the multiple-cloning site in pUC19 lies within the *lacZ′* gene (Fig. 2.5). An *E. coli* host is used that can synthesize the part of β-galactosidase (LacZω fragment) that combines with the product of the *lacZ′* gene (LacZα fragment) encoded on pUC19 to form a functional enzyme. When cells carrying recircularized pUC19 are grown in the presence of isopropyl-β-D-thiogalactopyranoside (IPTG),

which is an inducer of the *lac* operon, the protein product of the *lacI* gene (the LacI repressor) is prevented from binding to the promoter–operator region of the *lacZ′* gene, so the *lacZ′* gene in the plasmid is transcribed and translated (Fig. 2.8B). The LacZα fragment combines with a host LacZω fragment to form an active hybrid β-galactosidase. If the substrate 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal) is present in the medium, it is hydrolyzed by the hybrid β-galactosidase to form a blue product (5,5′-dibromo-4,4′-dichloro-indigo). Under these conditions, colonies containing recircularized pUC19 appear blue (Fig. 2.8A). In contrast, host cells that carry a plasmid-cloned DNA construct produce white colonies on the same medium. The reason for this is that target DNA inserted into a restriction endonuclease site within the multiple-cloning site usually disrupts the correct sequence of DNA codons (reading frame) of the *lacZ′* gene and prevents the production of a functional LacZα fragment, so no active hybrid β-galactosidase is produced (Fig. 2.8B). In the absence of β-galactosidase activity, the X-Gal in the medium is not converted to the blue compound, so these colonies remain white (Fig. 2.8A). The white (positive) colonies subsequently must be confirmed to carry a specific target DNA sequence.

A number of selection systems have been devised to identify cells carrying vectors that have been successfully inserted with target DNA. In addition to ampicillin, other antibiotics such as tetracycline, kanamycin, and streptomycin are used as selective agents for various cloning vectors. Some vectors carry a gene that encodes a toxin that kills the cell (Table 2.2). The toxin gene is under the control of a regulatable promoter, such as the promoter for the *lacZ* gene that is activated only when the inducer IPTG is supplied in the culture medium. Insertion of a target DNA fragment into the multiple-cloning site within the toxin gene prevents the production of a functional toxin protein in the presence of the inducer. Only cells that carry a vector with the target DNA survive under these conditions.

In addition to *E. coli*, other bacteria, such as *Bacillus subtilis*, often are host cells. For many applications, cloning vectors that function in *E. coli* may be provided with a second origin of replication that enables the plasmid to replicate in the alternative host cell. With these shuttle

Figure 2.8 **(A)** Strategy for selecting host cells that have been transformed with pUC19 carrying cloned target DNA. *E. coli* host cells transformed with the products of pUC19–target DNA ligation are selected on medium containing ampicillin, X-Gal, and IPTG. Nontransformed cells and cells transformed with circularized target DNA do not have a gene conferring resistance to ampicillin and therefore do not grow on medium containing ampicillin (recircularized target DNA also does not carry an origin of replication, and therefore, the plasmids are not propagated in host cells even in the absence of ampicillin). However, *E. coli* cells transformed with recircularized pUC19 or pUC19 carrying cloned target DNA are resistant to ampicillin and therefore form colonies on the selection medium. The two types of ampicillin-resistant transformants are differentiated by the production of functional β-galactosidase. **(B)** IPTG in the medium induces expression of the *lacZ′* gene by binding to the LacI repressor and preventing LacI from binding to the *lacO* operator sequence. This results in production of the β-galactosidase LacZα fragment in cells transformed with recircularized pUC19. The LacZα fragment combines with the LacZω fragment of β-galactosidase encoded in the host *E. coli* chromosome to form a functional hybrid β-galactosidase. β-Galactosidase cleaves X-Gal, producing a blue product, and the colonies on the plate appear blue. Insertion of target DNA into the multiple-cloning site of pUC19 alters the reading frame of the *lacZ′* gene, thereby preventing production of a functional β-galactosidase in cells transformed with this construct. Colonies that appear white on medium containing X-Gal and IPTG carry the cloned target gene **(A)**.


A
Mixture of pUC19-target DNA ligation products is transformed into competent E. coli host cells
Transformation mixture is plated on medium containing ampicillin, X-Gal, and IPTG
No plasmid Amp^s, no growth on ampicillin
Circularized target DNA Amp^s, no growth on ampicillin
pUC19 with cloned target DNA Amp^r, no β-galacto-sidase activity
Recircularized pUC19 Amp^r, β-galacto-sidase activity

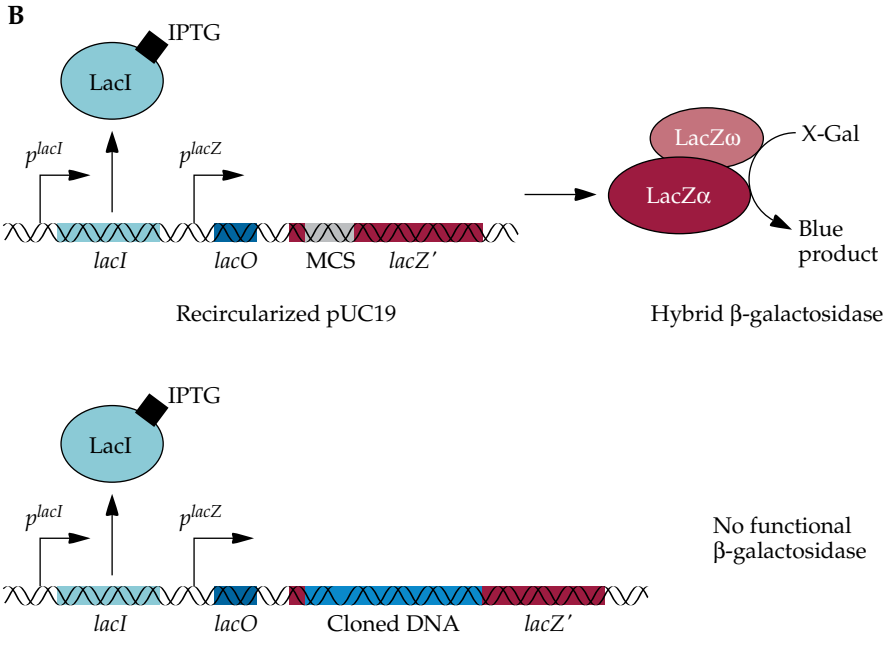

B
IPTG
LacI
p^lacI
p^lacZ
lacI
lacO
MCS
lacZ′
LacZω
LacZα
X-Gal
Blue product
Recircularized pUC19
Hybrid β-galactosidase
IPTG
LacI
p^lacI
p^lacZ
lacI
lacO
Cloned DNA
lacZ′
No functional β-galactosidase

Table 2.2 Some toxin genes used to select for successful insertion of target DNA into a vector

Toxin gene	Source	Mode of action
sacB	*Bacillus subtilis*	Encodes the enzyme levansucrase, which converts sucrose to levans that are toxic to Gram-negative bacteria
ccdB	*E. coli* F plasmid	Encodes a toxin that inhibits DNA gyrase activity and therefore DNA replication
barnase	*Bacillus amyloliquefaciens*	Encodes a ribonuclease that cleaves RNA
eco47IR	*E. coli*	Encodes an endonuclease that cleaves DNA at the sequence GGWCC (W = A or T) when it is not methylated
hsv-tk	Herpes simplex virus	Encodes thymidine kinase that converts the synthetic nucleotide analogue ganciclovir into a product that is toxic to insect cells

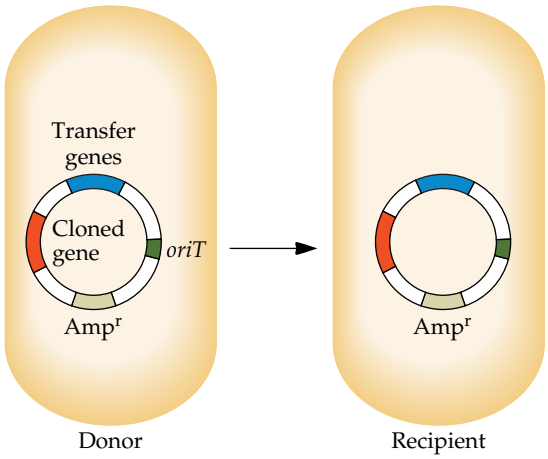
cloning vectors, the initial cloning steps are generally conducted using *E. coli* before the final construct is introduced into a different host cell. In addition, a number of plasmid vectors have been constructed with a single broad-host-range origin of DNA replication instead of a narrow-host-range origin of replication. These vectors can be used with a variety of microorganisms.

Broad-host-range vectors can be transferred among different bacterial hosts by exploiting a natural system for transmitting plasmids known as conjugation. There are two basic genetic requirements for transfer of a plasmid by conjugation: (i) a specific origin-of-transfer (*oriT*) sequence on the plasmid that is recognized by proteins that initiate plasmid transfer and (ii) several genes encoding the proteins that mediate plasmid transfer. The genes encoding the transfer proteins may be present on the transferred plasmid (Fig. 2.9A), in the genome of the plasmid donor cell, or supplied on a helper plasmid (Fig. 2.9B). Some of these proteins form a pilus that extends from the donor cell and, following contact with a recipient cell, retracts to bring the two cells into close contact. A specific endonuclease cleaves one of the two strands of the plasmid DNA at the *oriT*, and as the DNA is unwound, the displaced single-stranded DNA is transferred into the recipient cell through a conjugation pore made up of proteins encoded by the transfer genes. A complementary strand is synthesized in both the donor and recipient cells, resulting in a copy of the plasmid in both cells.

Cloning Eukaryotic Genes

Bacteria lack the molecular machinery to excise the introns (intervening, non-protein-coding sequences) from RNA that is transcribed from eukaryotic genes. Therefore, before a eukaryotic sequence is cloned for the purpose of producing the encoded protein in a bacterial host, the intron sequences must be removed. Functional eukaryotic mRNA does not contain introns because they have been removed by the eukaryotic cell's splicing machinery. Purified mRNA molecules are used as a starting point for cloning eukaryotic genes but must be converted to double-stranded DNA before they are inserted into a vector that provides bacterial sequences for transcription and translation.

A Biparental conjugation



B Triparental conjugation

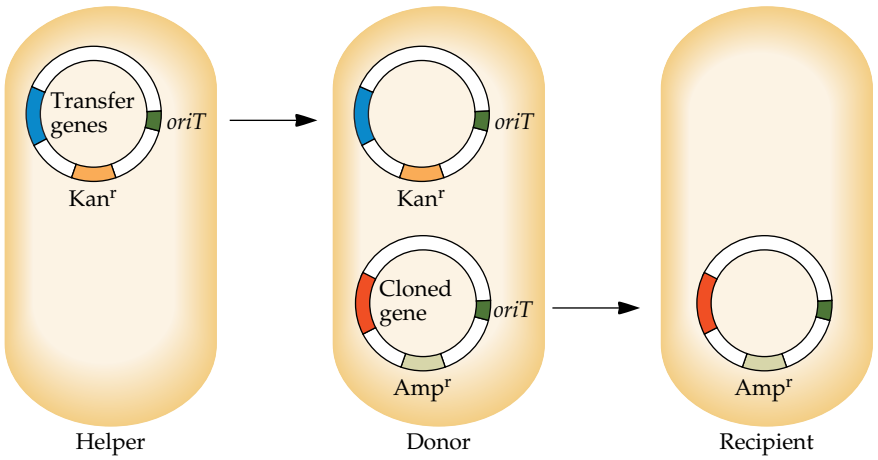


Figure 2.9 Plasmid transfer by biparental **(A)** or triparental **(B)** conjugation. A plasmid carrying a cloned gene, an origin of transfer (*oriT*), and transfer genes is transferred by biparental mating from a donor cell to a recipient cell **(A)**. Proteins encoded by the transfer genes mediate contact between donor and recipient cells, initiate plasmid transfer by nicking one of the DNA strands at the *oriT*, and form a pore through which the nicked strand is transferred from the donor cell to the recipient cell. In a cloning experiment, the donor is often a strain of *E. coli* that does not grow on minimal medium, allowing selection of recipient cells that grow on minimal medium. Acquisition of the plasmid by recipient cells is determined by resistance to an antibiotic, such as ampicillin in this example. If the plasmid carrying the cloned gene does not possess genes for plasmid transfer, these can be supplied by a helper cell **(B)**. In triparental mating, the helper plasmid is first transferred to the donor cell, where the proteins that mediate transfer of the plasmid carrying the cloned gene are expressed. Although the plasmid carrying the cloned gene does not possess transfer genes, it must have an *oriT* in order to be transferred. Neither the helper nor the donor cells grow on minimal medium, and therefore, the recipient cells can be selected on minimal medium containing an antibiotic such as ampicillin. To ensure that the helper plasmid was not transferred to the recipient cell, sensitivity to kanamycin is determined.

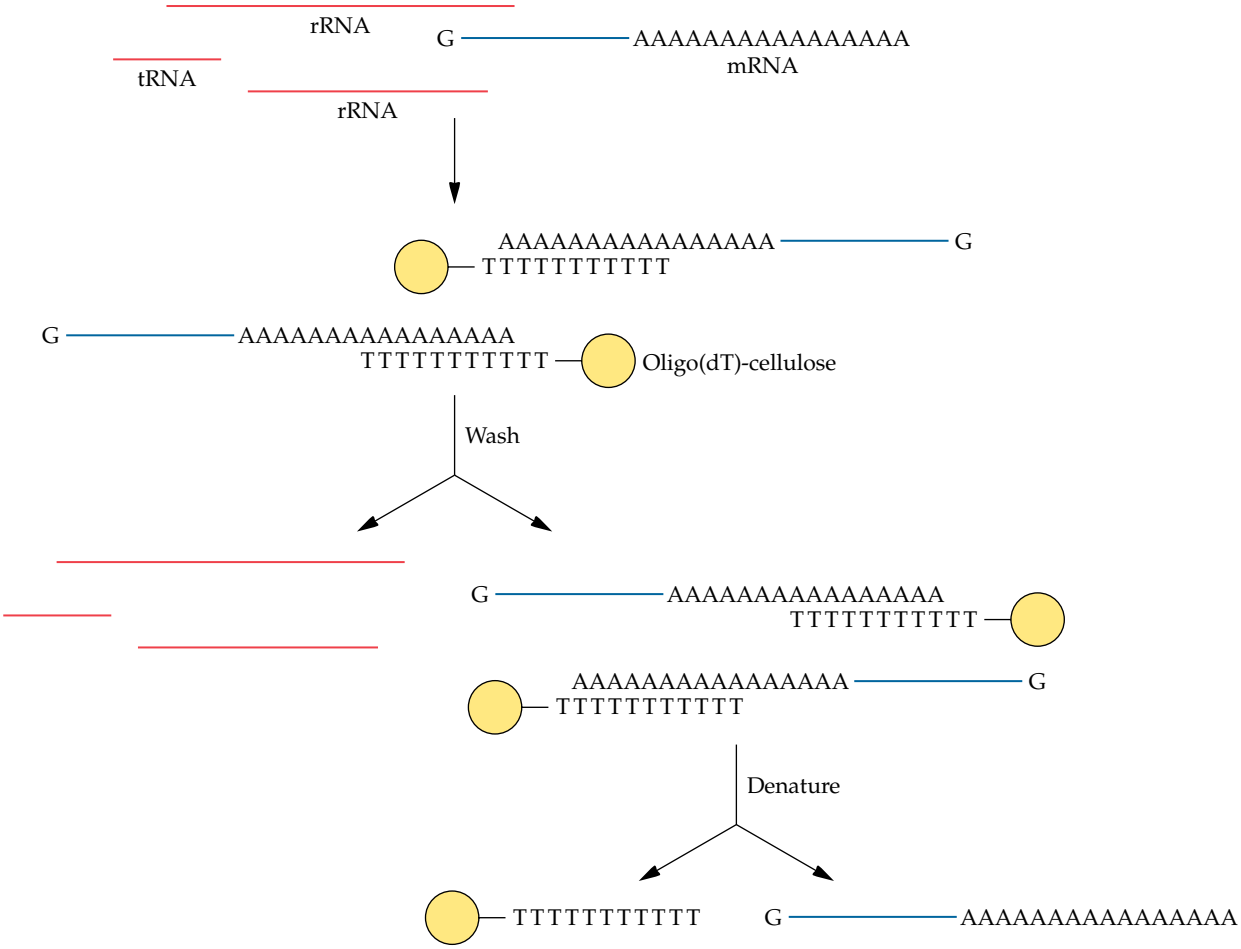


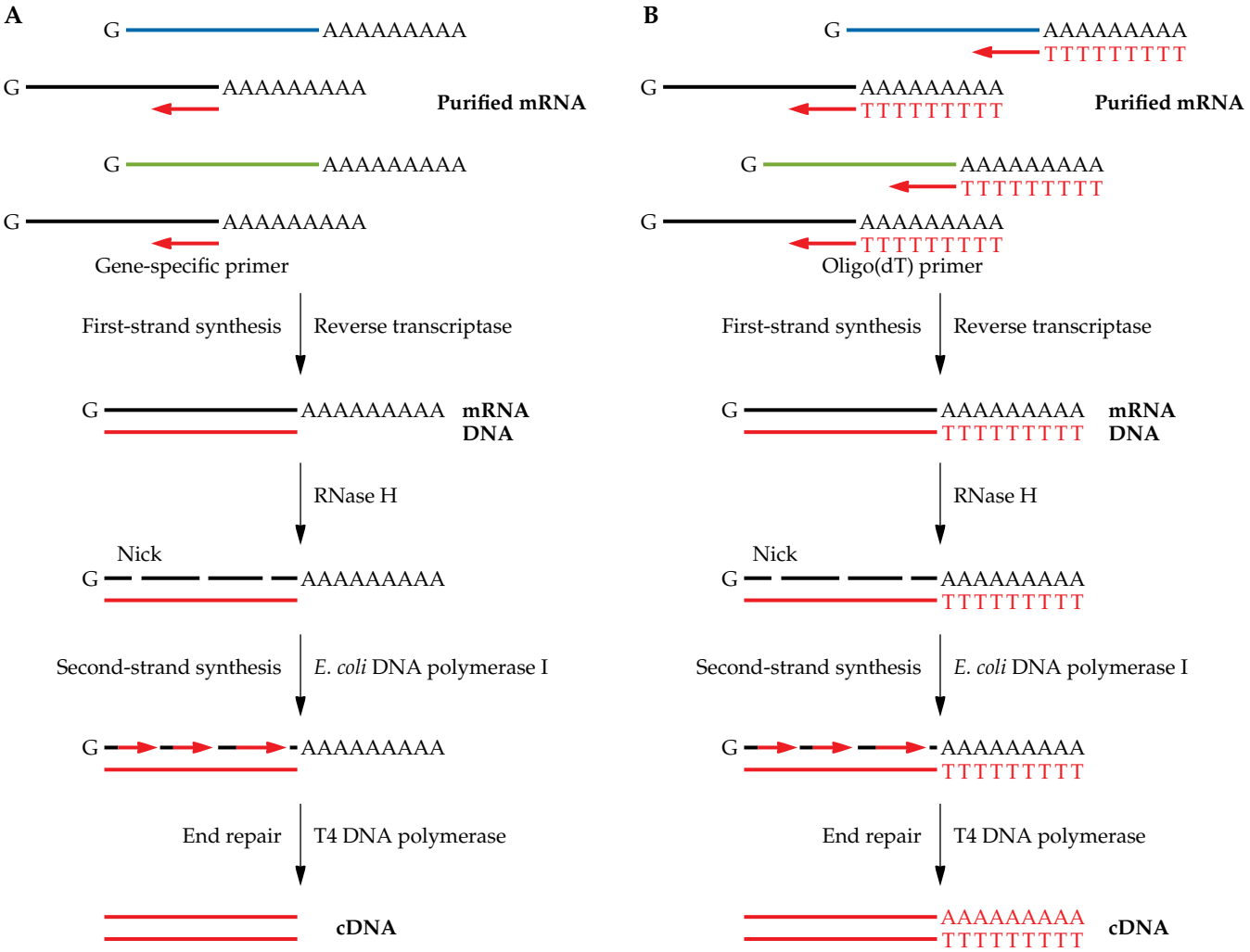
Figure 2.10 Schematic representation of oligo(dT)-cellulose separation of polyadenylated mRNA from total cellular RNA. G, 7-methylguanosine cap on the 5′ end of eukaryotic mRNA. Cellulose beads are indicated by yellow circles.

Purified mRNA can be obtained from eukaryotic cells by exploiting the tract of up to 200 adenine monophosphates (polyadenylic acid [poly(A)] tail) that are added to the 3′ ends of mRNAs before they are exported from the nucleus (Fig. 2.10). The poly(A) tail provides the means for separating the mRNA fraction of a tissue from the more abundant ribosomal RNA (rRNA) and transfer RNA (tRNA). Short chains of 15 thymidine monophosphates (oligodeoxythymidylic acid [oligo(dT)]) are attached to cellulose beads, and the oligo(dT)–cellulose beads are packed into a column. Total RNA extracted from eukaryotic cells or tissues is passed through the oligo(dT)–cellulose column, and the poly(A) tails of the mRNA molecules bind by base-pairing to the oligo(dT) chains. The tRNA and rRNA molecules, which lack poly(A) tails, pass through the column. The mRNA is removed (eluted) from the column by treatment with a low-salt buffer that breaks the A:T hydrogen bonds.

To convert mRNA to double-stranded DNA for cloning, the enzyme reverse transcriptase, encoded by certain RNA viruses (retroviruses), is used to catalyze the synthesis of complementary DNA (cDNA) from

an RNA template. If the sequence of the target mRNA is known, a short (~20 nucleotides), single-stranded DNA molecule known as an oligonucleotide primer that is complementary to a sequence at the 3′ end of the target mRNA is synthesized (Fig. 2.11A). The primer is added to a sample of purified mRNA that is extracted from eukaryotic cells known to produce the mRNA of interest. This sample of course contains all of the different mRNAs that are produced by the cell; however, the primer will specifically base-pair with its complementary sequence on the target mRNA. Not only is the primer important for targeting a specific mRNA,

Figure 2.11 Synthesis of double-stranded cDNA using gene-specific primers **(A)** or oligo(dT) primers **(B)**. A short oligonucleotide primer is added to a mixture of purified mRNA and anneals to a complementary sequence on the mRNA. Reverse transcriptase catalyzes the synthesis of a DNA strand from the primer using the mRNA as a template. To synthesize the second strand of DNA, the mRNA is nicked by RNase H, which creates initiation sites for *E. coli* DNA polymerase I. The 5′ exonuclease activity of DNA polymerase I removes RNA sequences that are encountered as DNA synthesis proceeds. The ends of the cDNA are blunted using T4 DNA polymerase prior to cloning.



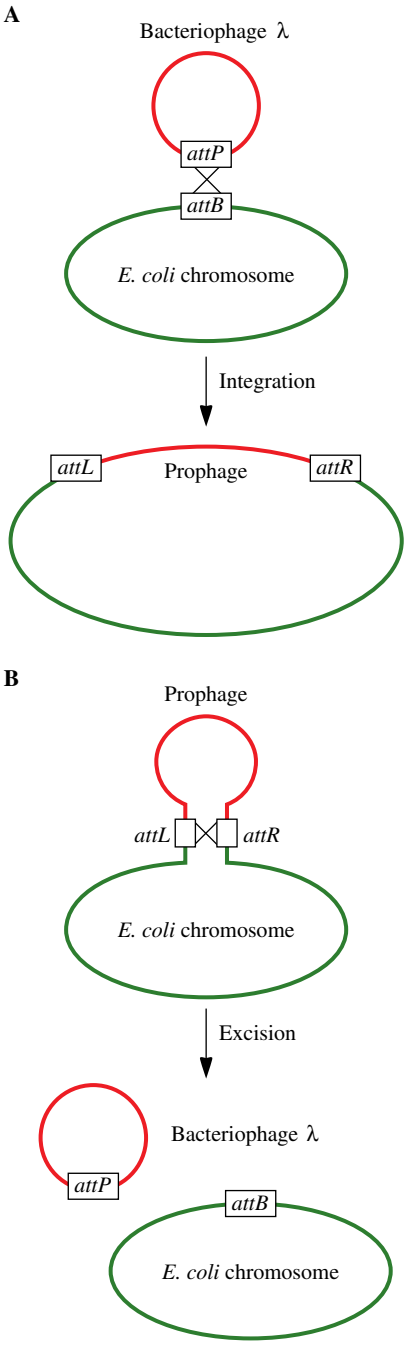


Figure 2.12 Integration **(A)** and excision **(B)** of bacteriophage λ DNA into and from the *E. coli* genome via recombination (represented by an X) between attachment (*att*) sites in the bacterial and bacteriophage DNA.

but also, it provides an available 3′ hydroxyl group to prime the synthesis of the first cDNA strand. In the presence of the four deoxyribonucleotides, reverse transcriptase incorporates a complementary nucleotide into the growing DNA strand as determined by the sequence of the template mRNA strand. To generate a double-stranded DNA molecule, the RNA:DNA (heteroduplex) molecules are treated with RNase H, which nicks the mRNA strands, thereby providing free 3′ hydroxyl groups for initiation of DNA synthesis by DNA polymerase I. As the synthesis of the second DNA strand progresses from the 3′ ends of the nicked mRNA fragments, the 5′ exonuclease activity of DNA polymerase I removes the ribonucleotides of the mRNA. After synthesis of the second DNA strand is completed, the ends of the cDNA molecules are blunted (end repaired and polished) with T4 DNA polymerase, which removes 3′ extensions and fills in from 3′ recessed ends. The double-stranded cDNA carrying only the exon sequences encoding the eukaryotic protein can be cloned directly into a suitable vector by blunt-end ligation. Alternatively, chemically synthesized short double-stranded DNA adaptors that contain a restriction endonuclease recognition sequence (which is not present in the cDNA) can be ligated to the ends of the cDNA molecules and then digested with the restriction endonuclease prior to insertion into a vector via sticky-end ligation.

When the sequence of the target mRNA intended for cloning is not known or when several target mRNAs in a single sample are of interest, cDNA can be generated from all of the mRNAs using an oligo(dT) primer rather than a gene-specific primer (Fig. 2.11B). The mixture of cDNAs, ideally representing all possible mRNA produced by the cell, is cloned into a vector to create a cDNA library that can be screened for the target sequence(s) (described below).

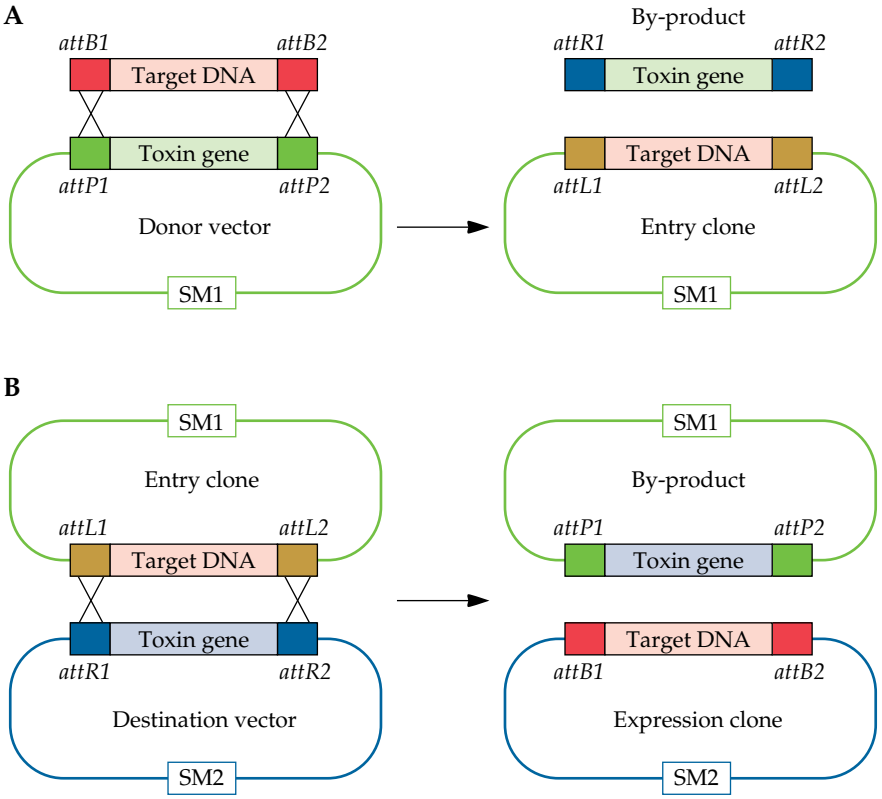
Recombinational Cloning

Recombinational cloning is a rapid and versatile system for cloning sequences without restriction endonuclease and ligation reactions. It is particularly useful when a large number of DNA fragments are to be cloned into one type of vector, for example, to introduce protein-coding sequences into an expression vector for the production and purification of thousands of different proteins in parallel to facilitate the creation of a proteomic microarray (described later in this chapter). One method, known as Gateway cloning technology, exploits the mechanism used by the double-stranded DNA bacteriophage λ to integrate its DNA into the host bacterial genome during infection. Bacteriophage λ integrates into the *E. coli* chromosome at a specific sequence (25 bp) in the bacterial genome known as the attachment bacterium (*attB*) site. The bacteriophage genome has a corresponding attachment phage (*attP*) sequence (243 bp) that can recombine with the bacterial *attB* sequence with the help of the bacteriophage λ recombination protein integrase and an *E. coli*-encoded protein called integration host factor (Fig. 2.12A). Recombination between the *attP* and *attB* sequences results in insertion of the phage genome into the bacterial genome to create a prophage with attachment sites *attL* (100 bp) and *attR* (168 bp) at the left and right ends of the integrated bacteriophage λ DNA, respectively. For subsequent excision of the bacteriophage λ DNA from the bacterial chromosome, recombination between the *attL* and *attR* sites is mediated by integration host factor, integrase, and the bacteriophage λ

enzyme, excisionase (Fig 2.12B). The recombination events occur at precise locations without either the loss or gain of nucleotides.

For recombinational cloning, a modified *attB* sequence is added to each end of the target DNA. The *attB* sequences are modified so that they will only recombine with specific *attP* sequences. For example, *attB1* recombines only with *attP1*, and *attB2* recombines with *attP2*. The target DNA with flanking *attB1* and *attB2* sequences is mixed with a vector (donor vector) that has *attP1* and *attP2* sites flanking a toxin gene that will be used to select for recombination following transformation into a host cell (Fig. 2.13A). Integrase and integration host factor are added to the mixture

Figure 2.13 Recombinational cloning. **(A)** Recombination (represented by an X) between attachment sites (*attB1* and *attB2*) flanking a target gene and *attP1* and *attP2* sites on either side of a toxin gene on a donor vector results in an entry clone where the target gene is flanked by *attL1* and *attL2* sites. The selectable marker (SM1) enables selection of cells transformed with an entry clone. The protein encoded by the toxin gene kills cells transformed with nonrecombined donor vectors. The origin of replication of the donor vector is not shown. **(B)** Recombination between the entry clone with flanking *attL1* and *attL2* sites and a destination vector with *attR1* and *attR2* sites results in an expression clone with *attB1* and *attB2* sites flanking the target gene. The selectable marker (SM2) enables selection of transformed cells with an expression clone. The second plasmid, designated as a by-product, has the toxin gene flanked by *attP1* and *attP2* sites. Cells with an intact destination vector that did not undergo recombination or that retain the by-product plasmid are killed by the toxin. Transformed cells with an entry clone, which lacks the SM2 selectable marker, are selected against. The origins of replication and the sequences for expression of the target gene are not shown.



of DNA molecules to catalyze *in vitro* recombination between the *attB1* and *attP1* sites and between the *attB2* and *attP2* sites. As a consequence of the two recombination events, the toxin gene sequence between the *attP1* and *attP2* sites on the donor vector is replaced by the target gene. The recombination events create new attachment sites flanking the target gene sequence (designated *attL1* and *attL2*), and the plasmid with the *attL1*–target gene–*attL2* sequence is referred to as an entry clone. The mixture of original and recombinant DNA molecules is transformed into *E. coli*, and cells that are transformed with donor vectors that have not undergone recombination retain the toxin gene and therefore do not survive (referred to as negative selection). Host cells carrying the entry clone are positively selected by the presence of a selectable marker on the vector.

The advantage of this procedure is the ability to easily transfer the target gene to a variety of vectors that have been developed for different purposes. For example, to produce high levels of the protein encoded on the cloned gene, the target DNA can be transferred to a destination vector that carries a promoter and other expression signals. An entry clone is mixed with a destination vector that has *attR1* and *attR2* sites flanking a toxin gene (Fig. 2.13B). In the presence of integration host factor, integrase, and bacteriophage λ excisionase, the *attL1* and *attL2* sites on the entry clone recombine with the *attR1* and *attR2* sites, respectively, on the destination vector. This results in the replacement of the toxin gene on the destination vector with the target gene from the entry clone, and the resultant plasmid is designated an expression clone. The reaction mixture is transformed into *E. coli*, and a selectable marker is used to isolate transformed cells that carry an expression clone. Cells that carry an intact destination vector or the exchanged entry plasmid (known as a by-product plasmid) will not survive, because these carry the toxin gene. Destination vectors are available for maintenance and expression of the target gene in various host cells such as *E. coli* and yeast, insect, and mammalian cells.

Genomic Libraries

A genomic library is a collection of DNA fragments, each cloned into a vector, that represents the entire genomic DNA, or cDNA derived from the total mRNA, in a sample. For example, the genomic library may contain fragments of the entire genome extracted from cells in a pure culture of bacteria or from tissue from a plant or animal. A genomic library can also contain the genomes of all of the organisms present in a complex sample such as from the microbial community on human tissue. Such libraries are known as metagenomic libraries. Whole-genome libraries may be used to identify sequences that contain specific nucleotides, encode particular functions, or interact with other molecules.

To create a genomic library, the DNA extracted from the cells (cell cultures or tissues) of a source organism (or a community of organisms for a metagenomic library) is first digested with a restriction endonuclease. Often, a restriction endonuclease that recognizes a sequence of four nucleotides, such as Sau3AI, is used. Although four-cutters will theoretically cleave the DNA approximately once every 256 bp, the reaction conditions are set to give a partial, not a complete, digestion to generate fragments of all possible sizes (Fig. 2.14). This is achieved by using either a low concentration of restriction endonuclease or shortened incubation

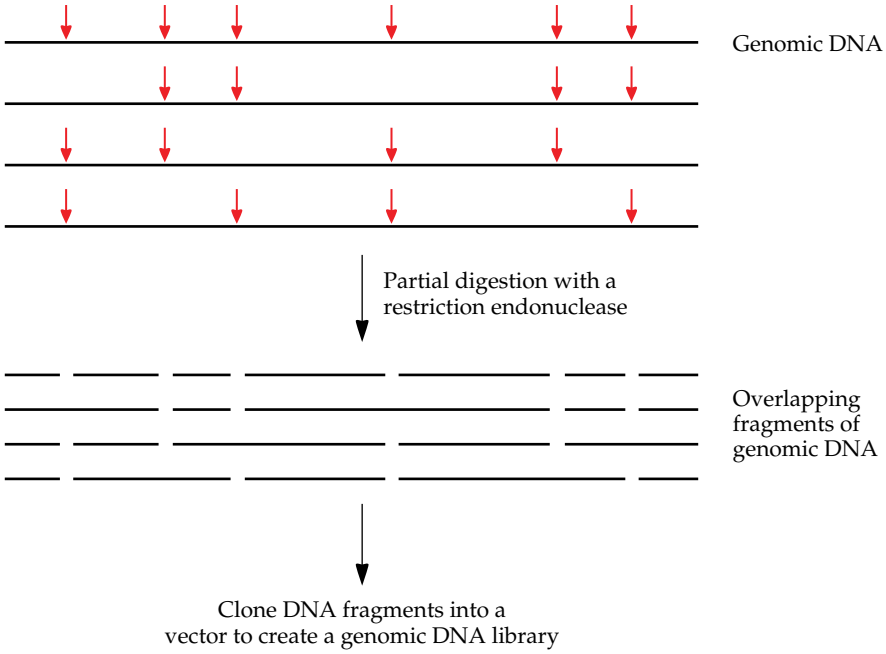


Figure 2.14 Construction of a genomic DNA library. Genomic DNA extracted from cells or tissues is partially digested with a restriction endonuclease. Conditions are set so that the enzyme does not cleave at all possible sites. This generates overlapping DNA fragments of various lengths that are cloned into a vector.

times, and usually some optimization of these parameters is required to determine the conditions that yield fragments of suitable sizes. The range of fragment sizes depends on the goal of the experiment. For example, for genome sequencing, large (100- to 200-kb) fragments are often desirable. To identify genes that encode a particular enzymatic function, that is, genes that are expressed to produce proteins in the size range of an average protein, smaller (~3- to 40-kb) fragments are cloned.

The number of clones in a genomic library depends on the size of the genome of the organism, the average size of the insert in the vector, and the average number of times each sequence is represented in the library (sequence coverage). To ensure that the entire genome, or most of it, is contained within the clones of a library, the sum of the inserted DNA in all of the clones of the library should be at least three times the amount of DNA in the genome. For example, the size of the *E. coli* genome is approximately 4.6×10^6 bp; if inserts of an average size of 1,000 bp are desired, then 13,800 clones are required for 3-fold coverage, i.e., $3[(4.6 \times 10^6)/10^3]$. For the human genome, which contains 3.3×10^9 bp, about 80,000 clones with an average insert size of 150,000 bp are required for 4-fold coverage, i.e., $4[(3.3 \times 10^9)/(15 \times 10^4)]$. Statistically, the number of clones required for a comprehensive genomic library can be estimated from the relationship $N = \ln(1 - P)/\ln(1 - f)$, where N is the number of clones, P is the probability of finding a specific gene, and f is the ratio of the length of the average insert to the size of the entire genome. On this basis, about 700,000 clones are required for a 99% chance of discovering a particular sequence in a human genomic library with an average insert size of 20 kb.

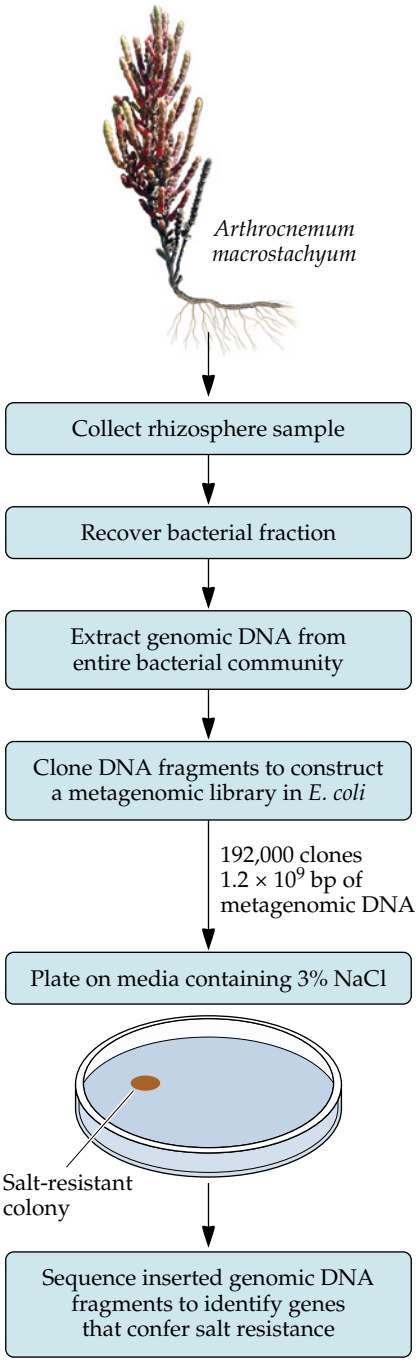


Figure 2.15 Isolation of genes involved in salt resistance from a metagenomic DNA library of halophilic bacteria that thrive in high-salt environments. Genomic DNA extracted from bacteria found in the soil around the roots of a plant (rhizosphere) growing in hypersaline soil was fragmented and cloned into a plasmid to generate a metagenomic library. *E. coli* host cells carrying the cloned DNA were plated on a solid medium containing 3% sodium chloride. The *E. coli* strain normally does not grow on 3% sodium chloride. The cloned DNA fragments from colonies that developed on the plates were sequenced to identify genes that confer resistance to salt. Data from Mirete et al., *Front. Microbiol*. **6:**1121, 2015.

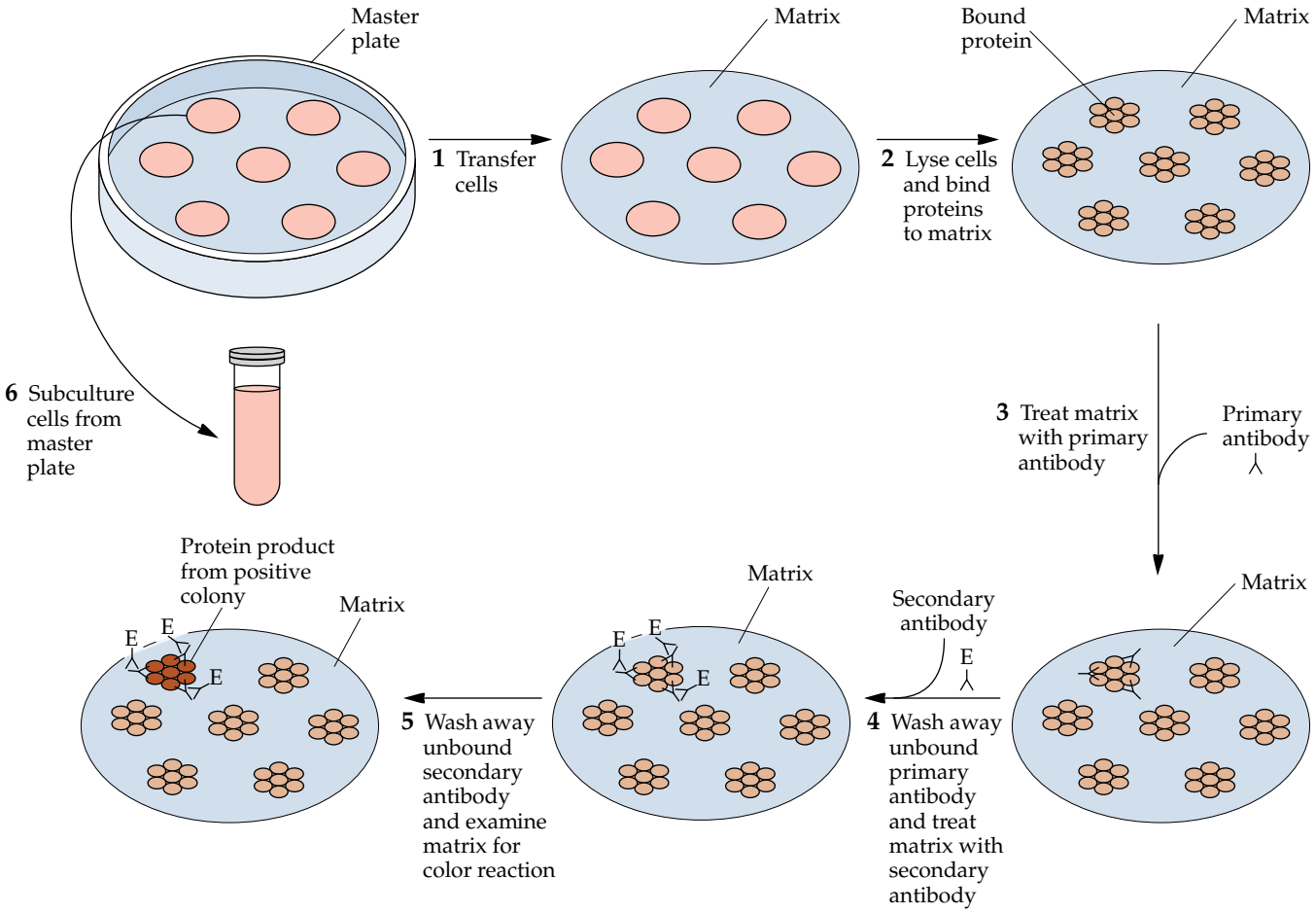
Several strategies can be used to identify target DNA in a genomic library. Genomic or metagenomic libraries can be screened to identify members of the library that carry a gene encoding a particular protein function. In functional complementation, the host cell does not have the protein activity of interest, in some cases because the host gene encoding the protein carries a mutation that abolishes the activity of the protein. A DNA library is constructed that carries fragments of genomic DNA from an organism that has the desired protein activity. Host cells with the genetic deficiency are transformed with plasmids of the DNA library, and transformed cells that have restored normal protein function are selected. The genomic DNA that is used to prepare the library can be from a variety of donor organisms, such as the wild-type strain of the host bacterium that carries a functional copy of the gene encoding the protein, a different organism that can be either another prokaryote or a eukaryote (the latter requires preparation of a cDNA library), or uncultured organisms that are present in an environmental sample.

Many genes encoding enzymes that catalyze specific reactions have been isolated from a variety of organisms by plating the cells of a genomic library on medium supplemented with a specific substrate. For efficient screening of thousands of clones, colonies that carry a cloned gene encoding a functional catabolic enzyme must be readily identifiable, often by production of a colored product or a zone of substrate clearance around the colony. Alternatively, genes may be identified that confer desired properties on host cells. In one example, genes encoding proteins that confer resistance to salt were isolated from a metagenomic library of DNA from bacteria present in a hypersaline environment (Fig. 2.15). DNA was extracted from bacteria collected from the saline soil around the roots of the halophyte *Arthrocnemum macrostachyum*. A library of 192,000 clones with an average insert size of 3 kb representing 1.2×10^9 bp of metagenomic DNA was constructed in an osmosensitive strain of *E. coli*. The library was screened by plating the transformed *E. coli* cells on medium containing 3% sodium chloride, which is normally lethal for the host *E. coli* strain. Eight different genes that conferred salt resistance were identified. Based on sequence similarity to known genes, these are predicted to encode enzymes involved in nucleic acid structure and repair, an outer membrane protein, a glycerol permease, a proton pump, and two hypothetical proteins whose functions are not known.

The presence of particular proteins produced by a genomic library can also be detected using an immunological assay. Rather than screening for the function of a protein, the library is screened using an antibody that

specifically binds to the protein encoded by a target gene. The colonies are arrayed on a solid medium, transferred to a matrix, and then lysed to release the cellular proteins (Fig. 2.16). An antibody (primary antibody) is applied that specifically binds to the target protein on the matrix. Following the interaction of the primary antibody with the target protein, any unbound antibody is washed away, and the matrix is treated

Figure 2.16 Screening of a genomic DNA library using an immunological assay. Transformed cells are plated onto solid agar medium under conditions that permit transformed but not nontransformed cells to grow. (1) From the discrete colonies formed on this master plate, a sample from each colony is transferred to a solid matrix such as a nylon membrane. (2) The cells on the matrix are lysed, and their proteins are bound to the matrix. (3) The matrix is treated with a primary antibody that binds only to the target protein. (4) Unbound primary antibody is washed away, and the matrix is treated with a secondary antibody that binds only to the primary antibody. (5) Any unbound secondary antibody is washed away, and a colorimetric (or chemiluminescent) reaction is carried out. The reaction can occur only if the secondary antibody, which is attached to an enzyme (E) that performs the reaction, is present. (6) A colony on the master plate that corresponds to a positive response on the matrix is identified. Cells from the positive colony on the master plate are subcultured because they may carry the plasmid–insert DNA construct that encodes the protein that binds the primary antibody.



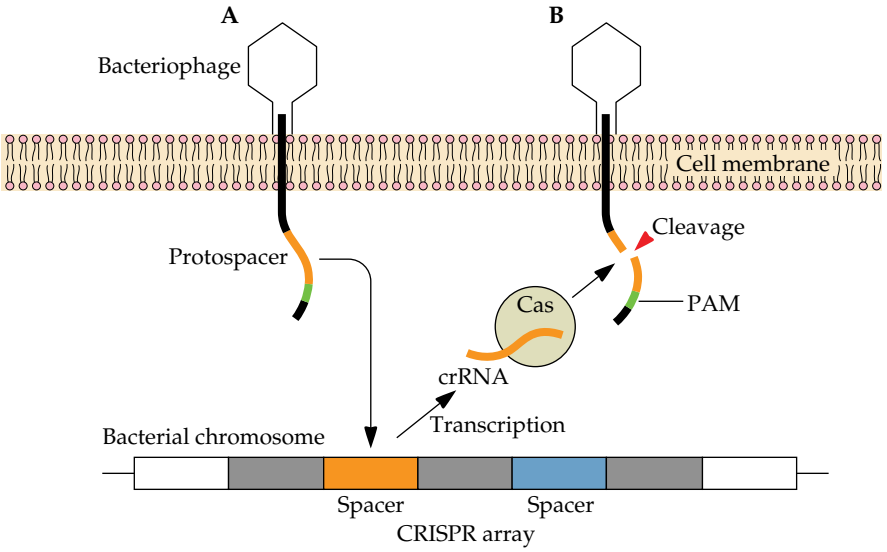
with a second antibody (secondary antibody) that is specific for the primary antibody. The secondary antibody is attached to an enzyme, such as alkaline phosphatase, that converts a colorless substrate to a colored or light-emitting (chemiluminescent) product that can readily identify positive interactions.

Genome Engineering Using CRISPR Technology

Researchers have designed strategies to insert, replace, or disrupt sequences at targeted sites in intact genomes *in vivo*. One method is based on a prokaryotic system that protects bacteria against invasion by foreign DNA such as bacteriophage genomic DNA and plasmids. This is a type of bacterial adaptive immune system that consists of genomic clustered regularly interspaced short palindromic repeats (CRISPR) containing fragments of foreign DNA molecules that the bacterium was previously exposed to and CRISPR-associated (Cas) proteins, including an endonuclease that cleaves homologous foreign DNA upon subsequent exposures. The CRISPR-Cas system has been adapted to introduce or replace genes in the genomes of a variety of organisms, both prokaryotes and eukaryotes, and also to edit genomes, that is, remove or alter targeted nucleotides (see chapters 6, 11, and 12).

In the natural bacterial CRISPR-Cas systems, short sequences ("protospacers") from an invading DNA molecule are incorporated as "spacers" between repeat sequences in the CRISPR locus of the bacterial genome (Fig. 2.17A). Thus, the CRISPR locus contains an array of spacers separated by repeat sequences that are a record of past foreign DNA invasions from which the bacterium survived. When the bacterium is subsequently

Figure 2.17 Bacterial CRISPR-Cas system for protection against invading bacteriophage. **(A)** Fragments of bacteriophage DNA (protospacer) are incorporated into the host bacterial genome as spacers between repeat sequences (gray) in the CRISPR array. **(B)** On subsequent invasion, the spacer DNA is transcribed to produce CRISPR RNA (crRNA) that guides an endonuclease (Cas) to a sequence in the invading DNA that is homologous to the spacer sequence and is adjacent to a protospacer-adjacent motif (PAM). The viral genome is cleaved.



Discovery of the CRISPR-Cas System for Genome Editing

milestone

When Japanese Ph.D. student Yoshizumi Ishino looked at the sequence of the *E. coli* gene he had just determined, he noticed something unusual, a cluster of five nearly identical sequences of 29 nucleotides separated by short spacer regions (Ishino et al., *J. Bacteriol*. **169:**5429–5433, 1987). The significance of this was not apparent, and Ishino did not investigate it further. A few years later, Ph.D. student Francisco Mojica at the University of Alicante in Spain noted a similar phenomenon in the genomes of two halophilic archaea that contained 14 regularly spaced repeat sequences (Mojica et al., *Mol. Microbiol*. **17:**85–93, 1995), which he later named "clustered regularly interspaced short palindromic repeats" or CRISPR for short. Soon after, repeat sequences separated by spacers were found in the genomes of several species of archaea and bacteria, as well as adjacent CRISPR-associated (*cas*) genes with nuclease- and helicase-like motifs (Jansen et al., *Mol. Microbiol*. **43:**1565–1575, 2002). Mojica proposed that the spacers were viral sequences that constituted a bacterial immune system, which was confirmed by French food scientists Rodolphe Barrangou and Philippe Horvath (Barrangou et al., *Science* **315:**1709–1712, 2007). Initially, it was speculated that RNA interference (see chapter 6) was the mechanism by which the CRISPR system defended bacteria from invading viruses (Makarova et al., *Biol. Direct* **1:**7, 2006). However, Luciano Marraffini and Erik Sontheimer at Northwestern University in Chicago showed that while CRISPR "interference" with foreign DNA was mediated by a small RNA transcribed from a spacer, introduced DNA containing a sequence identical to the spacer, and not the RNA transcribed from it, was the target site for the small RNA (Marraffini and Sontheimer, *Science* **322:**1843–1845, 2008). Perceptively, they suggested that "the ability to direct the specific addressable destruction of DNA that contains any given 24- to 48-nucleotide target sequence could have considerable functional utility."

Jennifer Doudna and her lab at the University of California at Berkeley in collaboration with Emmanuelle Charpentier and her lab at the University of Vienna were the first to work out the mechanism by which the CRISPR system destroyed target DNA. Doudna's lab began by applying their expertise in protein biochemistry to purify and resolve the structure of proteins encoded by the *cas* genes to determine their function (e.g., Wiedenheft et al., *Structure* **17:**904–912, 2009). They demonstrated that a Cas endonuclease specifically bound to and processed CRISPR RNA (crRNA), which is transcribed as a longer pre-crRNA from the CRISPR locus (Haurwitz et al., *Science* **329:**1355–1358, 2010). The Cas endonuclease cleavage site in invading foreign DNA is determined by base-pairing between the Cas-bound crRNA and a complementary DNA sequence (protospacer) only when the target site is near a short protospacer-adjacent motif (Sternberg et al., *Nature* **507:**62–67, 2014). During this time, Charpentier's group discovered (by deleting it) that *trans*-activating CRISPR RNA (tracrRNA), encoded in the CRISPR array, was essential for processing of crRNA by some Cas endonucleases, such as Cas9 from *Streptococcus pyogenes* (Deltcheva et al., *Nature* **471:**602–607, 2011). By joining forces, Doudna's and Charpentier's research groups showed that three molecules are necessary and sufficient for Cas9-mediated cleavage of DNA *in vitro*: (i) a crRNA, processed and oriented precisely by (ii) tracrRNA, guides (iii) the Cas9 endonuclease to generate a double-stranded break in purified DNA at a targeted site. The crRNA and tracrRNA were trimmed to essential nucleotides and fused together to create a single-guide RNA to facilitate practical applications of the system. The system could be programmed via the sequence of the single-guide RNA to cleave any DNA molecule at a specific sequence. In their seminal article (Jinek et al., *Science* **337:**816, 2012), the authors stated that the CRISPR-Cas system offers "considerable potential for gene-targeting and genome-editing applications." Doudna and Charpentier received the Nobel Prize in Chemistry in 2020 for the discovery of this new, simple, and powerful method for genome editing. Some scientists have described the CRISPR-Cas system as one of the most important discoveries in the history of biology.

invaded by a virus or plasmid whose DNA contains a sequence that is homologous to a spacer sequence, the spacer DNA is transcribed, producing a CRISPR RNA (crRNA) molecule that binds to and guides the Cas endonuclease complex to the target sequence in the invading DNA, which is cleaved (Fig. 2.17B). Recognition of the target sequence on the invading DNA requires that it is adjacent to a short specific sequence known as a protospacer-adjacent motif (PAM). For example, the *Streptococcus pyogenes* endonuclease Cas9 recognizes a target sequence that is complementary to

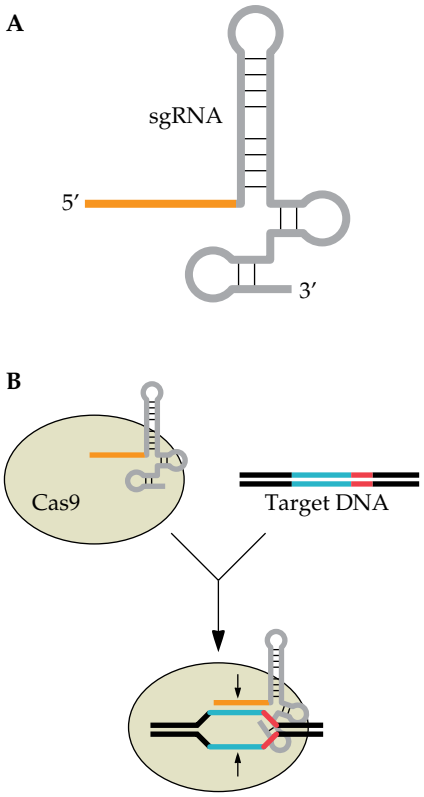


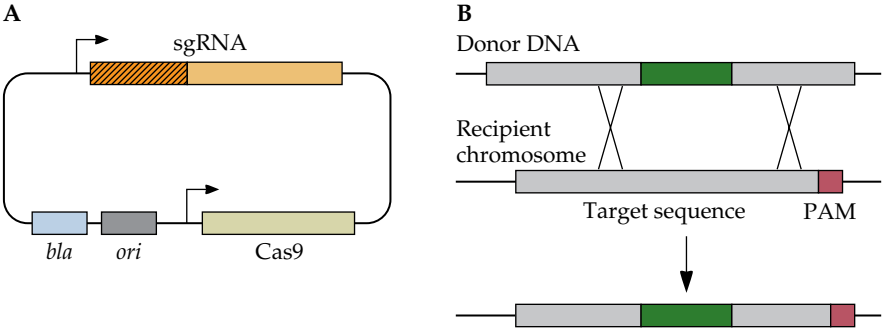
Figure 2.18 CRISPR-Cas system for genome editing. **(A)** An 80- to 100-nucleotide-long single guide RNA (sgRNA) is constructed that contains a 20-nucleotide guide sequence (orange) that is complementary to the target site. The secondary structure, stabilized by intramolecular base pairing between regions of the fused crRNA and tracrRNA sequences, is required for binding to the Cas9 endonuclease. **(B)** The sgRNA guides Cas9 to the target sequence (blue) in the genome. Target recognition requires an adjacent PAM sequence (red) NGG and complementarity between the guide sequence and the target sequence. Cas9 makes a double-stranded break in the target DNA (arrows) which is repaired by homologous recombination or nonhomologous end joining. The repair systems generate deletions and insertions at the target site.

a crRNA only if it is immediately upstream of the motif NGG (where N is any nucleotide). PAMs are also important for selection of protospacers during spacer acquisition. The PAM requirement prevents cleavage of the bacterium's own genome at sequences that are complementary to the crRNA, including the site in the CRISPR array from which the crRNA was transcribed, which lacks a PAM.

Because of its relative simplicity compared to systems in other bacteria, the CRISPR-Cas system from *S. pyogenes* has been adapted for use as a genome engineering tool. In the natural *S. pyogenes* system, two RNA molecules, crRNA and transactivating crRNA (tracrRNA), form a crRNA:tracrRNA hybrid that directs the Cas9 endonuclease to the target site. For ease of use in genome engineering, the two RNAs are combined into a single guide RNA (sgRNA) that is 80 to 100 nucleotides long. The sgRNA is designed to include a 20-nucleotide sequence that is complementary to the target site (which is located adjacent to a PAM) and the fused crRNA:tracrRNA sequence that forms a stem-loop structure involved in endonuclease binding (Fig. 2.18A). Following binding to the target sequence, the endonuclease makes a double-stranded break in the target DNA (Fig. 2.18B). This damage activates the cellular systems for DNA repair either by homologous recombination, in which DNA sequences with sufficient similarity are exchanged, or by nonhomologous end joining, in which sequences are deleted or inserted. The repair systems can be harnessed to disrupt, insert, or replace a DNA sequence at a targeted site.

Insertion of a DNA sequence (donor sequence) into a target site requires introduction of the sequences for the sgRNA, the Cas9 endonuclease, and the donor DNA into a recipient cell. The sgRNA and Cas9 coding sequences may be introduced on a vector (Fig. 2.19A), or the sgRNA and Cas9 mRNA

Figure 2.19 Vector for production of sgRNA and Cas9 in host cells **(A)**. The gene encoding sgRNA contains a 20-nucleotide sequence (hatched region) that is complementary to the target site in the host genome. Promoters (arrows) for the sgRNA and Cas9 genes, and codon usage for Cas9, must be suitable for expression in host cells. An origin of replication (*ori*) and a selectable marker (e.g., *bla* encoding β-lactamase, which confers resistance to the antibiotic ampicillin) are included for initial vector construction in *E. coli*. The vector and donor DNA are introduced into a recipient cell. Following expression, the sgRNA guides the Cas9 endonuclease to the target sequence in the recipient cell chromosome, and the endonuclease makes a double-stranded break in the target DNA. **(B)** The donor DNA sequence (green) is flanked by regions that are homologous to the target site (gray) for insertion by homologous recombination. Activation of recombinases that mediate DNA repair results in recombination between homologous sequences on the vector and in the recipient chromosome and, thereby, insertion of the donor DNA into the genome at the target site.



may be directly injected along with the donor DNA. When the genes are introduced on a vector, the promoters that drive expression of the sgRNA and endonuclease, and the coding sequence (e.g., codon usage) for the endonuclease, are optimized for expression in the chosen host. The donor DNA sequence is flanked by sequences that are homologous to the target genomic site for insertion by homologous recombination (Fig. 2.19B). The vector is introduced into the recipient cell, and following expression of the sgRNA and the endonuclease, the recipient cell genome is cleaved at the target site. Activation of recombinases that mediate DNA repair results in recombination between homologous sequences on the vector and in the recipient genome and, thereby, insertion of the donor DNA into the genome at the target site (Fig. 2.19B).

Polymerase Chain Reaction

The polymerase chain reaction (PCR) is a simple, efficient procedure for synthesizing large quantities of a specific DNA sequence *in vitro*. The reaction exploits the mechanism used by living cells to accurately replicate a DNA template. PCR can be used to produce millions of copies from a single template molecule and to detect a specific sequence in a complex mixture of DNA.

Amplification of DNA by PCR

The essential components for PCR amplification are (i) a template sequence in a DNA sample that is targeted for amplification and is 100 to 3,000 bp in length (larger regions can also be amplified, but with reduced efficiency), (ii) two synthetic oligonucleotide primers (~20 nucleotides each) that are complementary to regions on opposite strands that flank the target DNA sequence and that, after annealing to the sample DNA, have their 3′ hydroxyl ends oriented toward each other, (iii) a thermostable DNA polymerase that remains active after repeated heating to 95°C or higher and copies the DNA template with high fidelity, (iv) the four deoxyribonucleotides, and (v) a reaction buffer that provides optimal pH and osmotic conditions and cofactors (e.g., magnesium) required for DNA polymerase activity.

Replication of a specific DNA sequence by PCR requires three successive steps as outlined below. Amplification is achieved by repeating the three-step cycle 25 to 40 times. All steps in a PCR cycle are carried out in an automated block heater that is programmed to change temperatures after a specified period of time.

1. **Denaturation.** The first step in a PCR is the thermal denaturation of the double-stranded DNA template to separate the strands. This is achieved by raising the temperature of a reaction mixture to 95°C. The reaction mixture comprises the sample DNA that contains the target DNA to be amplified, a vast molar excess of the two oligonucleotide primers, a thermostable DNA polymerase (e.g., *Taq* DNA polymerase, isolated from the bacterium *Thermus aquaticus*), four deoxyribonucleotides, and the reaction buffer.
2. **Annealing.** In the second step, the primers base-pair, or anneal, with their complementary sequences in the DNA template. The temperature at which this step of the reaction is performed is

Specific Enzymatic Amplification of DNA *In Vitro*: the Polymerase Chain Reaction

milestone

PCR, which is the invention of Kary Mullis (U.S. patent 4,683,202), has had a tremendous impact on many research areas, including molecular biotechnology. The power of the method is in its simplicity, sensitivity, and specificity. It utilizes a mechanism similar to that used by our cells to accurately replicate a DNA template, it can detect and produce millions of copies from a single template molecule in a few hours, and, under the appropriate conditions, it can amplify a specific sequence in a complex mixture of DNA molecules even when other similar sequences are present.

PCR was a unique idea that did not replace any existing technology. In the early 1980s, Kary Mullis was trying to solve the problem of using synthetic oligonucleotides to detect single-nucleotide mutations in sequences that were present in low concentrations. He needed a method to increase the concentration of the target sequence. He reasoned that if he mixed heat-denatured DNA with two oligonucleotides that bound to opposite strands of the DNA at an arbitrary distance from each other and added some DNA polymerase and deoxynucleoside triphosphates, the polymerase would add the deoxynucleoside triphosphates to the hybridized oligonucleotides. The reaction did not yield the expected products. Mullis then heated the reaction products to separate the extended oligonucleotides from the template DNA and then repeated the process with fresh polymerase, hypothesizing that after each cycle, the number of molecules carrying the specific sequence between the primers would double. Despite the skepticism of his colleagues, Mullis proved that his reasoning was correct, albeit the hard way. By manually cycling the reaction through temperatures required to denature the DNA and anneal and extend the oligonucleotides, each time adding a fresh aliquot of a DNA polymerase isolated from *E. coli*, he was able to synthesize unprecedented amounts of target DNA (Mullis et al., *Cold Spring Harbor Symp. Quant. Biol*. **51:**263–273, 1986). Thermostable DNA polymerases that obviate the need to add fresh polymerase after each denaturation step and automated cycling have since made PCR a routine and indispensable laboratory procedure.

The capability of generating large amounts of DNA by amplification from segments of cloned or genomic DNA has facilitated the cloning of DNA versions of rare mRNA molecules, screening of gene libraries, diagnostic testing for gene mutations, sequencing of genomes, and a myriad of other applications. In fact, the first study using PCR described a diagnostic test for sickle-cell anemia (Saiki et al., Science **230:**1350–1354, 1985). Mullis received the Nobel Prize in Chemistry for his work on PCR in 1993.

determined by the nucleotide sequence of the primer that forms hydrogen bonds with complementary nucleotides in the target DNA. Typical annealing temperatures are in the range of 45 to 68°C, although optimization is often required to achieve the desired outcome, that is, a product consisting of fragments of target DNA sequence only.

3. **Extension.** In the third step, the temperature is raised to ~70°C, which is optimal for the catalytic activity of *Taq* DNA polymerase. DNA synthesis is initiated at the 3′ hydroxyl end of each annealed primer, and nucleotides are added to extend the complementary strand using the sample DNA as a template.

To understand how the PCR protocol succeeds in amplifying a discrete segment of DNA, it is important to keep in mind the location of each primer annealing site and its complementary sequence within the strands that are synthesized during each cycle. During the extension phase of the first cycle, the newly synthesized DNA from each primer is extended beyond the endpoint of the sequence that is complementary to the second primer. These new strands form "long templates" that are used in the second cycle (Fig. 2.20).

During the second cycle, the original sample DNA strands and the new strands synthesized in the first cycle (long templates) are denatured and

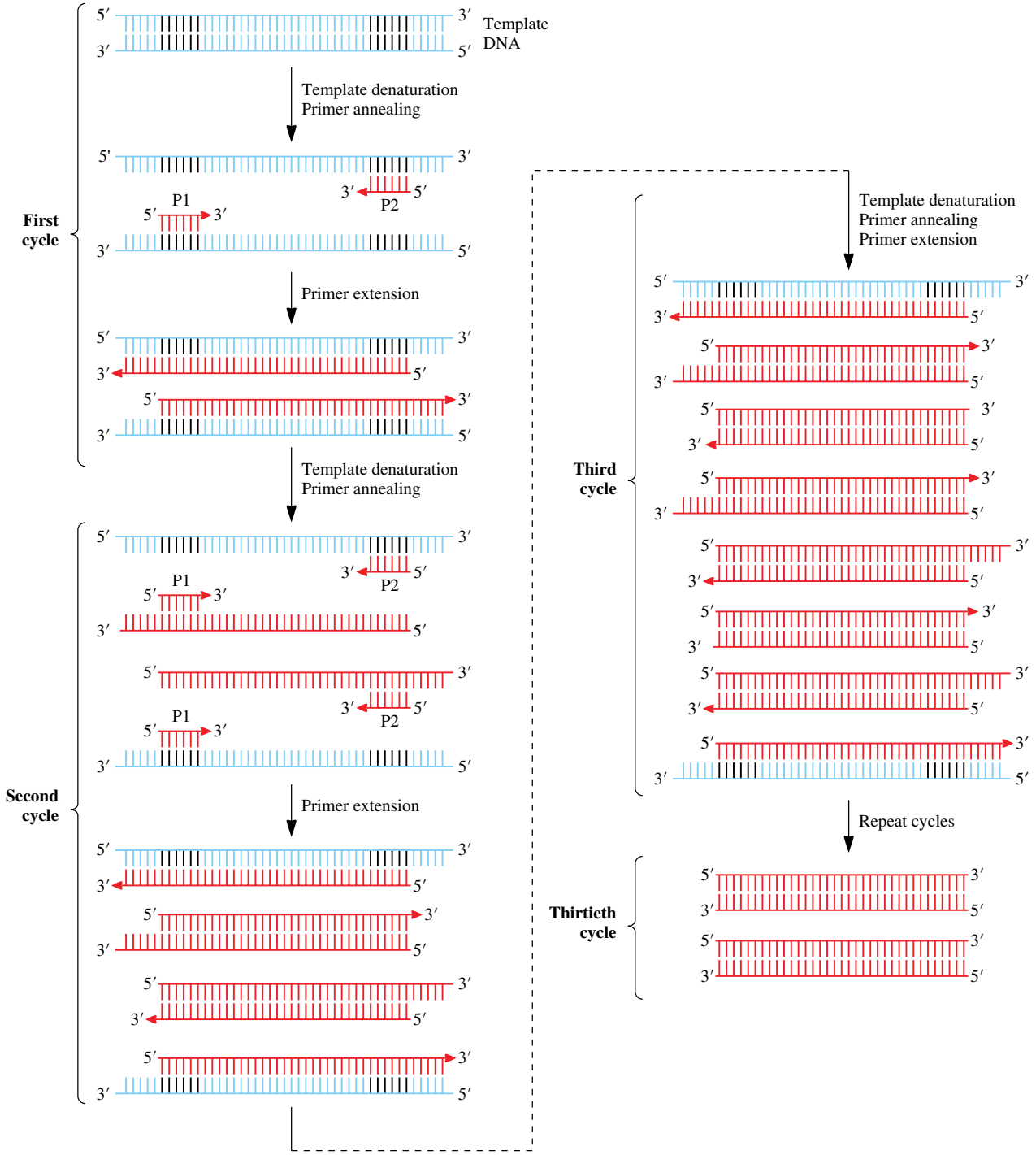


Figure 2.20 PCR. During a PCR cycle, the template DNA is denatured by heating and then slowly cooled to enable two primers (P1 and P2) to anneal to complementary (black) bases flanking the target DNA. The temperature is raised to about 70°C, and in the presence of the four deoxyribonucleotides, *Taq* DNA polymerase catalyzes the synthesis of a DNA strand extending from the 3′ hydroxyl end of each primer. In the first PCR cycle, DNA synthesis continues past the region of the template DNA strand that is complementary to the other primer sequence. The products of this reaction are two long strands of DNA that serve as templates for DNA synthesis during the second PCR cycle. In the second cycle, the primers hybridize to complementary regions in both the original strands and the long template strands, and DNA synthesis produces more long DNA strands from the original strands and short strands from the long template strands. A short template strand has a primer sequence at one end and the sequence complementary to the other primer at its other end. During the third PCR cycle, the primers hybridize to complementary regions of original, long template, and short template strands, and DNA synthesis produces long strands from the original strands and short strands from both long and short templates. By the end of the 30th PCR cycle, the products (amplicons) consist predominantly of short double-stranded DNA molecules that carry the target DNA sequence delineated by the primer sequences. Note that in the figure, newly synthesized strands are differentiated from template strands by a terminal arrow.

then hybridized with the primers. The large molar excess of primers in the reaction mixture ensures that they will hybridize to the template DNA before complementary template strands have the chance to reanneal to each other. A second round of synthesis produces long templates from the original strands as well as some DNA strands that have a primer sequence at one end and a sequence complementary to the other primer at the other end ("short templates") from the long templates (Fig. 2.20).

During the third cycle, short templates, long templates, and original strands all hybridize with the primers and are replicated (Fig. 2.20). In subsequent cycles, the short templates accumulate, and by the 30th cycle, these molecules, which are the desired PCR product, are about a million times more abundant than either the original or long template strands.

The specificity, sensitivity, and simplicity of PCR have rendered it a powerful technique that is central to many applications in molecular biotechnology, as illustrated throughout this book. For example, it is used to obtain large amounts of insert DNA for cloning, to detect specific mutations that cause genetic disease, to confirm biological relatives, to identify individuals suspected of committing a crime, and to diagnose infectious diseases (see chapter 4). Specific viral, bacterial, or fungal pathogens can be detected in samples from infected patients containing complex microbial communities by utilizing PCR primers that anneal to a sequence that is uniquely present in the genome of the pathogen. This technique is often powerful enough to discriminate among very similar strains of the same species of pathogenic microorganisms, which can assist in epidemiological investigations.

Cloning PCR Products

PCR is commonly used to amplify target DNA for cloning into a vector. To facilitate the cloning process, restriction enzyme recognition sites are added to the 5′ end of each of the primers that are complementary to sequences that flank the target sequence in a genome (Fig. 2.21). This is especially useful when suitable restriction sites are not available in the regions flanking the target DNA. Although the end of the primer containing the restriction enzyme recognition site lacks complementarity and therefore does not anneal to the target sequence, it does not interfere with DNA synthesis. Base-pairing between the 20 or so complementary nucleotides at the 3′ end of the primer and the template molecule is sufficiently stable for primer extension by DNA polymerase. At the end of the first cycle of PCR, the noncomplementary regions of the primer remain single stranded in the otherwise double-stranded DNA product. However, after the extension step of the second cycle, the newly synthesized complementary strand extends to the 5′ end of the primer sequence on the template strand, and therefore the PCR product contains a double-stranded restriction enzyme recognition site at one end (Fig. 2.21). Subsequent cycles yield DNA products with double-stranded restriction enzyme sites at both ends that can be cleaved to generate sticky ends for insertion into a vector. Alternatively, PCR products can be cloned using the single deoxyadenosine monophosphate (dAMP) that is added to the 3′ ends by *Taq* DNA polymerase, which lacks the proofreading activity of many DNA polymerases to correct mispaired bases. A variety of linearized vectors have been constructed that possess a single complementary 3′ deoxythymidine monophosphate (dTMP) overhang to facilitate cloning without using restriction enzymes (Fig. 2.22).

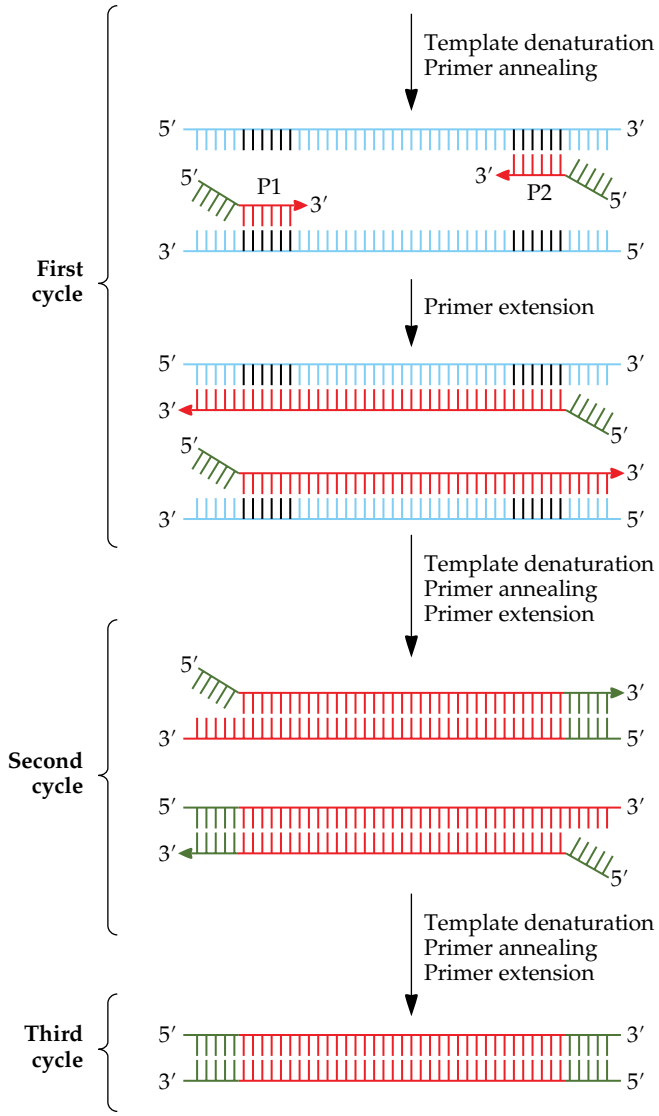


Figure 2.21 Addition of restriction enzyme recognition sites to PCR-amplified target DNA to facilitate cloning. Each of the two oligonucleotide primers (P1 and P2) has a sequence of approximately 20 nucleotides in the 3′ end that is complementary to a region flanking the target DNA (shown in black). The sequence at the 5′ end of each primer consists of a restriction endonuclease recognition site (shown in green) that does not base-pair with the template DNA during the annealing steps of the first and second PCR cycles. However, during the second cycle, the long DNA strands produced in the first cycle serve as templates for synthesis of short DNA strands (indicated by a terminal arrow) that include the restriction endonuclease recognition sequences at both ends. DNA synthesis during the third and subsequent PCR cycles produces double-stranded DNA molecules that carry the target DNA sequence flanked by restriction endonuclease recognition sequences. These linear PCR products can be cleaved with the restriction endonucleases to produce sticky ends for ligation with a vector. Note that not all of the DNA produced during each PCR cycle is shown.

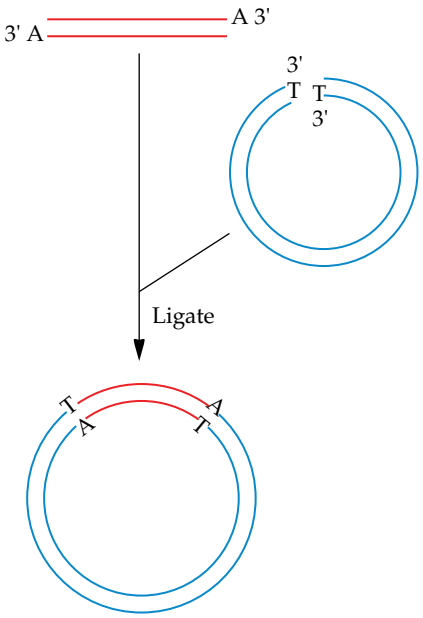


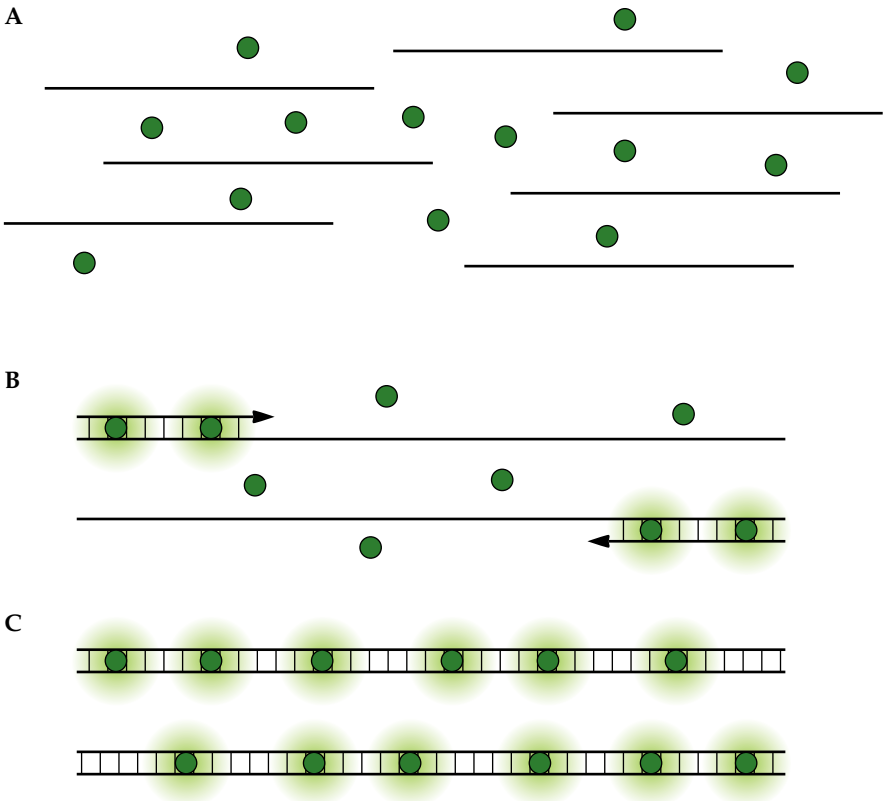
Figure 2.22 Cloning of PCR products without using restriction endonucleases. *Taq* DNA polymerase adds a single dAMP (A) to the ends of PCR-amplified DNA molecules. These extensions can base-pair with complementary single dTMP (T) overhangs on a specially constructed linearized cloning vector. Ligation with T4 DNA ligase results in insertion of the PCR product into the vector.

Quantitative PCR

PCR protocols have been developed to quantify the number of target DNA molecules, or RNA molecules after conversion to cDNA, present in a sample. Quantitative PCR is based on the principle that under optimal conditions, the number of DNA molecules doubles after each cycle. Typically, the amount of DNA present after each PCR cycle is measured in real time as the amount of fluorescence emitted by a fluorescent dye bound to the double-stranded DNA product. Thus, the fluorescence intensity increases in proportion to the concentration of double-stranded DNA (Fig. 2.23).

A real-time PCR occurs in four phases (Fig. 2.24A). In the first, or linear, phase (generally about 10 to 15 cycles), fluorescence emission at each cycle has not yet risen above the background level and therefore cannot be quantified accurately. In the second, or early exponential, phase, a sufficient amount of double-stranded DNA has been produced to increase the amount of fluorescence above a threshold level that is significantly higher than the background. The cycle at which this occurs is known as the threshold cycle (C_T). The C_T value is inversely correlated with the amount of target DNA in the original sample. During the third, exponential, phase, the amount of fluorescence continues to double as the DNA products of the reaction double in each cycle under ideal conditions. However, in the final, plateau, phase, the reaction components become limited, and measurements of the fluorescence intensity are no longer useful.

Figure 2.23 The fluorescent dye SYBR green does not bind to single-stranded DNA **(A)**, binds to double-stranded DNA as it is synthesized **(B)**, and is bound to the double-stranded amplified DNA **(C)**. Only the dye-bound DNA fluoresces.



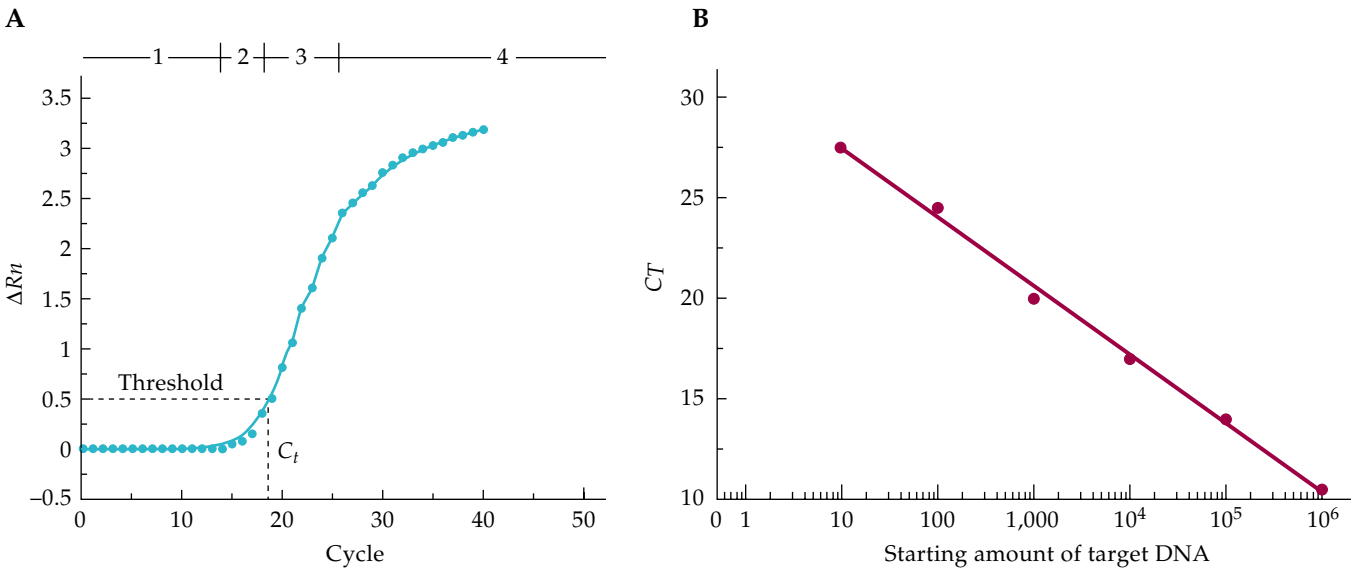


Figure 2.24 **(A)** Plot of normalized fluorescence (ΔRn) versus cycle number in a real-time PCR experiment. Four phases of PCR are shown: (1) a linear phase, where fluorescence emission is not yet above background level, (2) an early exponential phase, where the fluorescence intensity becomes significantly higher than the background (the cycle at which this occurs is generally known as C_T), (3) an exponential phase, where the amount of product doubles in each cycle, and (4) a plateau phase, where reaction components are limited and amplification slows down. **(B)** Plot of C_T versus the starting amount of a target nucleotide sequence. Fluorescence detection is linear over several orders of magnitude.

To quantify the amount of target DNA in a test sample, a standard curve is first generated by serially diluting a control sample with a known number of copies of the target DNA, and assuming all dilutions are amplified with equal efficiency, the C_T values for each dilution are plotted against the known starting amount of DNA (Fig. 2.24B). The number of copies of a target DNA in a test sample can then be determined by obtaining the C_T value for the test sample and extrapolating the starting amount from the standard curve. Since the amount of DNA doubles with each cycle during the exponential phase, a sample that has four times the number of starting copies of the target sequence compared to another sample would require two fewer cycles of amplification to generate the same number of product molecules. Often, a melt curve is generated to assess the specificity of the products, which denature at a characteristic temperature that is determined by their nucleotide sequence.

Among its many applications, quantitative real-time PCR has been used to monitor microorganisms that cause a range of infectious diseases. For example, it has been used to quantify *Salmonella enterica* contamination in food samples. In this case, food samples (chicken and mung beans were tested) were rinsed with water or a saline solution, and the liquid was filtered to collect the bacterial cells. The bacterial cells were removed from the filter membrane, lysed, and subjected to real-time PCR. The entire procedure took only approximately 3 hours and was able to detect and quantify as few as 700 *S. enterica* cells per 100 ml of liquid.

Chemical Synthesis of Genes

The ability to synthesize DNA with a specific sequence of nucleotides easily, inexpensively, and rapidly is essential for many of the methodologies of molecular biotechnology. There are a number of applications for synthetic genes, including large-scale production of proteins, testing protein function after changing specific codons, and creating nucleotide sequences that encode proteins with novel properties. Oligonucleotides are the key components for assembling genes. Chemical synthesis of short (10- to 100-nucleotide) single-stranded DNA oligonucleotides is technically straightforward, and two complementary oligonucleotides can be annealed to produce short fragments of double-stranded DNA. Oligonucleotides are enzymatically assembled into larger double-stranded DNA molecules to produce entire genes or whole chromosomes.

Assembling Oligonucleotides into Genes

One method for building a synthetic gene utilizes a set of overlapping oligonucleotides that are about 60 nucleotides in length with approximately 20-base overlaps (Fig. 2.25). After complementary 3′ and 5′ sequences are annealed, large gaps remain, but the base-paired regions are both long enough and stable enough to hold the structure together. After all the oligonucleotides are combined, the gaps are filled by enzymatic DNA synthesis with DNA polymerase I (usually from *E. coli*). This enzyme uses the 3′ hydroxyl groups as replication initiation points and the single-stranded regions as templates. After the enzymatic synthesis is completed, the nicks are sealed with T4 DNA ligase. For larger genes (≥1,000 bp), small sections of the gene are first assembled into units of about 500 bases in length, and then these are combined with other 500-base units. In turn, these larger kilobase segments are joined together until the entire sequence is completed. Computer programs are available both commercially and freely on the Internet which make it easier to determine the best set of oligonucleotides and overlaps for gene construction as well as allowing the user to select a particular codon usage, change any codon, and designate restriction endonuclease sites at specific locations. Finally, it is absolutely essential that a chemically synthesized gene have the correct sequence of nucleotides. Consequently, small synthetic genes are sequenced directly and, for larger genes, the sequences of each of the 500-base building blocks are determined before assembly.

Assembling PCR Products into Genes

A rapid method for synthesis of a gene begins with a series of overlapping PCR products that are assembled using a 5′ exonuclease together with DNA polymerase and DNA ligase (Fig. 2.26). This method is often referred to as Gibson assembly, named after Daniel Gibson, the scientist at the J. Craig Venter Institute who developed it. The PCR products contain sequences that span the entire gene and are synthesized using primers with sequences (15 to 25 nucleotides) that are complementary to the primers of adjacent fragments. Multiple PCR fragments are mixed together with the enzymes in a single reaction tube. The exonuclease removes several nucleotides

Synthetic Genomes

milestone

Chemically synthesized oligonucleotides have been assembled not only into genes but also into whole genomes. The first genome to be produced synthetically was the cDNA encoding the small (7,500 bp) single-stranded RNA genome of poliovirus (Cello et al., *Science* **297:**1016–1018, 2002). The poliovirus genome sequence was known and facilitated the design of 70-nucleotide-long, single-stranded oligonucleotides. Overlapping complementarity at the termini of neighboring oligonucleotides enabled their assembly into 400- to 600-bp fragments. The fragments were ligated into three larger segments that were subsequently digested with a restriction endonuclease and cloned in the correct order and orientation into a plasmid. Expression of the full-length cDNA from a suitable promoter in HeLa cells resulted in production of viral RNA and proteins that were assembled into infectious poliovirus particles.

Construction of a much larger synthetic bacterial genome presented a greater challenge (Gibson et al., *Science* **319:**1215–1220, 2008). The genome of the bacterium *Mycoplasma genitalium* was chosen because it has the smallest genome (a single chromosome of 580,076 bp) currently known for a free-living bacterium. The chromosome was initially produced in 101 segments of about 5 to 7 kb that were each assembled from synthetic oligonucleotides. The sequence of each segment overlapped its neighbor by approximately 80 bp, and therefore, following brief treatment with an exonuclease to generate sticky ends, four neighboring segments were joined *in vitro* by complementary base pairing at their termini and enzymatic repair of gaps. These 24-kb fragments were cloned into a bacterial artificial chromosome, which can carry large DNA inserts, and propagated in *E. coli*. This *in vitro* recombination method was repeated to assemble the fragments into successively larger pieces. Large segments carrying half- and full-genome sequences could not be cloned in *E. coli*, and therefore, quarter- and then half-genome fragments were finally assembled into a 582,970-bp sequence in the yeast *Saccharomyces cerevisiae* by *in vivo* recombination between overlapping homologous sequences.

These results demonstrated that a small bacterial genome can be constructed entirely from synthetic oligonucleotides. The next step was to show that a genome produced "from scratch" can direct the survival and growth of a living bacterium. The 1,077,947-bp genome of *Mycoplasma mycoides* was synthesized in a manner similar to that used to construct the *M. genitalium* genome (Gibson et al., *Science* **329:**52–56, 2010); *M. mycoides* was chosen because it grows at a faster rate than the extremely slow-growing *M. genitalium*. Briefly, overlapping synthetic oligonucleotides were assembled into 1,080-bp fragments. These fragments were recombined into 10-kb and then into 100-kb segments and, finally, into a full-length genome by *in vivo* homologous recombination in yeast. The intact *M. mycoides* genome was extracted from yeast and transplanted into a related species, *Mycoplasma capricolum*, replacing the recipient cell's chromosome. Remarkably, the synthetic genome was self-replicating and controlled the functions of a living bacterium. The bacterium was able to grow logarithmically and exhibited cellular and colony morphologies similar to those of *M. mycoides*. "Watermark" sequences (unique synthetic DNA sequences) were inserted in four regions to enable differentiation of the synthetic and natural genomes.

A major motivation for producing a synthetic bacterium is to understand the minimal genetic requirements for life, that is, to identify the minimum set of essential genes that can support survival and reproduction of a cell. Because it takes less energy and fewer resources to maintain and propagate a small genome, more resources can be directed to the synthesis of high yields of useful products from cloned genes.

from the 5′ ends of each strand to generate single-stranded 3′ overhangs that anneal to the complementary ends of another PCR fragment. DNA polymerase fills in gaps in the annealed fragments, and DNA ligase catalyzes the formation of a phosphodiester bond to seal nicks in the DNA backbone. The end result is a double-stranded DNA molecule that may be hundreds of kilobases long. The PCR fragments that represent the first and last sequences of the gene can be generated with primers that are complementary to the ends of a linearized plasmid to assemble and clone the gene in a single reaction (Fig. 2.27).

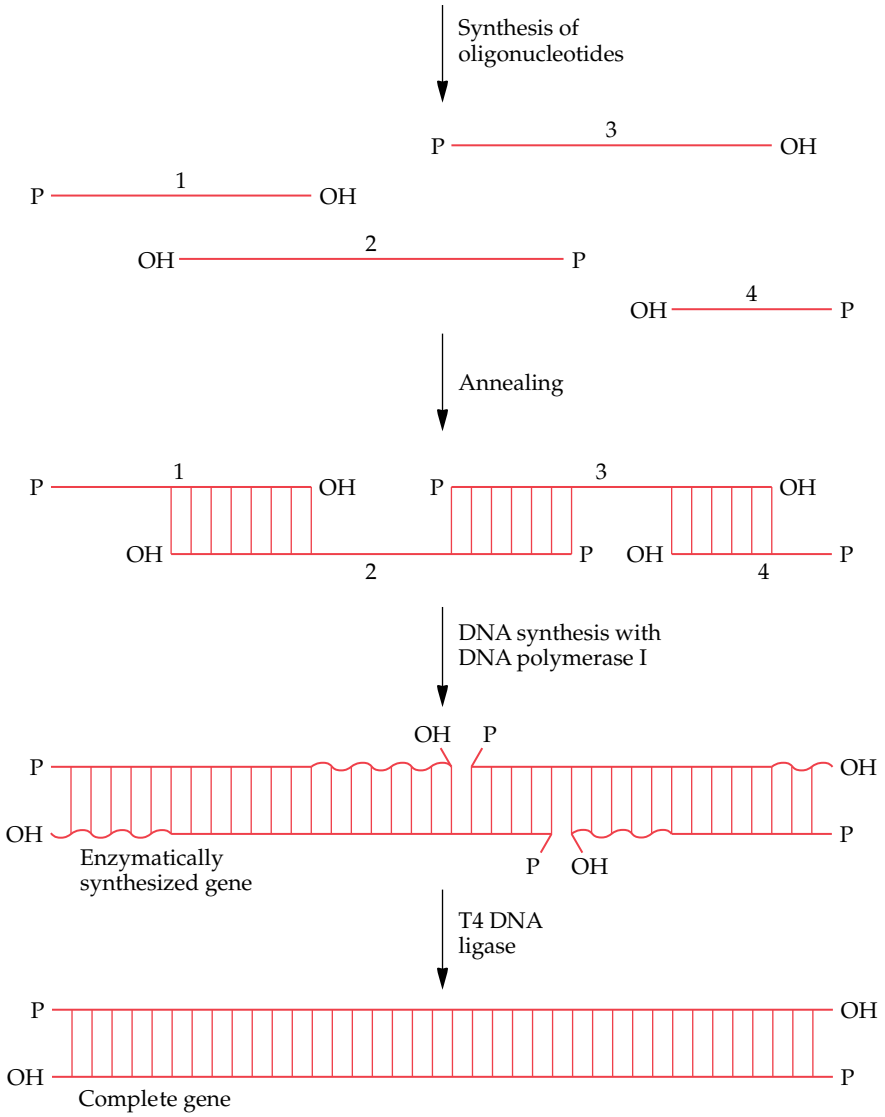


Figure 2.25 Assembly of a gene from oligonucleotides. Individual oligonucleotides are synthesized chemically and then hybridized. The sequences of the oligonucleotides are designed to enable them to form a stable molecule with base-paired regions separated by single-stranded regions (gaps). The gaps are filled in by *in vitro* enzymatic DNA synthesis. The nicks are sealed with T4 DNA ligase.

DNA Sequencing Technologies

Determination of the nucleotide composition and order in a gene or genome is a foundational technique in molecular biotechnology. Cloned or PCR-amplified genes and entire genomes are routinely sequenced. DNA sequences can often reveal something about the function of the protein encoded in a gene, for example, from predicted cofactor-binding sites, transmembrane domains, receptor recognition sites, or DNA-binding regions. The nucleotide sequences in noncoding regions that do not encode a protein or RNA molecule may provide information about the regulation of a gene. Comparison of gene sequences among individuals can reveal mutations

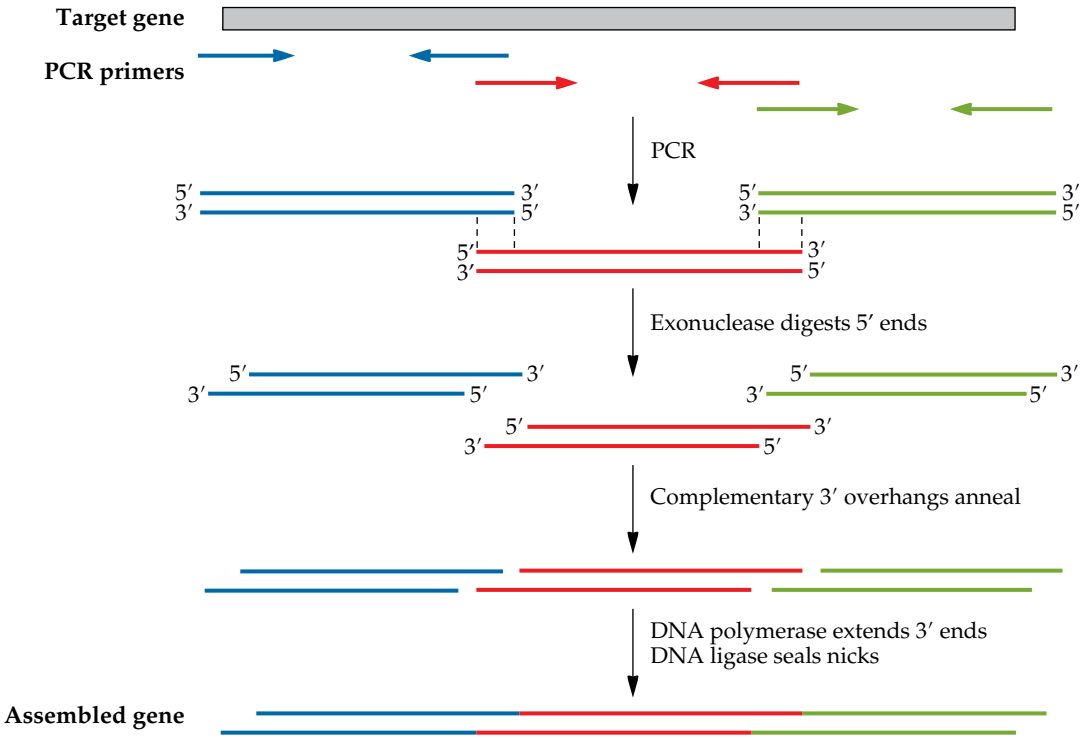


Figure 2.26 Assembly of a gene from PCR products (Gibson assembly). PCR fragments spanning the entire target gene contain sequences (15 to 25 nucleotides) at their ends that are homologous to the ends of adjacent fragments (indicated by broken lines). The fragments are mixed together and incubated with an exonuclease that digests the 5′ ends of all fragments to create single-stranded 3′ overhangs that are complementary to the 3′ overhang on an adjacent fragment. The complementary regions anneal, and DNA polymerase catalyzes the addition of nucleotides to close the gaps. DNA ligase catalyzes the formation of phosphodiester bonds to seal nicks in the backbone of the assembled DNA.

that contribute to phenotypic differences. For example, identification of nucleotide differences (polymorphisms) in a gene in individuals with a particular disease, but not in healthy individuals, may be used to predict disease susceptibility. Comparison of gene sequences among different organisms can lead to the development of hypotheses about the evolutionary relationships among organisms.

For more than three decades, the dideoxynucleotide procedure developed by the English biochemist Frederick Sanger has been used for DNA sequencing. This includes sequencing of DNA fragments containing one to a few genes and also the entire genomes from many different organisms, including the human genome. However, the interest in sequencing large numbers of DNA molecules in less time and at a lower cost has driven the development of sequencing technologies that can process thousands to millions of sequences concurrently. Many different faster and more efficient sequencing technologies have been developed. In general, most involve (i) enzymatic addition of nucleotides to a primer based on complementarity to a template DNA fragment and (ii) detection and identification of the nucleotide(s) added. These sequencing-by-synthesis methods employ

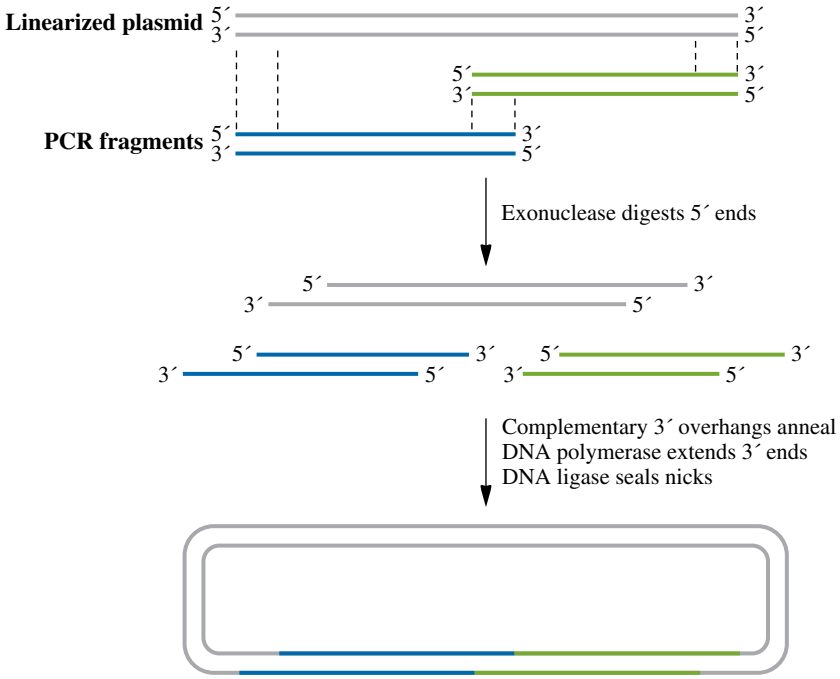


Figure 2.27 Gibson cloning. One or more PCR fragments may be inserted into a plasmid vector using the Gibson assembly procedure described in Fig. 2.26. The ends of the PCR fragments that will be adjacent to the plasmid contain a region of 15 to 25 nucleotides that is homologous to the plasmid sequence at the insertion site (homologous regions are indicated by broken lines). The plasmid sequences are included on the primers used to generate the PCR products. The plasmid is linearized (for example, by restriction enzyme digestion) and mixed with the PCR fragments. Following incubation with a 5′ exonuclease, the 3′ overhangs on the plasmid anneal to complementary overhangs on the PCR fragments to generate a circular plasmid carrying the cloned PCR fragments.

DNA polymerase to catalyze the addition of single nucleotides; however, the techniques differ in the method by which the addition is detected. Newer, more rapid and economical methods are being developed that do not require DNA synthesis and/or can generate a long sequence from a single DNA molecule.

Dideoxynucleotide Sequencing

The dideoxynucleotide procedure for DNA sequencing, also known as Sanger sequencing, is based on the principle that during DNA synthesis, addition of a nucleotide triphosphate requires a free hydroxyl group on the 3′ carbon of the sugar of the last nucleotide of the growing DNA strand (Fig. 2.28A). However, if a synthetic dideoxynucleotide that lacks a hydroxyl group at the 3′ carbon of the sugar moiety is incorporated at the end of the growing chain, DNA synthesis stops because a phosphodiester bond cannot be formed with the next incoming nucleotide (Fig. 2.28B). The termination of DNA synthesis is the defining feature of the dideoxynucleotide DNA sequencing method.

DNA Sequencing with Chain-Terminating Inhibitors

milestone

New techniques are the lifeblood of science. They enable researchers to acquire information that was previously inaccessible and that, in turn, generates insights that stimulate new research and lead to new discoveries. For molecular biotechnology, DNA sequencing is a powerful procedure that has become a laboratory mainstay. The most definitive form of the molecular characterization of a gene or genome is its sequence. Among other things, the coding content of a gene, potential primer sequences for a PCR, and the presence of mutations can be determined by DNA sequencing.

The earliest methods of nucleic acid sequencing were limited to very short RNA molecules. The sequencing of a DNA molecule required transcribing a DNA fragment into RNA with RNA polymerase and then sequencing the RNA product. In general, RNA sequencing entailed treating a radiolabeled RNA molecule with different ribonucleases, chromatographically separating the digestion products, redigesting the separated products, hydrolyzing the products of the second digestion with alkali, chromatographically separating the hydrolysis products, determining the sequence of the oligonucleotides, and constructing the sequence based on overlapping stretches of nucleotides. This approach was time-consuming and tedious. Subsequently, Allan Maxam and Walter Gilbert devised a method to sequence DNA directly by base-specific chemical cleavage (*Proc. Natl. Acad. Sci. USA* **74:**560–564, 1977). The Maxam and Gilbert method involved labeling one end of one strand of a DNA molecule with radioactive phosphorus and then chemically fragmenting the labeled molecules in four separate reactions that each cleaved specifically at one of the four nucleotide bases, about once per DNA molecule. The different-sized labeled fragments were separated by electrophoresis and visualized by X-ray film autoradiography to infer the sequence from the order of the radioactive bands. This was a significant improvement over earlier methods, as it allowed researchers to sequence DNA molecules of several hundred nucleotides in a couple of days. Although widely used for a period of 5 to 10 years, this method became obsolete with the development of a relatively simple, accurate, and reliable method using chain elongation inhibitors by Frederick Sanger (Sanger et al., *Proc. Natl. Acad. USA* **74:**5463–5467, 1977). The Sanger dideoxynucleotide method described in this chapter superseded the Maxam and Gilbert sequencing procedure when the M13 bacteriophage cloning system was developed, which provided single-stranded DNA templates required for sequencing. The M13 system was no longer required following the introduction of PCR-based cycle sequencing, which generates single-stranded DNA templates during a DNA denaturation step. Sanger and Gilbert received the Nobel Prize in Chemistry in 1980 for their "contributions concerning the determination of base sequences in nucleic acids."

In a dideoxynucleotide DNA sequencing procedure, a synthetic oligonucleotide primer (~17 to 24 nucleotides) anneals to a predetermined site on the strand of the DNA to be sequenced (Fig. 2.29A). The oligonucleotide primer defines the beginning of the region to be sequenced and provides a 3′ hydroxyl group for the initiation of DNA synthesis. The reaction tube contains a mixture of the four deoxyribonucleotides (deoxyadenosine triphosphate [dATP], deoxycytidine triphosphate [dCTP], deoxyguanosine triphosphate [dGTP], and deoxythymidine triphosphate [dTTP]) and four dideoxynucleotides (dideoxyadenosine triphosphate [ddATP], ddCTP, ddGTP, and ddTTP). Each dideoxynucleotide is labeled with a different fluorescent dye. The concentration of the dideoxynucleotides is optimized to ensure that during DNA synthesis a modified DNA polymerase incorporates a dideoxynucleotide into the mixture of growing DNA strands at every possible position. Thus, the products of the reaction are DNA molecules of all possible lengths, each of which includes the primer sequence at its 5′ end and a fluorescently labeled dideoxynucleotide at the 3′ terminus (Fig. 2.29B).

PCR-based cycle sequencing is performed to minimize the amount of template DNA required for sequencing. Multiple cycles of denaturation, primer annealing, and primer extension produce large amounts of

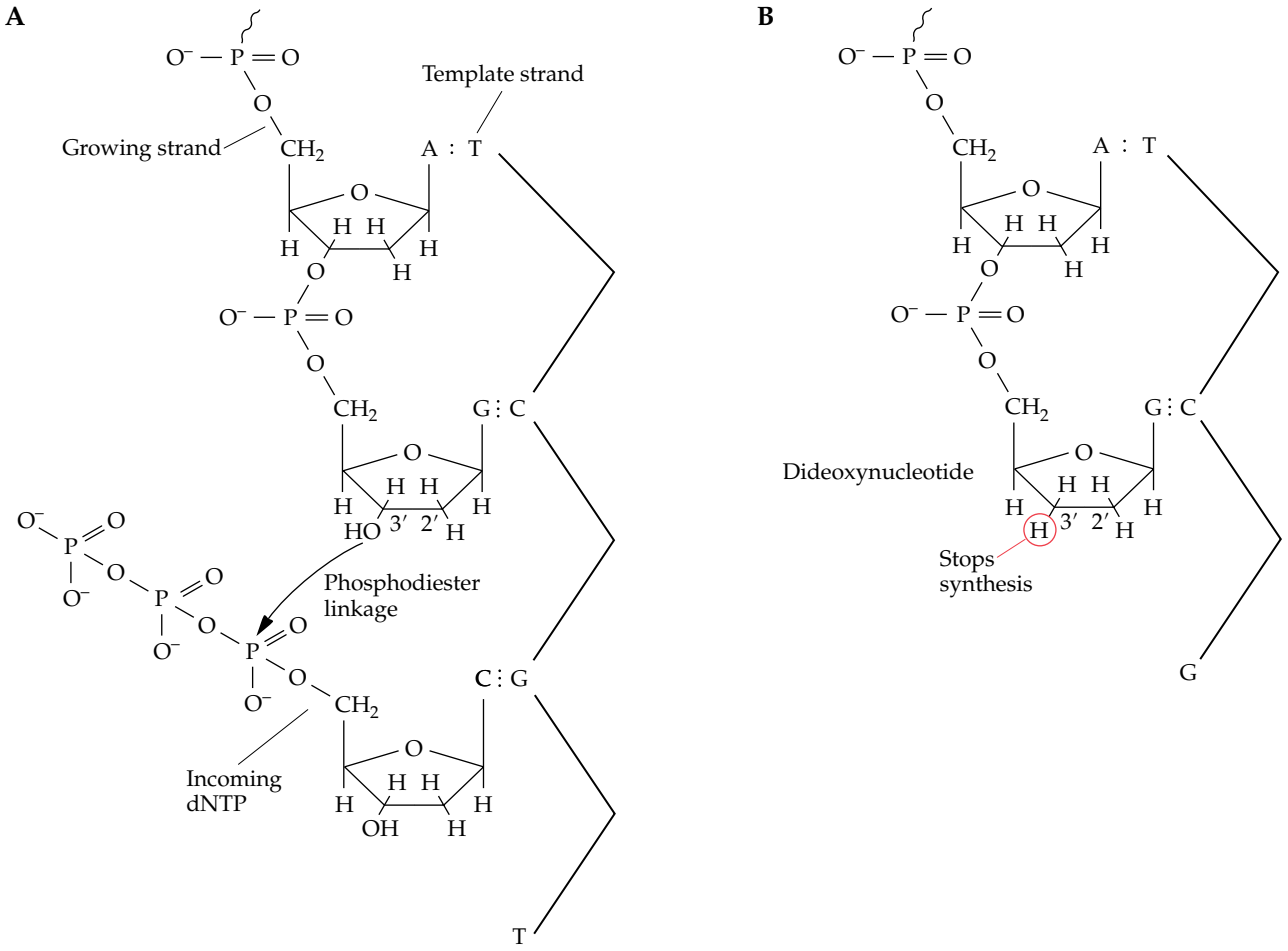


Figure 2.28 Incorporation of a dideoxynucleotide terminates DNA synthesis. **(A)** Addition of an incoming deoxyribonucleoside triphosphate (dNTP) requires a hydroxyl group on the 3′ carbon of the last nucleotide of a growing DNA strand. **(B)** DNA synthesis stops if a synthetic dideoxyribonucleotide that lacks a 3′ hydroxyl group is incorporated at the end of the growing chain because a phosphodiester bond cannot be formed with the next incoming nucleotide.

dideoxynucleotide-terminated fragments. These are applied to a polymer in a long capillary tube that enables separation of DNA fragments that differ in size by a single nucleotide. The fluorescently labeled fragments move through the polymeric matrix in an electric field based on their size, from smallest to largest, and pass by a laser that excites the fluorescent dye. Each of the four different fluorescent dyes emits a characteristic wavelength of light that represents a particular nucleotide, and the order of the fluorescent signals corresponds to the sequence of nucleotides (Fig. 2.29C). Generally, automated systems that employ this sequencing technology can determine about 500 to 1,000 bases per run (the read length or read). Although slower than the newer technologies, dideoxynucleotide sequencing is still a preferred method for small-scale sequencing projects due to its high accuracy (Table 2.3).

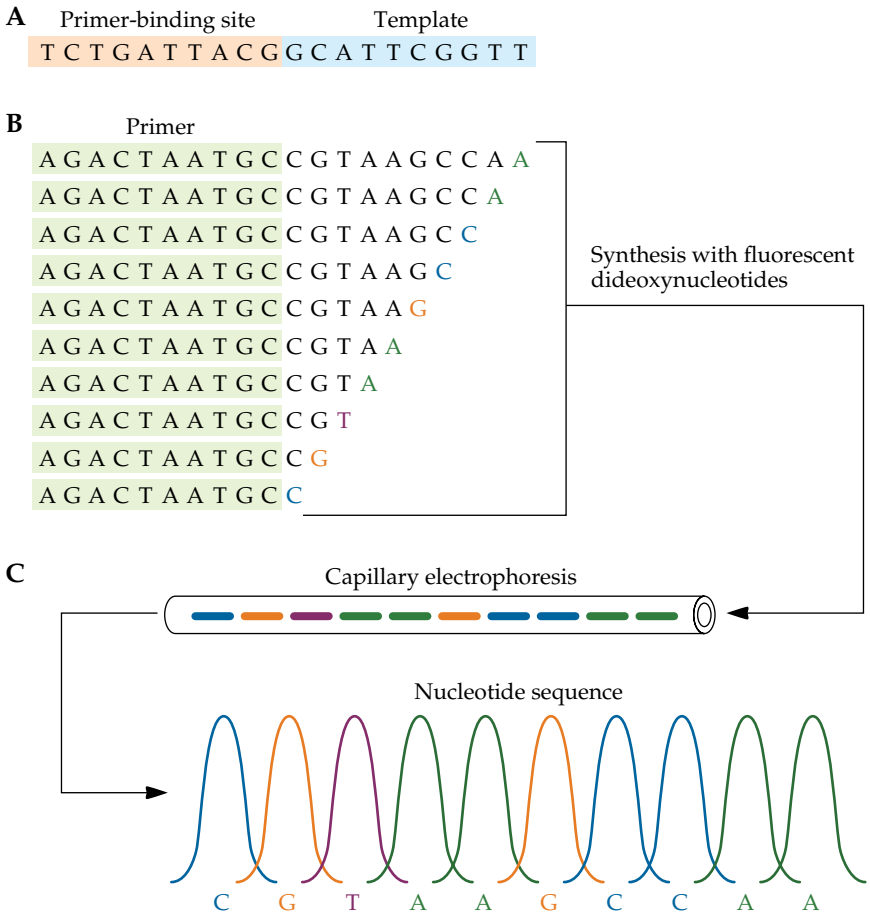


Figure 2.29 Dideoxynucleotide method for DNA sequencing. An oligonucleotide primer binds to a complementary sequence adjacent to the region to be sequenced in a single-stranded DNA template **(A)**. As DNA synthesis proceeds from the primer, dideoxynucleotides are randomly added to the growing DNA strands, thereby terminating strand extension. This results in DNA molecules of all possible lengths that have a fluorescently labeled dideoxynucleotide at the 3′ end **(B)**. DNA molecules of different sizes are separated by capillary electrophoresis, and as each molecule passes by a laser, a fluorescent signal that corresponds with one of the four dideoxynucleotides is recorded. The successive fluorescent signals are represented as a sequencing chromatogram (colored peaks) **(C)**.

Sequencing Using Reversible Chain Terminators

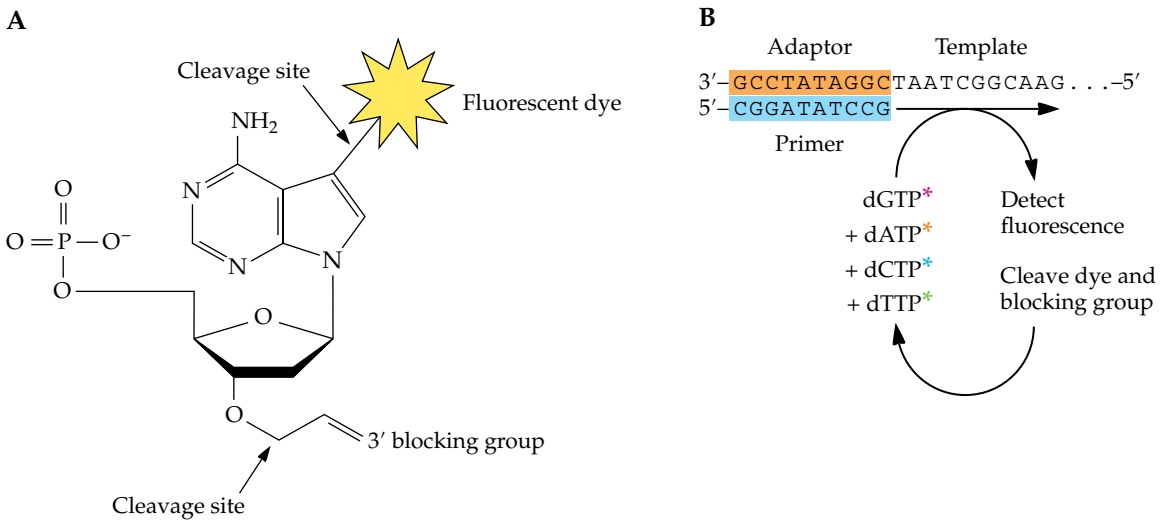
One next-generation sequencing-by-synthesis method, referred to as Illumina sequencing, utilizes reversible chain terminators. These are deoxynucleoside triphosphates with two important modifications: (i) a chemical blocking group is added to the 3′ carbon of the sugar moiety to prevent addition of more than one nucleotide during each round of sequencing, and (ii) a different fluorescent dye is added to each of the four nucleotides to enable identification of the incorporated nucleotide (Fig. 2.30A). The fluorophore is added at a position that does not interfere with the formation of hydrogen bonds between complementary base pairs or phosphodiester

Table 2.3 Comparison of sequencing technologies

Technology	Method	Read length (kb)	Error rate (%)	Advantages
Dideoxynucleotide sequencing (Sanger)	Synthesis	0.5–1	0.001	High accuracy
Sequencing using reversible chain terminators (Illumina)	Synthesis	0.1–0.3	0.1	Low cost, accurate
Single-molecule real-time sequencing (PacBio)	Synthesis	30–50	5–15	Rapid, continuous long reads, template amplification not required
Nanopore sequencing	Nanopore	10–200	5–20	Very long reads, portable, template amplification not required

bonds to incorporate the nucleotide during DNA synthesis. A short DNA adaptor that serves as a binding site for a sequencing primer is added to the 3′ end of the DNA template. DNA polymerase is employed to catalyze the addition of a modified nucleotide to the sequencing primer as specified by the DNA template sequence (Fig. 2.30B). After recording fluorescent emissions from the incorporated nucleotide, the fluorescent dye and the 3′ blocking group are removed. The blocking group is removed in a manner

Figure 2.30 Sequencing using reversible chain terminators. **(A)** Reversible chain terminators are modified nucleotides that have a removable blocking group on the oxygen of the 3′ position of the deoxyribose sugar to prevent addition of more than one nucleotide per sequencing cycle. To enable identification, a different fluorescent dye is attached to each of the four nucleotides via a cleavable linker. Shown is the fluorescent dye attached to adenine. **(B)** An adaptor sequence is added to the 3′ end of the DNA sequencing template that provides a binding site for a sequencing primer. All four modified nucleotides are added in a single cycle, and a modified DNA polymerase extends the growing DNA chain by one nucleotide per cycle. Fluorescence is detected, and then the dye and the 3′ blocking group are cleaved before the next cycle. Removal of the blocking group restores the 3′ hydroxyl group for addition of the next nucleotide.



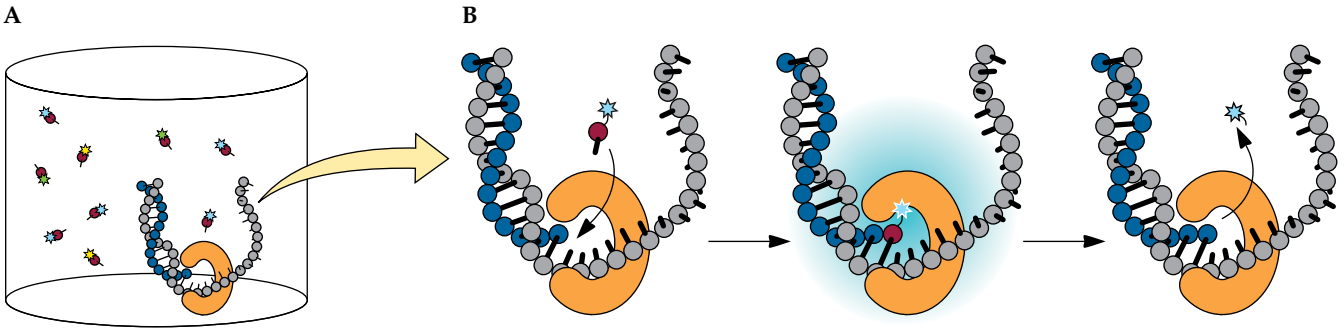
that restores the 3′ hydroxyl group of the sugar to enable subsequent addition of another nucleotide in the next cycle. Cycles of nucleotide addition to the growing DNA strand by DNA polymerase, acquisition of fluorescence data, and chemical cleavage of the blocking and dye groups are repeated to generate short read lengths, i.e., 100 to 300 nucleotides per run (Table 2.3).

Single-Molecule Real-Time Sequencing

To generate sufficiently high levels of a fluorescent signal for detection of nucleotide addition, the sequencing methods described above require large amounts of template DNA. A DNA amplification step is often required, which increases template preparation time and can introduce mutations that are interpreted as nucleotide variations. Recently, sequencing technologies have been developed to circumvent the amplification step. In one approach, known as Pacific Biosciences (PacBio) or single-molecule real-time (SMRT) sequencing, a single molecule of DNA polymerase is immobilized on a solid support (on the bottom of a nanoscale well) and captures a single DNA molecule that is bound to a primer (Fig. 2.31A). During the sequence acquisition stage, DNA polymerase extends the primer in a template-dependent fashion, and a signal corresponding to nucleotide addition is measured in a narrow volume at the bottom of the well (Fig. 2.31B).

The nucleotide added during the extension phase is detected in real time, as it is incorporated. For real-time sequencing, the nucleotides do not carry a blocking group on the 3′ hydroxyl group, and therefore DNA synthesis is continuous. A different fluorescent tag is attached to the terminal phosphate of each nucleoside triphosphate, in a manner that does not interfere with the activity of the DNA polymerase. With each nucleotide addition to the growing DNA chain, pyrophosphate is cleaved and, with it, the fluorescent tag. The laser and camera used to

Figure 2.31 Single-molecule real-time sequencing. One molecule of DNA polymerase (orange shape) is attached to the bottom of a nanoscale well. A single-stranded DNA molecule (gray strand) bound to a primer (blue strand) is captured in the active site of the polymerase **(A)**. Each of the four different nucleoside triphosphates is attached to a different fluorophore (colored stars) at the terminal phosphate, which is released during template-dependent nucleotide incorporation into the growing DNA strand. Fluorescence emission from a zeptoliter (10^{-21} liter) volume at the bottom of the well is detected by a laser before the cleaved pyrophosphate with attached fluorophore diffuses away **(B)**.



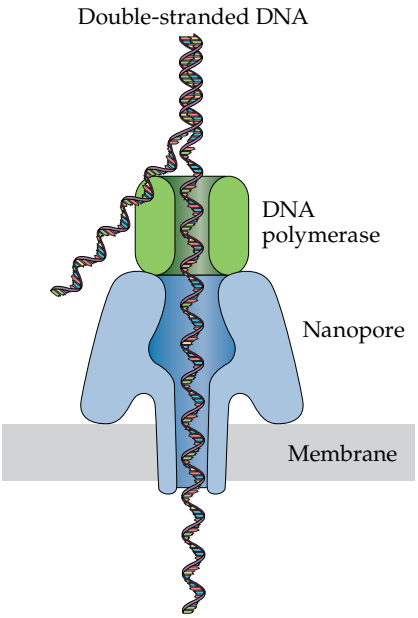


Figure 2.32 Nanopore sequencing. A double-stranded DNA molecule is bound to an adapter and a tethering protein that direct it to a protein nanopore embedded in a synthetic membrane. A motor protein such as DNA polymerase or helicase unwinds the DNA and threads one strand through the nanopore. As the DNA passes through the pore, an electrical current fluctuates in a manner that is characteristic of the nucleotides in the pore which enable their identification. The processivity of DNA polymerase facilitates the sequencing of long DNA molecules (several hundred kilobases) in a single run.

detect fluorescence are narrowly focused on the immobilized DNA polymerase and therefore record a pulse of fluorescence only in the brief time (tens of milliseconds) when the tagged nucleotide is held in the enzyme's active site (Fig. 2.31B). Following formation of a phosphodiester bond, the fluorescent tag cleaved from the nucleotide rapidly diffuses out of the range of the detector. Translocation of the DNA template positions DNA polymerase to accept the next nucleotide into the active site. The template DNA molecule is usually circularized to enable DNA polymerase to read it multiple times in a type of "rolling circle" amplification. Long sequence reads (up to 30 to 50 kb) can be generated rapidly by this method; however, accuracy is generally lower than that of other methods due to the short time interval between nucleotide additions, dissociation of a nucleotide before a phosphodiester bond forms, and simultaneous measurement of fluorescence from more than one nucleotide (Table 2.3). An advantage of the SMRT sequencing method is that, because it measures the rate of nucleotide addition, it can identify some epigenetic modifications such as methylated nucleotides which have a detectably different rate of incorporation compared to their unmodified counterparts.

Nanopore Sequencing

Nanopore sequencing is a newer single-molecule sequencing method that does not involve DNA synthesis. Instead, the DNA passes through a very small pore, and changes in electrical current are measured. A pore-forming protein, such as bacterial α-hemolysin, is inserted in a synthetic membrane across which an electrical current is applied. A long double-stranded DNA molecule is bound to a highly processive DNA polymerase, such as that from the bacteriophage phi29, that unwinds the DNA and threads one of the strands through the nanopore (Fig. 2.32). As the DNA strand passes through the pore, the electrical current is disrupted in a manner that is characteristic of the nucleotides in the pore. Changes in the electrical current are thereby interpreted as a nucleotide sequence. Although the error rate of this sequencing method is high (Table 2.3), it has the advantage of producing very long reads of more than 100 kb, can determine the sequences of both DNA and RNA molecules directly, and can detect epigenetic modifications. It is also available in a portable handheld format and is therefore useful for environmental monitoring, for surveillance during epidemics such as the SARS-CoV-2 pandemic, and in regions without access to sequencing facilities, including on the International Space Station.

Sequencing Whole Genomes

Just as the sequence of a gene can provide information about the function of the encoded protein, the sequence of an entire genome can contribute to our understanding of the nature of an organism. Tens of thousands of whole genomes have now been sequenced, from organisms of all domains of life. Initially, the sequenced genomes were relatively small, limited by the early sequencing technologies. The first DNA genome to be sequenced was from the *E. coli* bacteriophage ΦX174 (5,375 bp) in 1977, while the first sequenced genome from a cellular organism was that of the bacterium

Haemophilus influenzae (1.8 Mbp) in 1995. Within 2 years, the sequence of the larger *E. coli* genome (4.6 Mbp) was reported, and the sequence of the human genome (3,000 Mbp), the first vertebrate genome, was completed in 2003.

Most of these first genome sequences were determined using a shotgun cloning approach. In this strategy, a clone library of randomly generated, overlapping genomic DNA fragments is constructed in a bacterial host. The plasmids are isolated, and then the cloned inserts are sequenced using the dideoxynucleotide method. Using this approach, the first human genome was sequenced in 13 years at a cost of $2.7 billion. The aspiration to acquire genome sequences faster and at a much lower cost has driven the development of new genome sequencing strategies. Today, many large-scale sequencing projects have been completed and many more are under way, motivated by compelling biological questions. Some will contribute to our understanding of the microorganisms that cause infectious diseases and to the development of new techniques for their detection and treatment. Others are aimed at helping us to understand what it means to be human and how we evolved. Understanding the nucleotide polymorphisms among individuals with and without a specific disease will help us to determine the genetic basis of disease.

Generally, DNA sequencing projects fall into two categories: *de novo* genome sequencing and resequencing. Sequencing the genome of an organism that has not previously been sequenced is *de novo* genome sequencing, whereas resequencing involves comparing a newly determined sequence with a known reference sequence. A large-scale sequencing project typically entails (i) preparing a library of template DNA fragments, (ii) amplifying the DNA fragments which will increase the detection signal from nucleotide addition during the sequencing reaction, (iii) sequencing the template DNA using one of the sequencing techniques described above, and (iv) assembling the sequences generated from the fragments in the order in which they are found in the original genome. Sequencing massive amounts of DNA required not only the development of new technologies for nucleotide sequence determination but also new methods to reduce the time for preparation and processing of large libraries of sequencing templates. High-throughput next-generation sequencing approaches have circumvented the cloning steps of the shotgun sequencing strategy by attaching, amplifying, and sequencing the genomic DNA fragments directly on a solid support. Single-molecule sequencing eliminates the need for an amplification step, further reducing the time for template preparation. In both cases, all of the templates are sequenced at the same time. The term used to describe this is "massive parallelization."

Preparation of Genomic DNA Sequencing Libraries

Although the shotgun cloning strategy has been used successfully to obtain the sequences of many whole genomes, preparation of clone libraries in bacterial cells is costly and time-consuming for routine sequencing of the large amounts of genomic DNA that are required for many research and clinical applications. To reduce the time and cost of large-scale sequencing, high-throughput next-generation sequencing strategies have been developed that use cell-free methods to generate a library of genomic DNA fragments. First, purified, genomic DNA is

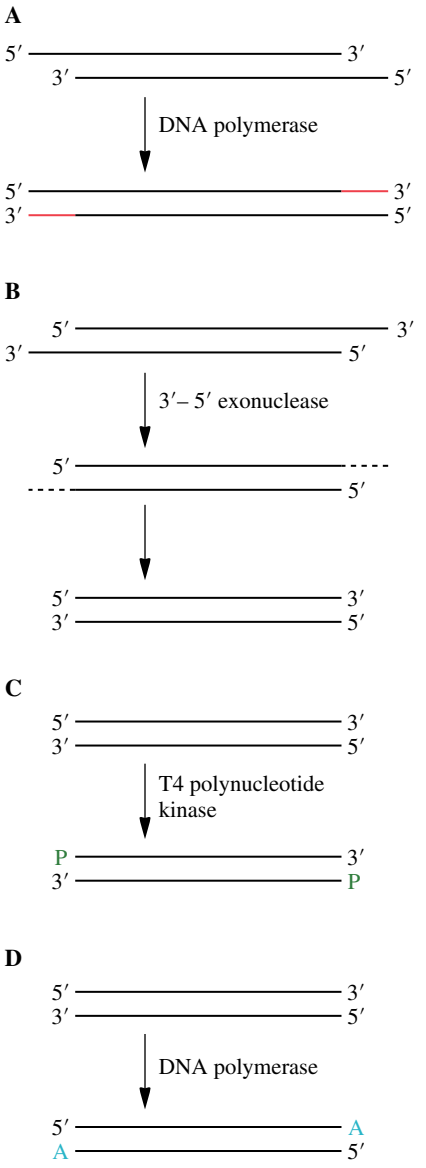
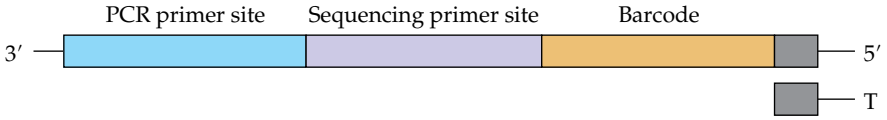


Figure 2.33 Preparation of genomic DNA fragments for ligation of adaptors. Ends of frayed DNA are repaired (end-repaired) using DNA polymerase to fill in from recessed 3′ ends **(A)** and a 3′ exonuclease to degrade 3′ extensions **(B)**. Note: Fragments with different combinations of extensions and recessed ends are not shown here. In all of these cases, the polymerase and/or 3′ exonuclease activities produce blunt-end DNA molecules. T4 polynucleotide kinase phosphorylates the 5′ ends of the blunt-end fragments **(C)**. A single deoxyadenosine monophosphate may be added to the 3′ ends of blunted fragments (A-tailed) by DNA polymerase to facilitate ligation of adaptors that have a single complementary 3′ deoxythymidine monophosphate overhang **(D)**.

fragmented either mechanically by sonication (applying high-frequency sound energy) or nebulization (forcing DNA through a small hole using compressed air), or by enzymatic digestion. Physical fragmentation tends to leave extended single-stranded ends that must be blunted (end-repaired or polished) by filling in 3′ recessed ends with DNA polymerase in the presence of the four deoxyribonucleotides (Fig. 2.33A) and removing protruding 3′ ends with an exonuclease (Fig. 2.33B). Next, different oligonucleotide adaptors are ligated to each end of the polished genomic fragments. To facilitate ligation, the 5′ ends of the DNA fragments are phosphorylated with T4 polynucleotide kinase (Fig. 2.33C). The 3′ ends may be adenylated (A-tailed) by enzymatic addition of a single deoxyadenosine monophosphate to facilitate ligation of adaptors that have a single complementary 3′ deoxythymidine monophosphate overhang (Fig. 2.33D). The adaptors have sequences that anneal to PCR primers for amplification of the genomic sequence and to sequencing primers that prime the sequencing reaction (Fig. 2.34). In addition, adaptors may contain an index (barcode) sequence for tagging a genomic library. A barcode is a short (usually 8 to 12 nucleotides), unique nucleotide sequence that is used to identify and sort sequence reads generated from a genomic library when multiple libraries are pooled prior to sequencing (multiplexing). Generally, genomic DNA fragments are size-selected by removing fragments above and below a certain size (typically 200 to 500 bp) to facilitate assembly of the genome sequence (described below). Following size selection, the libraries may be amplified by PCR to enrich for genomic DNA fragments that have adaptors ligated to both ends and to increase the amount of template for sequencing.

Figure 2.34 Features of an adaptor used for preparation and sequencing of genomic DNA fragments. A single 3′ deoxythymidine monophosphate overhang facilitates ligation of the adaptor to the ends of A-tailed genomic DNA fragments (Fig. 2.33D). The adaptor has a sequence that anneals to an oligonucleotide primer which captures and amplifies the genomic sequence on a solid support, a sequence that anneals to a sequencing primer that primes the sequencing reaction, and a unique barcode sequence that is used to tag a genomic library when multiple libraries are combined for sequencing (multiplexing).

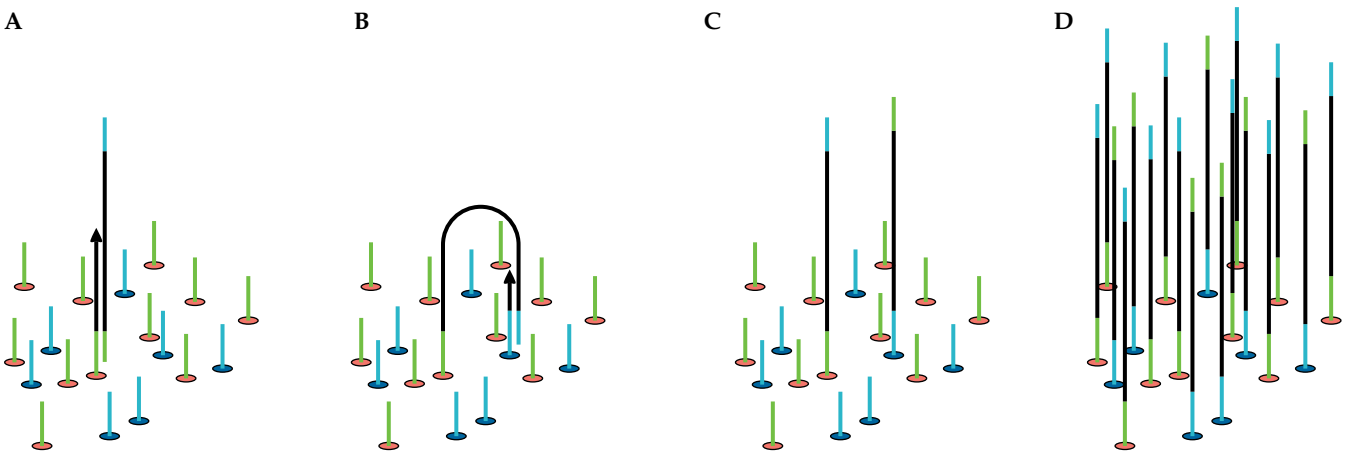


High-Throughput Next-Generation Sequencing

Most of the current commercially available, high-throughput next-generation sequencing strategies use PCR to generate clusters containing millions of copies of each DNA sequencing template. In one strategy, single-stranded sequencing templates (denatured library fragments) are captured by hybridization via the adaptor sequence that is complementary to oligonucleotides covalently bound to a solid surface, such as a glass slide (Fig. 2.35A). The oligonucleotides also act as primers for DNA polymerase to synthesize the strand complementary to the captured template strand (Fig. 2.35B). The resulting double-stranded DNA is denatured, and the original template is washed away. The remaining strand is anchored to the solid support via the bound oligonucleotide primer at one end and at the other end carries the adaptor sequence that is complementary to an oligonucleotide primer that is adjacent on the glass slide, to which it hybridizes (Fig. 2.35C). Following extension of the second primer (bridge PCR) and denaturation of the double-stranded product, two single-stranded DNA molecules are anchored to the solid support (Fig. 2.35D). The process is repeated many times to generate clusters of about a thousand copies of each sequencing template (Fig. 2.35E).

The nucleotide sequence of each template may be acquired by addition of a sequencing primer that is complementary to an adaptor sequence, DNA polymerase, and four differentially labeled fluorescent reversible chain terminators (Fig. 2.30). A single sequencing cycle consists of addition

Figure 2.35 Generation of clusters of sequencing templates. Denatured genomic DNA library fragments are captured on a glass slide by annealing to a bound oligonucleotide via a complementary adaptor sequence **(A)**. The oligonucleotide primes the synthesis of the complementary strand (denoted by the black arrow). The resulting double-stranded DNA is denatured, and the original library fragment is washed away. The newly synthesized strand remains anchored to the glass slide at one end and binds to an adjacent oligonucleotide primer by the adaptor sequence at the other end **(B)**. The complementary strand is synthesized by extension of the second primer in a process known as bridge amplification. Denaturation of the double-stranded product results in two single-stranded DNA molecules that are bound to the glass slide **(C)**. The process is repeated many times to generate clusters of about a thousand copies of each sequencing template **(D)**.



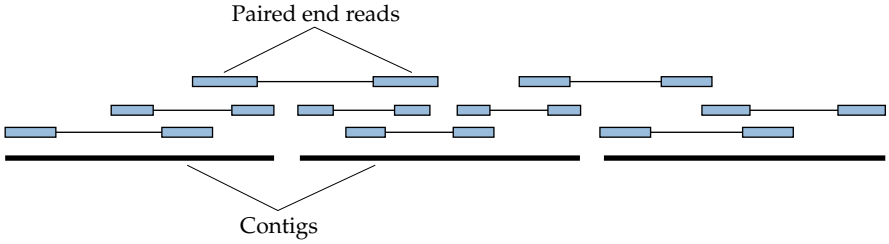
of these reagents to each cluster containing the PCR-amplified copies of a fragment of genomic DNA and then capturing the fluorescent signal generated by addition of a single nucleotide (as described above). The spectrum of fluorescence corresponds to the nucleotide added, and the cycle is repeated to determine the sequence of 50 to 150 nucleotides from one or both ends of each template. This process occurs simultaneously for hundreds of millions of clusters anchored to the solid support.

The accuracy of the sequence is assessed using a Phred quality score (Q), which indicates the probability (P) that a base is identified incorrectly, as described in the following equation: $Q = -10 \log_{10} P$. For example, a Q score of 30 (Q30) indicates that there is a 1 in 1,000 chance that a base is called incorrectly, or that the base call accuracy is 99.9%.

Genome Sequence Assembly

A genome sequence can be assembled by aligning the sequences of DNA reads with sequences from a previously determined and highly related (reference) genome. For example, reads from resequenced human genomes, that is, genomes from different individuals, are mapped to a reference human genome. Alternatively, when a reference sequence is not available, the reads can be assembled *de novo* using a computer program that aligns the matching ends of different reads. The process of generating successive overlapping sequences produces long, contiguous stretches of nucleotides called contigs. The presence of repetitive sequences in a genome can result in erroneous matching of overlapping sequences. This problem can be overcome by using the sequences from both ends of a DNA fragment (paired-end reads), which are a known distance apart (when genomic DNA fragments are size-selected prior to sequencing), to order and orient the reads and to assemble the contigs into larger scaffolds (Fig. 2.36). Many overlapping reads are required to ensure that the nucleotide sequence is accurate and assembled correctly. Each nucleotide site in a genome is generally sequenced many times from different fragments. The extent of sequencing redundancy, called coverage, or depth of coverage, varies from 10 to more than 100, depending on the error rate of the sequencing method, the read length (shorter reads require greater coverage), the complexity of the genome, the assembly method, and the goal of the sequencing project.

Figure 2.36 Genome sequence assembly. Sequence data generated from both ends of a DNA fragment are known as paired-end reads (paired ends are shown in blue for each fragment, and the distance between them is represented by a thin, black line). A large number of reads are generated and assembled into longer contiguous sequences (contigs) using a computer program that matches overlapping sequences. Paired ends help to determine the order and orientation of contigs as they are assembled into scaffolds. Shown is a scaffold consisting of three contigs.



The assembly process generates a draft sequence; however, small gaps may remain between contigs. Although a draft sequence is sufficient for many purposes, for example, in resequencing projects that map a sequence onto a reference genome, in some cases it is preferable to close the gaps to complete the genome sequence. For *de novo* sequencing of genomes from organisms that lack a reference genome, gap closure is desirable. Small gaps (less than 10 kb) can be closed by PCR amplification using primers designed to anneal to known sequences on each side of the gap and genomic DNA as a template and then sequencing the PCR product. Sequencing of additional libraries containing fragments of different sizes may be required to determine the sequence of large gaps.

Sequencing Metagenomes

Traditionally, the identification of microorganisms and characterization of their biological functions have required cultivating each strain in the laboratory. In the 1990s, with the emergence of techniques to extract DNA directly from environmental samples such as soil and seawater, researchers began to examine the sequence diversity of bacteria using the universal 16S ribosomal RNA gene as a taxonomic marker. These studies revealed that fewer than 1% of all bacterial species could be cultured, and therefore, novel genes that might be of considerable interest for basic and applied research were inaccessible using methods that depended on growth of bacteria in the laboratory. Considering the wealth of biotechnologically important genes and proteins that had been obtained from the relatively few culturable microorganisms, the possibility of harvesting useful genes from the much greater number of unculturable microorganisms was exciting, if not daunting. With the development of high-throughput next-generation sequencing and algorithms for assembling genome sequences, it has become possible to access the genomes of uncultured organisms from complex environmental and clinical samples. The study of the collective genomes in these samples is known as metagenomics.

The primary objective of a metagenomic project is to construct a comprehensive DNA library from all the microorganisms of a particular ecosystem or location (Fig. 2.37). The entire library is sequenced using a massively parallel approach and assembled into contigs as described above with the aim of determining the sequence of as many different genomes as possible and identifying both novel gene sequences and those that are similar (homologous) to known gene sequences. For example, metagenomic analysis of the microbiome associated with *Mycale hentscheli*, a marine sponge known for its production of diverse natural products, revealed several biosynthetic gene clusters for polyketides, a diverse group of secondary metabolites with significant pharmacological potential as antiviral and/or antitumor agents. DNA extracted from two sponge specimens and sequenced using Illumina technology yielded 366 Mbp that was assembled into 34 bacterial genomes with 60 to 100% completeness. The sequences revealed several large gene clusters encoding the biosynthetic enzymes for three classes of cytotoxic polyketides (mycalamides, pateamines, and pelorusides) that inhibit translation or microtubules (Fig. 2.38). In addition, numerous genes predicted to encode enzymes for other bioactive metabolites were identified. The gene clusters were distributed among diverse bacteria, including previously unknown, uncultivated bacteria

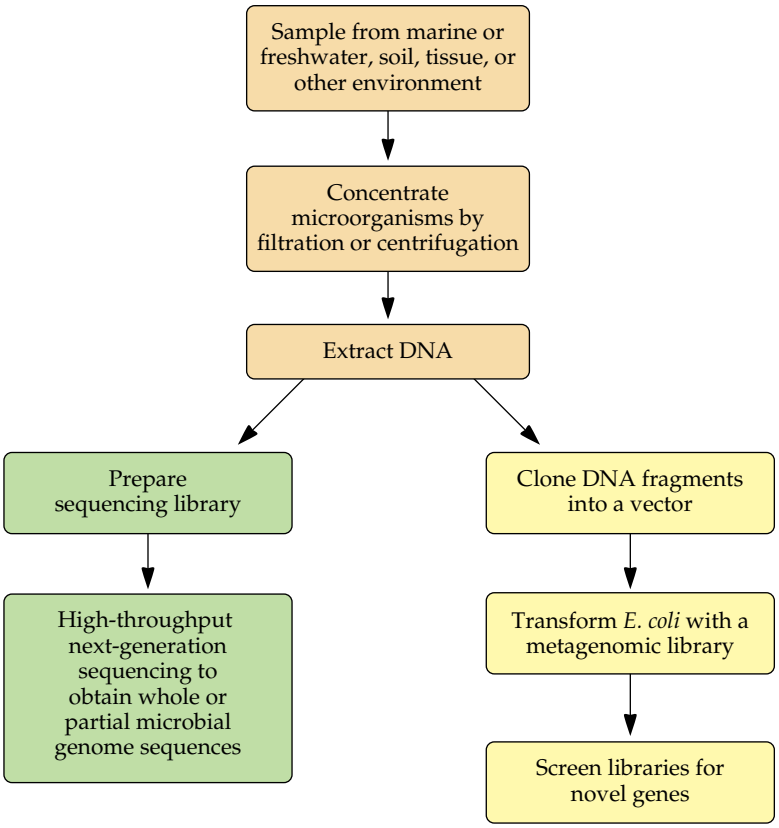


Figure 2.37 Construction of metagenomic libraries. Bacteria and/or viruses in samples from various environments or tissues are concentrated before extracting and then fragmenting the DNA. Libraries containing the DNA fragments are sequenced or screened for novel genes.

(provisionally assigned to the phylum *Candidatus*). Identification of the producer and knowledge of the biosynthetic gene sequences will facilitate production in a heterologous host and testing of these compounds. This research hints at the enormous and largely untapped reservoir of metabolites with therapeutic potential that can be revealed by metagenome sequencing.

Genomics

Genome sequence determination is only a first step in understanding an organism. The next steps require identification of the features encoded in a sequence and investigations of the biological functions of the encoded proteins and regulatory elements that determine the physiology and ecology of the organism. The area of research that generates, analyzes, and manages the massive amounts of information about genome sequences is known as genomics.

Sequence data are deposited and stored in databases that can be searched using computer algorithms to retrieve sequence information (data mining or bioinformatics). Public databases such as GenBank (National Center for Biotechnology Information, USA), the European

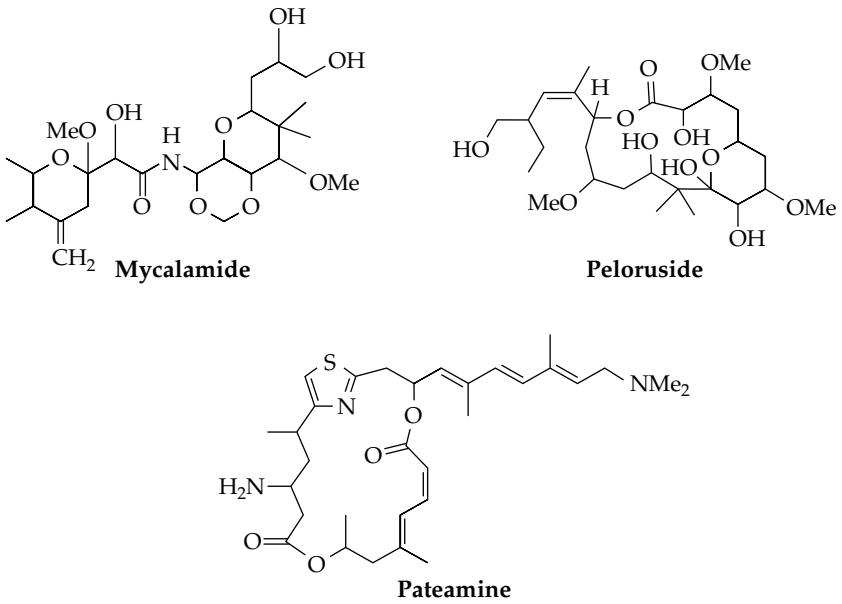


Figure 2.38 Polyketides produced by bacteria associated with the marine sponge *Mycale hentscheli* have therapeutic potential as antiviral and/or antitumor agents. Mycalamides bind to ribosomes and inhibit translation, pateamines inhibit translation initiation, and pelorusides stabilize microtubules. Data from Rust et al. *Proc. Natl. Acad. Sci. U S A* **117:**9508–9518, 2020.

Nucleotide Archive (European Molecular Biology Laboratory), and the DNA Data Bank of Japan (National Institute of Genetics) receive sequence data from individual researchers and from large sequencing facilities and share the data as part of the International Nucleotide Sequence Database Collaboration. Sequences can be retrieved from these databases via the Internet. Many specialized databases also exist, for example, for storing genome sequences from individual organisms, protein-coding sequences, regulatory sequences, sequences associated with human genetic diseases, gene expression data, protein structures, protein-protein interactions, and many other types of data.

One of the first analyses to be conducted on a new genome sequence is the identification of descriptive features, a process known as annotation. Some annotations are protein-coding sequences (open reading frames), sequences that encode functional RNA molecules (e.g., rRNA and tRNA), regulatory elements, and repetitive sequences. Annotation relies on algorithms that identify features based on conserved sequence elements such as translation start and stop codons, intron-exon boundaries, promoters, transcription factor-binding sites, and known genes (Fig. 2.39). It is important to note that annotations are often predictions of sequence function based on homology to sequences of known functions. In many cases, the function of the sequence remains to be verified through experimentation.

Comparison of a genome sequence to other genome sequences can reveal interesting and important sequence features. Comparisons among closely related genomes may reveal polymorphisms and mutations based on sequence differences. Association of specific polymorphisms with diseases can be used to predict, diagnose, and treat human diseases. The 100,000 Genomes Project undertaken by Genomics England aims to use

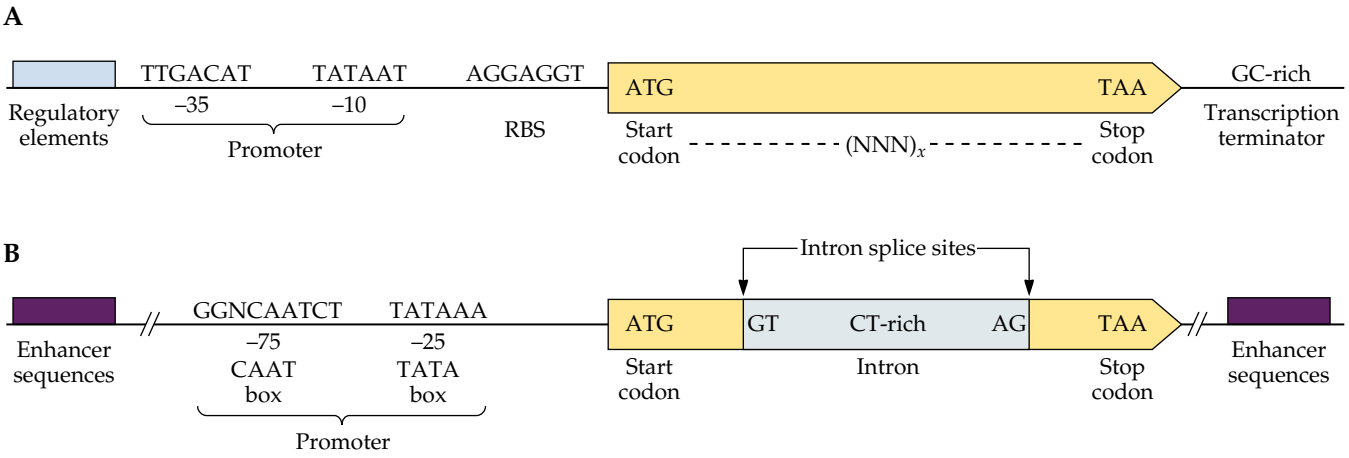


Figure 2.39 Genome annotation utilizes conserved sequence features. Predicting protein-coding sequences (open reading frames) in prokaryotes **(A)** and eukaryotes **(B)** requires identification of sequences that correspond to potential translation start (ATG or, more rarely, GTG or TTG) and stop (TAA, shown; also TAG or TGA) codons in mRNA. The number of nucleotides between the start and stop codons must be a multiple of three (i.e., triplet codons) and must be a reasonable size to encode a protein. In prokaryotes, a conserved ribosome-binding site (RBS) is often present 4 to 8 nucleotides upstream of the start codon **(A)**. Prokaryotic transcription regulatory sequences such as an RNA polymerase recognition (promoter) sequence and binding sites for regulatory proteins can often be predicted based on similarity to known consensus sequences. Transcription termination sequences are not as readily identifiable but are often GC-rich regions downstream of a predicted translation stop codon. In eukaryotes, protein coding genes typically have several intron sequences in primary RNA that are delineated by GU and AG and contain a pyrimidine-rich tract **(B)**. Introns are spliced from the primary transcript to produce mRNA. Transcription regulatory elements such as the TATA and CAAT boxes that are present in the promoters of many eukaryotic protein-coding genes can sometimes be predicted. Sequences that are important for regulation of transcription are often difficult to predict in eukaryotic genome sequences; for example, enhancer elements can be thousands of nucleotides upstream and/or downstream from the coding sequence that they regulate.

whole-genome sequencing to identify nucleotide polymorphisms associated with rare diseases and cancers. Traditionally, cancer genetic research has investigated specific genes that were hypothesized to play a role in tumorigenesis based on their known cellular functions, for example, genes encoding transcription factors that control expression of cell division genes. Although important, this gives an incomplete view of the genetic basis for cancers. Sequencing tumor genomes and comparing the sequences to those of normal cells have revealed point mutations, copy number mutations, and structural rearrangements associated with specific cancers. For instance, comparison of the genome sequences from acute myeloid leukemia tumor cells and normal skin cells from the same patient revealed eight previously unidentified mutations in protein-coding sequences that are associated with the disease. Comparison of the genomes of bacterial pathogens with those from closely related nonpathogens has led to the identification of virulence genes. Unique sequences can be used for pathogen detection, and genes encoding proteins that are unique to a pathogen are potential targets for antimicrobial drugs and vaccine development.

Genome comparisons among distantly related organisms enable scientists to make predictions about evolutionary relationships. For example, the Genome 10K Project aims to sequence and analyze the genomes of 10,000 vertebrate species, roughly 1 per genus. Comparison of these sequences will contribute to our understanding of the genetic changes that led to the diversity in morphology, physiology, and behavior in this group of animals.

Another goal of genomic analysis is to understand the function of sequence features. Gene function can sometimes be inferred by the pattern of transcription. Transcriptomics is the study of gene transcription profiles either qualitatively, to determine which genes are expressed, or quantitatively, to measure changes in the levels of transcription of genes. Proteomics is the study of the entire protein populations of various cell types and tissues and the numerous interactions among proteins. Some proteins, particularly enzymes, are involved in biochemical pathways that produce metabolites for various cellular processes. Metabolomics aims to characterize metabolic pathways by studying the metabolite profiles of cells. All of these "-omic" subdisciplines of genomics use a genome-wide approach to study the function of biological molecules in cells, tissues, or organisms at different developmental stages or under different physiological or environmental conditions.

Transcriptomics

Transcriptomics (gene expression profiling) aims to measure the levels of transcription of genes on a whole-genome basis under a given set of conditions. Transcription may be assessed as a function of medical conditions, as a consequence of mutations, in response to natural or toxic agents, in different cells or tissues, or at different times during biological processes such as cell division or development of an organism. Often, the goal of gene expression studies is to identify the genes that are up- or downregulated in response to a change in a particular condition. High-throughput next-generation RNA sequencing is currently the major experimental approach for measuring RNA transcript levels on a whole-genome basis, although DNA microarrays are used by some researchers and in clinical diagnostic tests (see chapter 4).

RNA Sequencing

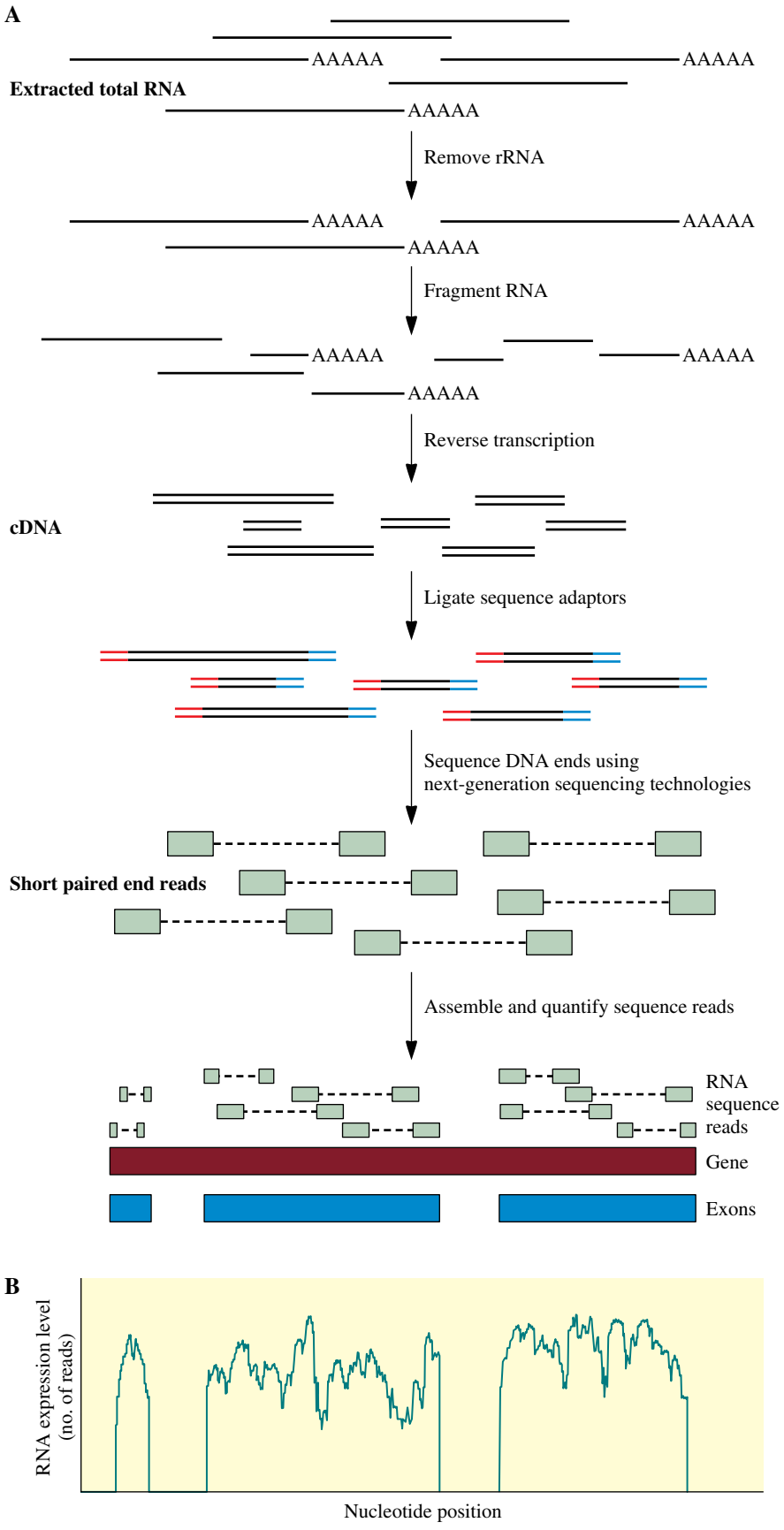
RNA sequencing detects and quantifies the complete set of gene transcripts produced by cells under a given set of conditions. In addition, RNA sequencing can delineate the beginning and end of genes, reveal posttranscriptional modifications such as variations in intron splicing that lead to variant proteins, and identify differences in the nucleotide sequence of a gene among samples. This approach does not require prior knowledge of the genome sequence. Traditionally, RNA sequencing approaches required generating cDNA libraries from isolated mRNA and sequencing the cloned inserts, or the end(s) of the cloned inserts (expressed sequence tags), using the dideoxynucleotide method. New developments in sequencing technologies circumvent the requirement for preparation of a clone library and enable high-throughput sequencing of cDNA or RNA directly.

In an RNA sequencing experiment, total RNA is isolated from two or more sources for which expression profiles are compared. RNA may be

extracted from a population of cells, for example, bacterial cells in culture or specific tissues in multicellular organisms, or from individual cells to assess heterogeneity among cells in a population. The RNA is converted to cDNA using reverse transcriptase and a mixture of oligonucleotide primers composed of six random bases (random hexamers) that bind to multiple sites on all of the template RNA molecules (Fig. 2.40A). Because rRNA makes up a large fraction (>80%) of the total cellular RNA and its levels are not expected to change significantly under different conditions, these molecules are often removed prior to cDNA synthesis by hybridization to complementary oligonucleotides that are covalently linked to magnetic beads for removal. Long RNA molecules are fragmented to pieces of about 200 bp by physical (e.g., nebulization), chemical (e.g., metal ion hydrolysis), or enzymatic (e.g., controlled RNase digestion) methods either before cDNA synthesis (RNA fragmentation) or after cDNA synthesis (cDNA fragmentation).

The cDNA fragments are ligated at one or both ends to an adaptor that serves as a binding site for a sequencing primer (Fig. 2.40A). High-throughput next-generation sequencing technologies are employed to sequence the cDNA fragments. The sequence reads are assembled in a manner similar to that for genomic DNA, which is by aligning the reads to a reference genome or by aligning overlapping sequences to generate contigs for *de novo* assembly when a reference genome is not available. The reads are expected to align uniformly across the transcript (Fig. 2.40A). Gene expression levels are determined by counting the reads that correspond to each nucleotide position in a gene and averaging these across the length of the transcript (Fig. 2.40B). Expression levels are typically normalized between samples by scaling to the total number of reads per sample (e.g., reads/kilobase pair/million reads). Appropriate coverage (i.e., the number of cDNA fragments sequenced) is more difficult to determine for RNA sequencing than for genome sequencing because the total complexity of the transcriptome is not known before the experiment. In general, larger genomes and genomes that have more RNA splicing variants have greater transcriptome complexity and therefore require greater coverage. Also, accurate measurement of transcripts from genes with low expression levels requires sequencing of a greater number of transcripts. Quantification may be confounded by the high GC content of some cDNA fragments which have a higher melting temperature and therefore are inefficiently sequenced, by overrepresentation of cDNA fragments from the 5′ end of transcripts due to the use of random hexamers, and by reads that map to more than one site in a genome due to the presence of repeated sequences. However, because each transcript is represented by many different reads, these biases are expected to have minimal effects on the quantification of a transcript.

Figure 2.40 High-throughput RNA sequencing. **(A)** Total RNA is extracted from a sample, and rRNA may be removed. The RNA is fragmented and then converted to cDNA using reverse transcriptase. Adaptors are added to the ends of the cDNA to provide binding sites for sequencing primers. High-throughput next-generation sequencing technologies are used to determine the sequences at the ends of the cDNA molecules (paired-end reads). The sequence reads are aligned to a reference genome or assembled into contigs using the overlapping sequences. Shown is the alignment of paired-end reads to a gene containing two introns. **(B)** RNA expression levels are determined by counting the reads that correspond to a gene.


A
AAAAA
AAAAA
Extracted total RNA
AAAAA
Remove rRNA
AAAAA
AAAAA
AAAAA
Fragment RNA
AAAAA
AAAAA
AAAAA
Reverse transcription
cDNA
Ligate sequence adaptors
Sequence DNA ends using
next-generation sequencing technologies
Short paired end reads
Assemble and quantify sequence reads
RNA
sequence
reads
Gene
Exons
B
RNA expression level
(no. of reads)
Nucleotide position

Whole-tissue and single-cell RNA sequencing were applied to investigate the mechanism of tissue regeneration in the earthworm *Eisenia andrei*, an animal model for regenerative medicine. The regenerative capacity of most animals is limited; however, earthworms can regenerate entire amputated posterior or anterior body portions, including brain and heart, in 2 weeks. To understand the early events in the regenerative process, total RNA was extracted from wound tissue at the site of head amputation at several time points up to 72 hours postamputation. Sequencing libraries were prepared as described above, and cDNA was sequenced using the Illumina platform to generate 150-bp paired-end reads that were mapped to the *E. andrei* genome sequence. Differentially expressed genes were identified by comparing transcript abundance in tissues at early stages of regeneration to control tissue obtained immediately after amputation. Genes whose expression was significantly upregulated included those involved in energy metabolism, signal transduction, regulation of cell growth, cell differentiation, and programmed cell death. For example, the upregulated *FOS* gene participates in wound response and maintenance of neoblasts (pluripotent stem cells). The types of cells that participate in earthworm head regeneration were determined from characteristic transcription profiles of single cells. Cells in wound tissue were separated using collagenase, and then single cells were encapsulated in a droplet of oil with a gel bead containing enzymes and primers for generation of cDNA (Fig. 2.41A). Importantly, the primers also contain unique sequences to enable sequencing reads to be assigned to individual cells. The cDNAs were amplified, fragmented, and sequenced using Illumina sequencing. Transcriptome data are often depicted graphically in a heat map in which gene expression levels are represented as variations in color intensity. Genes whose expression patterns are similar among samples are clustered to facilitate predictions of their cofunction in a pathway or process. In this example, genes that are markers of a given cell type are clustered to aid cell identification (Fig. 2.41B). Pluripotent stem cells, and to a lesser extent muscle and neuronal cells, were the dominant cell types in the regenerating tissue.

DNA Microarrays

A DNA microarray (DNA chip or gene chip) experiment consists of hybridizing a nucleic acid sample (target) derived from the mRNAs of a cell or tissue to single-stranded DNA sequences (probes) that are arrayed on a solid platform. Depending on the purpose of the experiment, the probes on a microarray may represent an entire genome, a single chromosome, selected genomic regions, or selected coding regions from one or several different organisms. Some DNA microarrays contain sets of oligonucleotides as probes, usually representing thousands of different genes, that are synthesized directly on a solid surface. Thousands of copies of an oligonucleotide with the same specific nucleotide sequence are synthesized in a predefined position on the array surface (probe cell). The probes are typically 20 to 70 nucleotides, although longer probes can also be used, and several probes with different sequences for each gene are usually present on the microarray to minimize errors. Probes are designed to be specific for their target sequences, to avoid hybridization with nontarget sequences, and to have similar melting (annealing) temperatures so that all target sequences can bind to their complementary probe sequence under

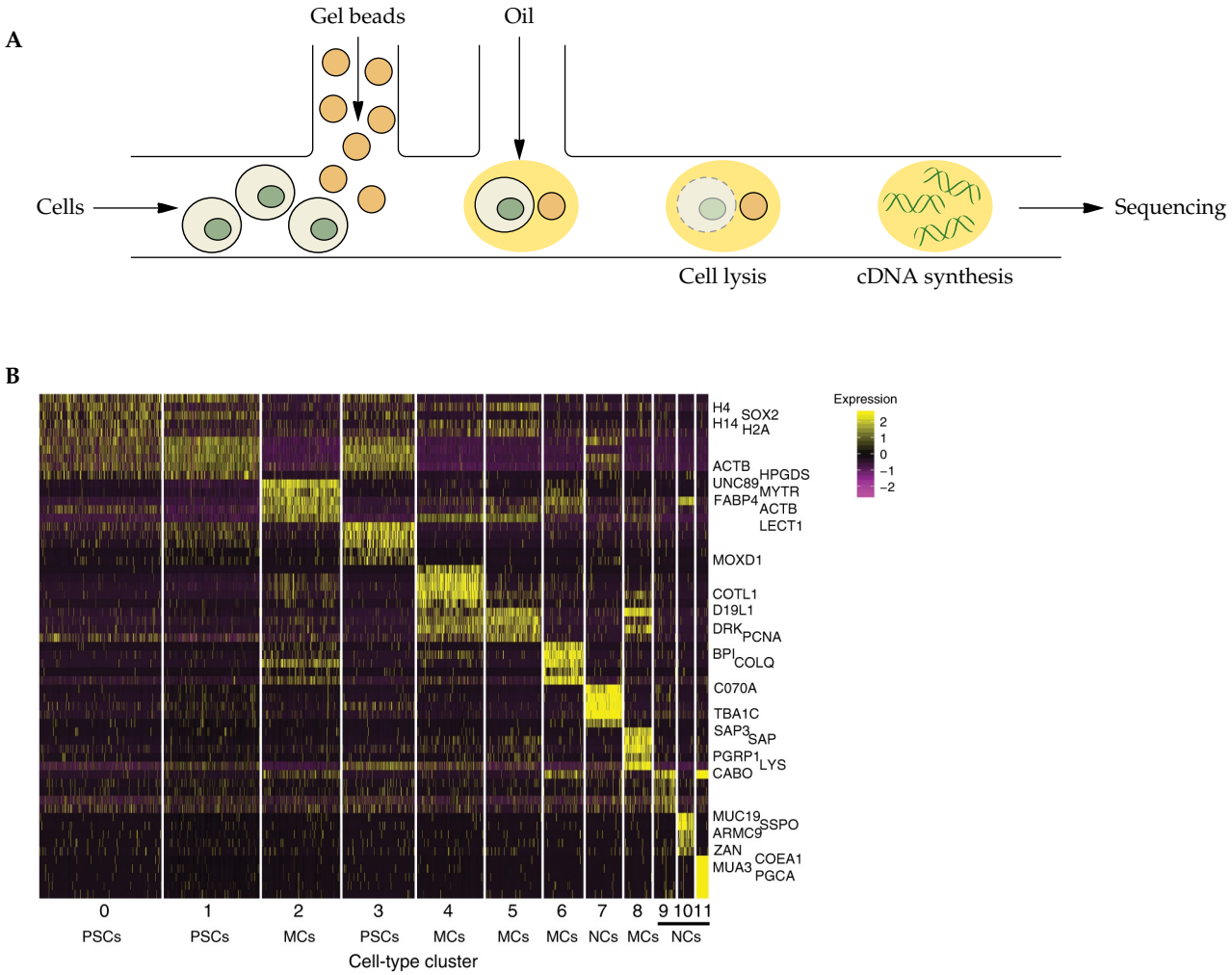


Figure 2.41 Single-cell RNA sequencing. **(A)** Cells in a tissue are dissociated, and a single cell is encapsulated in a droplet of oil with a gel bead containing primers and reverse transcriptase for cDNA synthesis, often within a microfluidic device as depicted here. In addition to sequences to prime cDNA synthesis, such as poly(T) that binds to the poly(A) tails of eukaryotic mRNA, the primers may contain unique identifier sequences, for example, to indicate the origin of the cell. Within the oil droplet, the cells are lysed and cDNA is synthesized from the released RNA. The cDNA can then be extracted from the oil, amplified, and sequenced as described in Fig. 2.40. **(B)** Heat map of gene expression in individual cells in regenerating earthworm tissue. Each row shows the expression level of a specific gene (listed on the right) in the cells. Highly expressed genes are indicated in yellow, and genes with low levels of expression are indicated in pink. Five genes with the greatest fold change in expression for each cell type are shown. PSCs, pluripotent stem cells; MCs, muscle cells; NCs, neuronal cells. Reprinted from Shao et al., *Nat. Commun.* **11:**2656, 2020, under license CC BY 4.0.

the same conditions. A complete whole-genome oligonucleotide array may contain more than 500,000 probes representing as many as 30,000 genes.

For most gene expression profiling experiments that utilize microarrays, mRNA is extracted from cells or tissues and used as a template

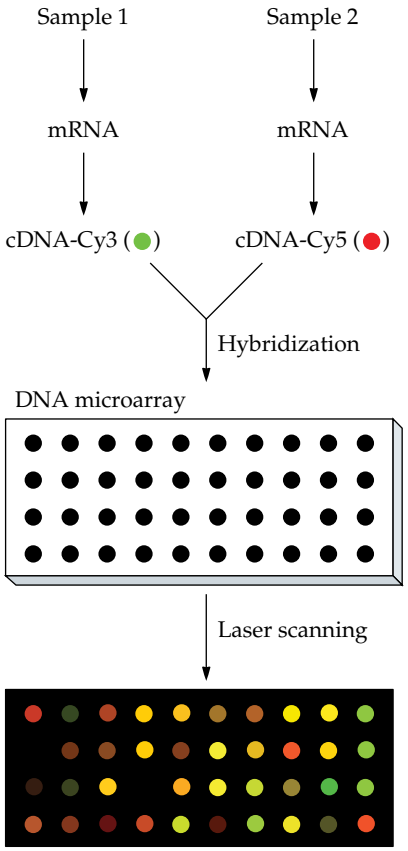


Figure 2.42 Gene expression profiling with a DNA microarray. The mRNA is extracted from two samples (sample 1 and sample 2), and during reverse transcription, the cDNA strands are labeled with the fluorescent dyes Cy3 and Cy5, respectively. The cDNA samples are mixed and hybridized to an ordered array of either gene sequences or gene-specific oligonucleotides. After the hybridization reaction, each probe cell is scanned for both fluorescent dyes and the separate emissions are recorded. Probe cells that produce only a green or red emission represent genes that are transcribed only in sample 1 or 2, respectively; yellow emissions indicate genes that are active in both samples, and the absence of emissions (black) represents genes that are not transcribed in either sample.

to synthesize cDNA using reverse transcriptase. Usually, mRNA is extracted from two or more sources for which expression profiles are compared, for example, from diseased versus normal tissue, or from cells grown under different conditions (Fig. 2.42). The cDNA from each source is labeled with a different fluorophore by incorporating fluorescently labeled nucleotides during cDNA synthesis. For example, a green-emitting fluorescent dye (Cy3) may be used for the normal (reference) sample, and a red-emitting fluorescent dye (Cy5), for the test sample. After labeling, the cDNA samples are mixed and hybridized to the same microarray. Replicate samples are independently prepared under the same conditions and hybridized to different microarrays. A laser scanner determines the intensities of Cy5 and Cy3 for each probe cell on a microarray. The ratio of red (Cy5) to green (Cy3) fluorescence intensity of a probe cell indicates the relative expression levels of the represented gene in the two samples. To avoid variation due to inherent and sequence-specific differences in labeling efficiencies between Cy3 and Cy5, reference and test samples are often reversed labeled and hybridized to another microarray. Alternatively, for some microarray platforms, the target sequences from reference and test samples are labeled with the same fluorescent dye and are hybridized to different microarrays. Methods to calibrate the data among microarrays in an experiment include using the fluorescence intensity of a gene that is not differentially expressed among different conditions as a reference point (i.e., a housekeeping gene), including spiked control sequences that are sufficiently different from the target sequences and therefore bind only to a corresponding control probe cell, and adjusting the total fluorescence intensities of all genes on each microarray to similar values under the assumption that a relatively small number of genes are expected to change among samples.

Genes whose expression changes in response to a particular biological condition are identified by comparing the fluorescence intensities for each gene, averaged among replicates, under two different conditions. The raw data of the fluorescence emissions of each gene are converted to a ratio, commonly expressed as fold change. Generally, positive ratios represent greater expression of the gene in the test sample than of that in the reference sample. Negative values indicate a lower level of expression in the test sample relative to the reference sample.

Proteomics

Proteins are the molecular machines of cells. They catalyze biochemical reactions, monitor the internal and external environments of the cell and mediate responses to perturbations, and are a major component of cell structures such as the cytoskeleton and extracellular matrix. Some proteins are present at more or less the same levels in all cells of a multicellular individual or a population of unicellular organisms under most conditions, for example, a subunit of RNA polymerase which is required for transcription. The levels of other proteins differ among cells according to the cells' functions or change in response to developmental or environmental cues. Thus, analysis of the proteins that are present under particular biological conditions can provide insight into the activities of a cell or tissue.

Proteomics is the comprehensive study of all the proteins of a cell, tissue, body fluid, or organism from a variety of perspectives, including structure, function, expression profiling, and protein–protein interactions. There are several advantages to studying the protein complement (proteome) of cells or tissues compared to other genomic approaches. Although analysis of genomic sequences can often identify protein-coding sequences, in many cases the function of a protein, and the posttranslational modifications that influence protein activity and cellular localization, cannot be predicted from the sequence. On the other hand, it may be possible to infer a protein's function by determining the conditions under which it is expressed and active. While expression profiles of protein-coding sequences can be determined using transcriptomics, mRNA levels do not always correlate with protein levels and do not indicate the presence of active proteins, and interactions between proteins cannot be assessed by these methods. Generally, mRNA is turned over rapidly, and therefore, transcriptomics measures actively transcribed genes, whereas proteomics monitors relatively more stable proteins. From a practical standpoint, proteomics may be used to identify proteins associated with a clinical disorder (protein biomarkers), especially in the early stages of disease development, that can aid in disease diagnosis or provide targets for treatment of disease.

Identification of Proteins

A cell produces a large number of different proteins that may first be separated to identify individual components of the proteome. To reduce the complexity, proteins are sometimes extracted from particular subcellular locations such as the cell membrane, nucleus, Golgi apparatus, endosomes, or mitochondria. Extracted proteins are typically separated based on differences in their molecular mass and/or net charge which determine the rate at which they move through a matrix in an electrical field. Two-dimensional polyacrylamide gel electrophoresis (2D PAGE) is an effective method to separate proteins in a sample. Proteins are first separated on the basis of their net charge by electrophoresis through an immobilized pH gradient in one dimension (the first dimension) (Fig. 2.43A). Some amino acids in a polypeptide have side chains with ionizable groups that contribute to the net charge of a protein; the degree of ionization (protonation) is influenced by the pH of the solution. In a gel to which an electric current is applied, proteins migrate through a pH gradient until they reach a specific pH (the isoelectric point) where the overall charge of the protein is zero and they no longer continue to migrate. A particular position in the pH gradient may be occupied by two or more different proteins that have the same isoelectric point. However, proteins with identical isoelectric points may have different molecular weights and can be further separated according to their molecular mass by electrophoresis at right angles to the first dimension (the second dimension) through a sodium dodecyl sulfate (SDS)-polyacrylamide gel (Fig. 2.43B). The separated proteins form an array of spots in the gel that is visualized using Coomassie blue, silver, or fluorescent protein stains.

Depending on the size of the two-dimensional polyacrylamide gel and the abundance of individual proteins, approximately 2,000 different

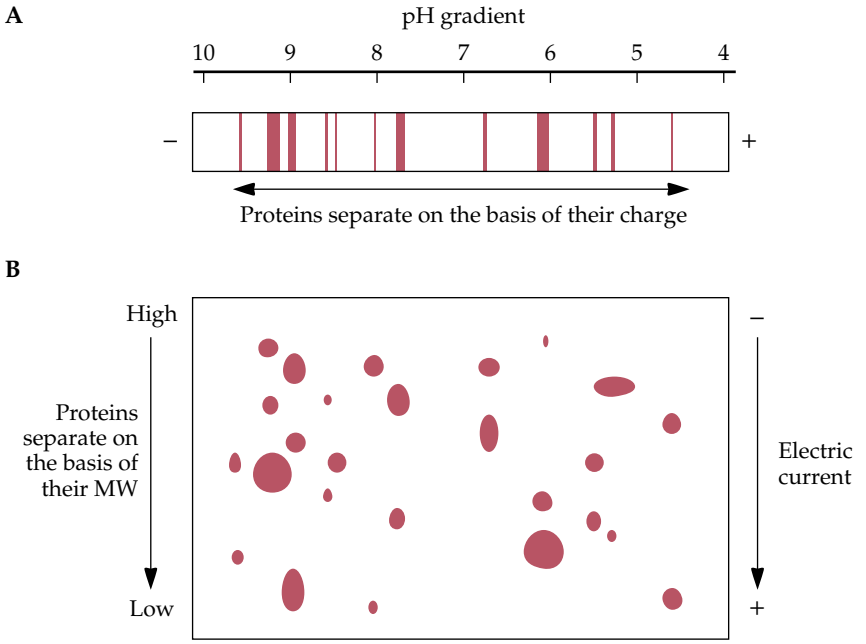


Figure 2.43 2D PAGE for separation of proteins. **(A)** First dimension. Isoelectric focusing is performed to first separate proteins in a mixture on the basis of their net charge. The protein mixture is applied to a pH gradient gel. When an electric current is applied, proteins will migrate toward either the anode (+) or cathode (–) depending on their net charge. As proteins move through the pH gradient, they will gain or lose protons until they reach a point in the gel where their net charge is zero. The pH in this position of the gel is known as the isoelectric point and is characteristic of a given protein. At that point, a protein no longer moves in the electric current. **(B)** Second dimension. Several proteins in a sample may have the same isoelectric point and therefore migrate to the same position in the gel in the first dimension. Therefore, proteins are further separated on the basis of differences in their molecular weights (MW) by electrophoresis, at a right angle to the first dimension, through a sodium dodecyl sulfate-polyacrylamide gel.

proteins can be resolved. The pattern of spots is captured by densitometric scanning of the gel. Databases have been established with images of two-dimensional polyacrylamide gels from some different cell types, and software is available for detecting spots, matching patterns between gels, and quantifying the protein content of the spots. Proteins with either low or high molecular weights, those with highly acidic or basic isoelectric points (such as ribosomal proteins and histones), and those that are found in cellular membranes are not readily resolved by 2D PAGE. Proteins that are present in relatively small amounts are not readily detected.

After separation, individual proteins are excised from the gel, and the identity of the protein is determined, usually by mass spectrometry (MS). A mass spectrometer detects the masses of the ionized form of a molecule. For identification, the protein is first fragmented into peptides by digestion with a protease, such as trypsin that cleaves at lysine and arginine residues (Fig. 2.44). The peptides are ionized and separated according to their mass-to-charge (m/z) ratio, and then the abundance and m/z ratios of the ions are measured. Several mass spectrometers are available that differ in the type of sample analyzed, the mode of ionization of the sample, the method for

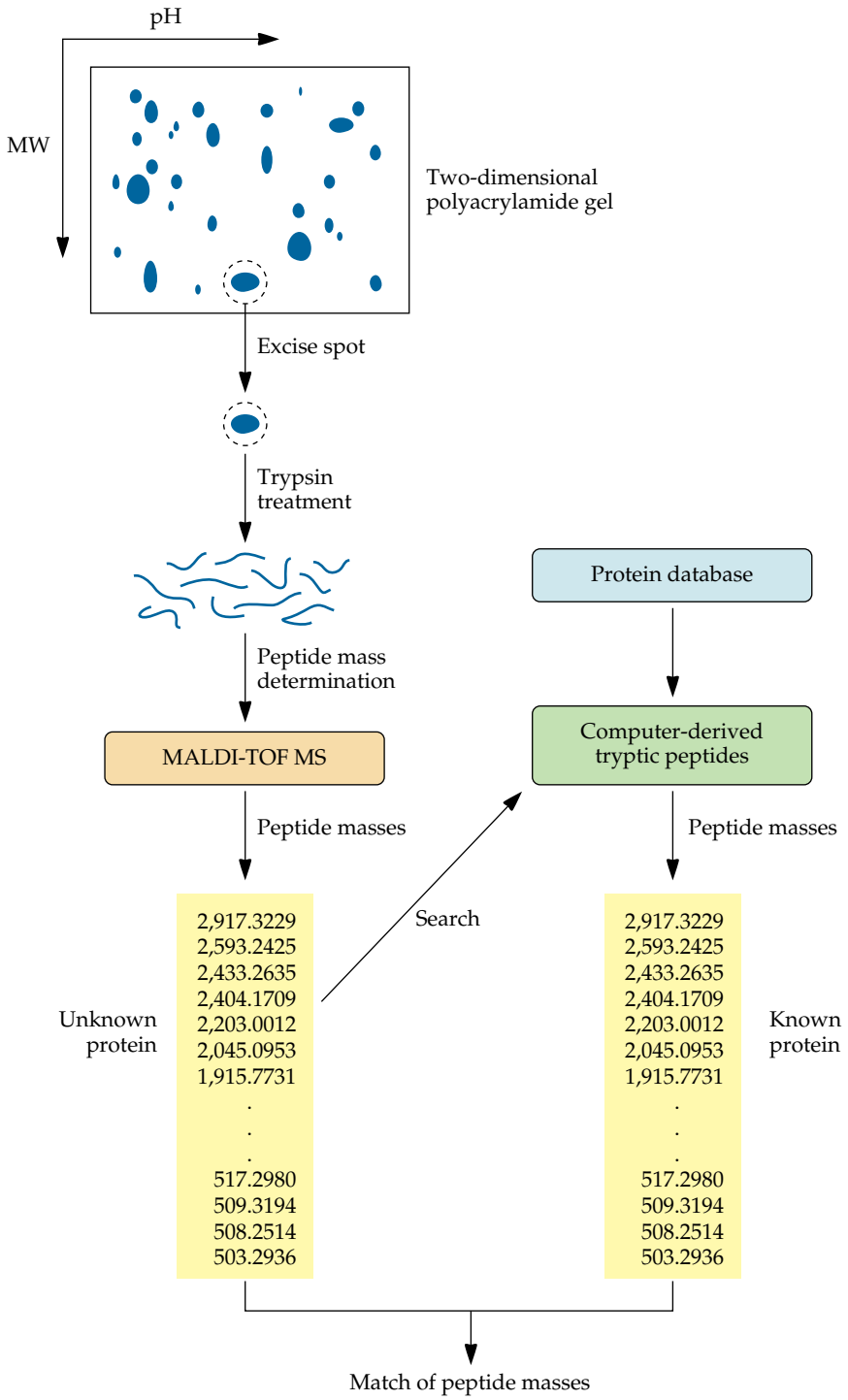


Figure 2.44 Peptide mass fingerprinting. A spot containing an unknown protein that was separated by 2D PAGE is excised from the gel and treated with trypsin. Purified trypsin peptides are separated by MALDI-TOF MS. The set of peptide masses from the unknown protein are used to search a database that contains the masses of tryptic peptides for every known sequenced protein, and the best match is determined. The trypsin cleavage sites of known proteins are determined from the amino acid sequence and, consequently, the masses of the tryptic peptides are easy to calculate. Only some of the tryptic peptide masses for the unknown protein are listed in this example.

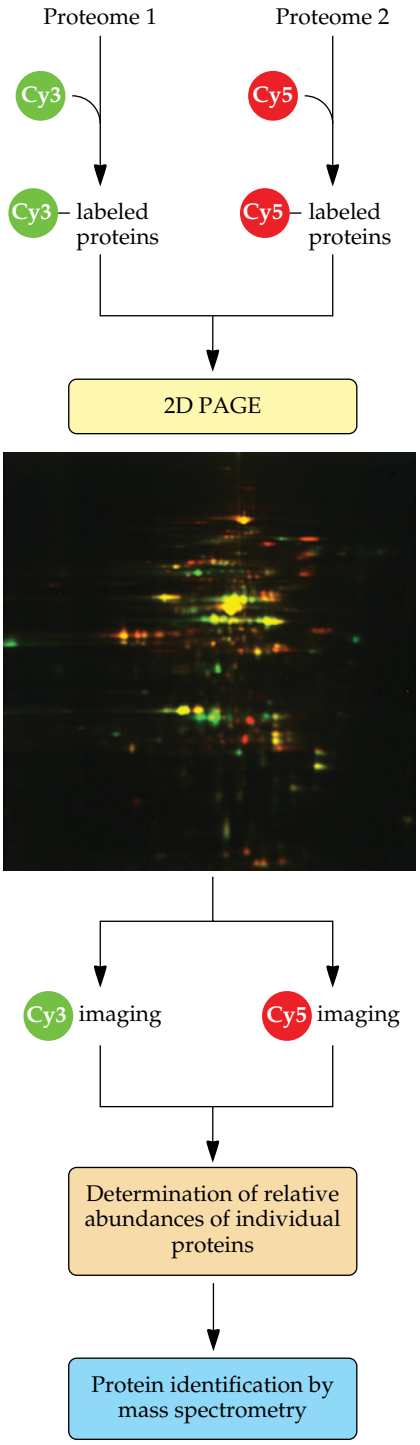


Figure 2.45 Protein expression profiling using 2D differential in-gel electrophoresis. The proteins of two different proteomes are labeled with fluorescent dyes Cy3 and Cy5, respectively. The labeled proteins from the two samples are combined and separated by 2D PAGE. The gel is scanned for each fluorescent dye, and the relative levels of the two dyes in each protein spot are recorded. Each spot with an unknown protein is excised for identification by MS.

generating the electromagnetic field that separates and sorts the ions, and the method of detecting the different masses. Peptide masses are usually determined by matrix-assisted laser desorption ionization–time of flight (MALDI-TOF) MS. To determine the *m/z* value of each peptide fragment generated from an excised protein by MALDI-TOF MS, the peptides are ionized by mixing them with a matrix consisting of an organic acid and then using a laser to promote ionization. The ions are accelerated through a tube using a high-voltage current, and the time required to reach the ion detector is determined by their molecular mass, with lower-mass ions reaching the detector first.

To facilitate protein identification, computer algorithms have been developed for processing large amounts of MS data. Databases have been established that contain the theoretical masses of tryptic peptides for all proteins encoded by a genome. The databases are searched to identify a protein whose peptide masses match the values of the peptide masses of an unknown protein that were determined by MALDI-TOF MS (Fig. 2.44). Each identified protein is ranked by a score representing the probability that the match is false. This type of analysis is called peptide mass fingerprinting. It is not always necessary to isolate a protein before identification by MS. All of the proteins in a complex mixture extracted from cells or tissues may be digested with a protease in solution, and then the masses of thousands of peptides are determined in a single MS run. Peptide mass fingerprinting is applied to identify as many of the proteins as possible, although the false identification rate can be high.

Protein Expression Profiling

Several methods have been developed to quantitatively compare the proteomes among samples. Two-dimensional differential in-gel electrophoresis is very similar to 2D PAGE; however, rather than separating proteins from different samples on individual gels and then comparing the maps of separated protein, proteins from two different samples are differentially labeled and then separated on the same two-dimensional polyacrylamide gel (Fig. 2.45). Typically, proteins from each sample are labeled with different fluorescent dyes, such as the green-emitting dye Cy3 and the red-emitting dye Cy5, which have higher sensitivity than many other protein stains. Then, the labeled samples are mixed and run together in the same gel, which overcomes the variability between separate gel runs. The two dyes carry the same mass and charge, and therefore, a protein labeled with Cy3 migrates to the same position as the identical protein labeled with Cy5. The Cy3 and Cy5 protein patterns are visualized separately by fluorescent excitation, the images are compared, and any differences are recorded. The ratio of Cy3 to Cy5 fluorescence intensity for each

spot is determined to detect proteins that are either up- or downregulated. Unknown proteins are identified by MS analysis.

Another powerful technique for comparing protein populations among samples utilizes protein microarrays. Protein microarrays consist of large numbers of proteins immobilized in a known position on a surface such as a glass slide in a manner that preserves the structure and function of the proteins. The proteins arrayed on the surface can be antibodies specific for a set of proteins in an organism, purified proteins that were expressed from a DNA or cDNA library, short synthetic peptides, or multiprotein samples from cell lysates or tissue specimens. The arrayed proteins are probed with samples that contain molecules that interact with the proteins. For example, the interacting molecules can be other proteins to detect protein–protein interactions, nucleic acid sequences to identify proteins that regulate gene expression by binding to DNA or RNA, substrates for specific enzymes, or small protein-binding compounds such as lipids or drugs.

Microarrays consisting of immobilized antibodies are used to detect and quantify proteins present in a complex sample. Antibodies directed against more than 1,800 human proteins have been isolated, characterized, and validated, and subsets of these that detect specific groups of proteins can be arrayed. Several antibody microarrays are commercially available to detect changes in cell signaling proteins associated with various cancers and protein biomarkers of some neurological, autoimmune, and infectious diseases. To compare protein profiles in two different samples, for example, in normal and diseased tissues, proteins extracted from the two samples are labeled with two different fluorescent dyes (e.g., Cy3 and Cy5) and then applied to one antibody microarray (Fig. 2.46). Proteins present in the samples bind to their cognate antibodies, and after washing to remove unbound proteins, the antibody-bound proteins are detected with a fluorescence scanner.

Proteins whose expression changes in response to a particular biological condition are identified by comparing the fluorescence intensities for each protein, averaged among replicates, under two different conditions. The ratio of fluorescence intensities, for example, red (Cy5) to green (Cy3) fluorescence, indicates the relative abundance of a protein in two samples and is commonly expressed as the fold change. Positive ratios represent greater abundance of the protein in a test sample than in a reference sample, and negative values indicate a lower level of abundance in a test sample relative to a reference sample.

To increase the sensitivity of the assay and therefore the detection of low-abundance proteins, or to detect a specific subpopulation of proteins, a "sandwich"-style assay is often employed (Fig. 2.47). In this case, unlabeled proteins in a sample are bound to an antibody microarray, and then a second, labeled antibody is applied. This approach has been used to determine whether particular posttranslational protein modifications such as phosphorylation of tyrosine or glycosylation are associated with specific diseases. Serum proteins are first captured by immobilized antibodies on a microarray. Then, an antiphosphotyrosine antibody is applied that binds only to tyrosine phosphorylated proteins (Fig. 2.47A). The antiphosphotyrosine antibody is tagged, for example, with a biotin molecule, and fluorescently labeled streptavidin, which binds specifically to biotin, is added to detect the phosphorylated protein. In a similar manner, glycosylated

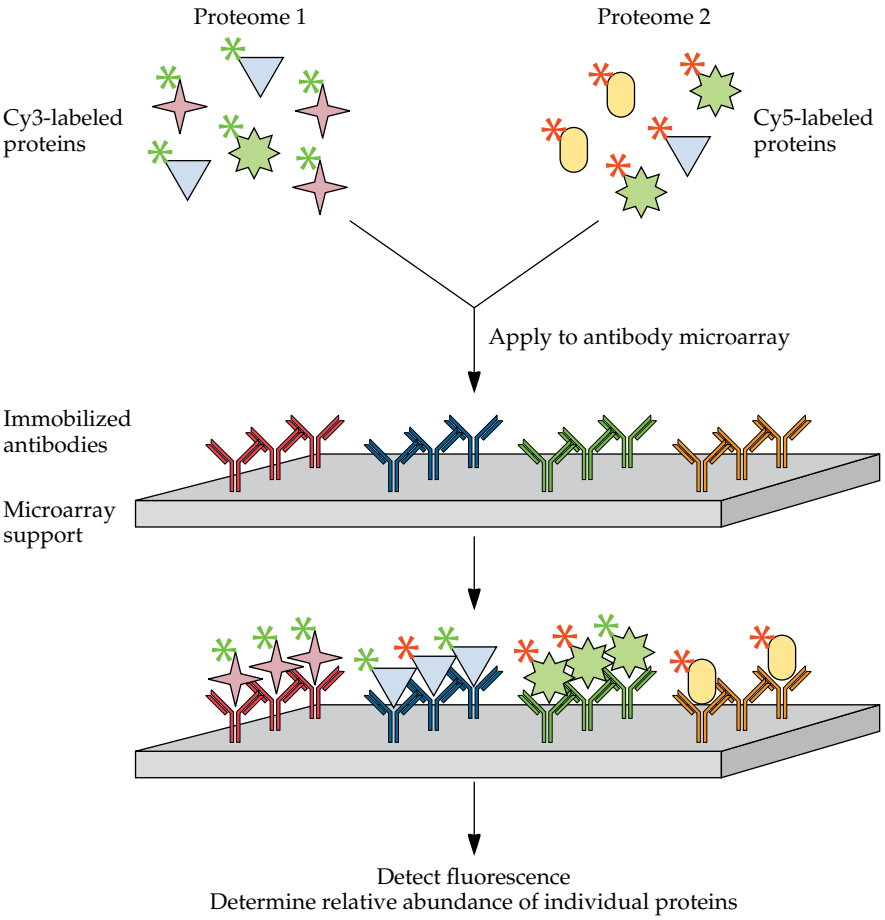


Figure 2.46 Protein expression profiling with an antibody microarray. Proteins extracted from two different samples are labeled with fluorescent dyes Cy3 and Cy5, respectively. The labeled proteins are mixed and incubated with an array of antibodies immobilized on a solid support. Proteins bound to their cognate antibodies are detected by measuring fluorescence, and the relative levels of specific proteins in each sample are determined.

proteins can be detected with lectins (Fig. 2.47B). Lectins are proteins that bind to specific carbohydrate moieties on the surface of proteins or cell membranes, and many different lectins with affinities for different glycosyl groups (glycans) are available.

In another type of microarray, purified proteins representing as many proteins of a proteome under study as possible are arrayed on a solid support and then probed with antibodies in serum samples collected from healthy (control) and diseased individuals. The purpose of these studies is to discover whether individuals produce antibodies that correlate with particular diseases or biological processes. For example, the differential expression of antibodies in serum samples from individuals with and without Alzheimer's disease was tested using a microarray consisting of more than 9,000 unique human proteins (Fig. 2.48). After incubation of the serum samples with the protein microarray, bound antibodies were detected using a fluorescently labeled secondary antibody that interacts specifically

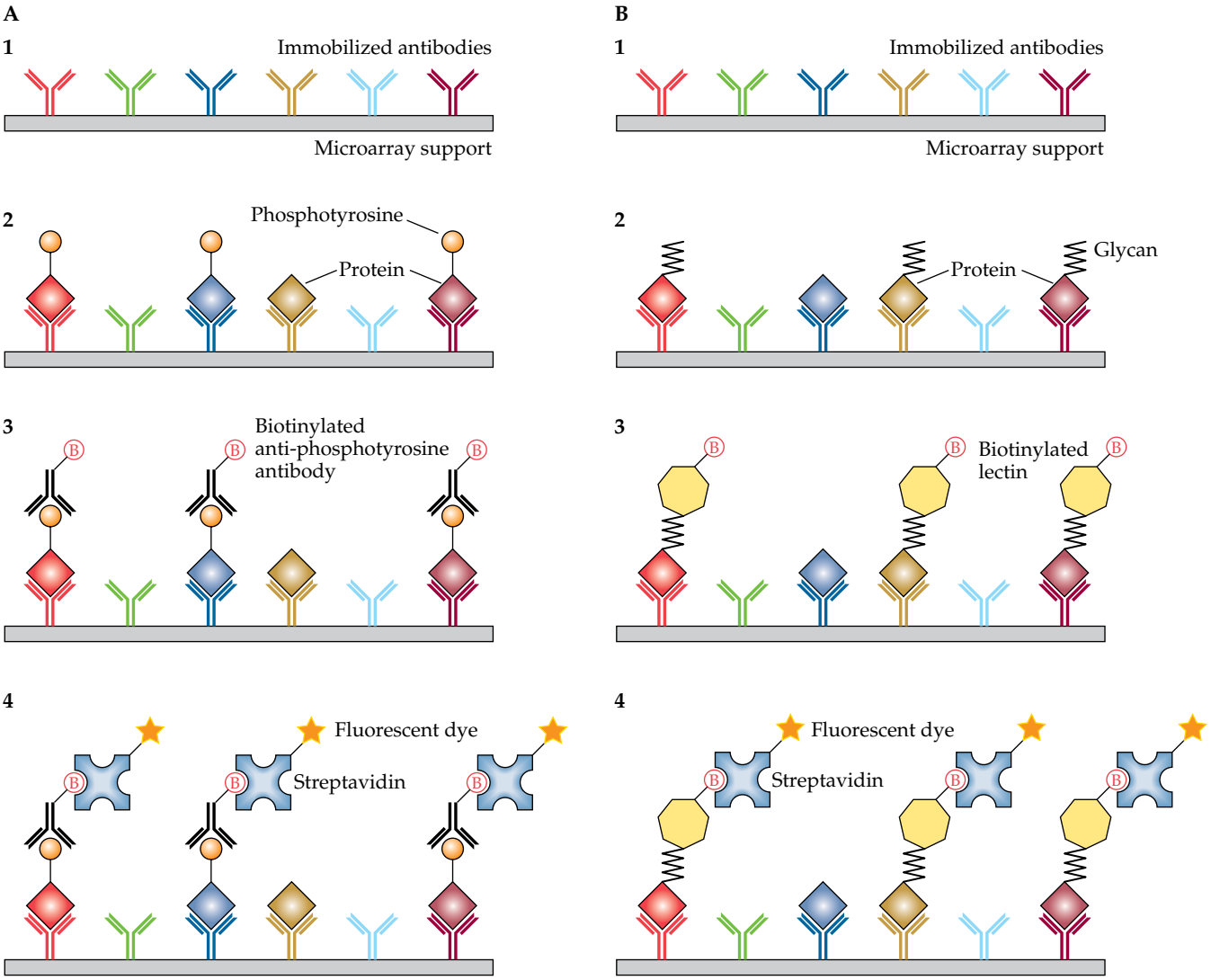


Figure 2.47 Detection of posttranslational modifications with antibody microarrays. **(A)** Detection of tyrosine phosphorylation. An antibody microarray **(1)** is incubated with a protein sample **(2)**. Biotinylated antiphosphotyrosine antibody is added **(3)**, and for visualization, a streptavidin-fluorescent dye conjugate attaches to the biotin of the antiphosphotyrosine antibody **(4)**. **(B)** Detection of glycan groups. An antibody microarray **(1)** is incubated with a protein sample **(2)**. A biotinylated molecule (e.g., lectin) that binds to a specific glycan is added **(3)** and, for visualization, a streptavidin-fluorescent dye conjugate attaches to the biotin of the lectin **(4)**.

with human antibodies. The screen resulted in the identification of 10 autoantibodies (i.e., directed against an individual's own protein) that may be used as biomarkers to diagnose Alzheimer's disease. Protein microarrays can also be used to identify proteins that interact with therapeutic drugs or other small molecules (Fig. 2.49). This can aid in determining the mechanism of action of a drug, for assessing responsiveness among various forms of a target protein (e.g., variants produced by different individuals), and for predicting undesirable side effects.

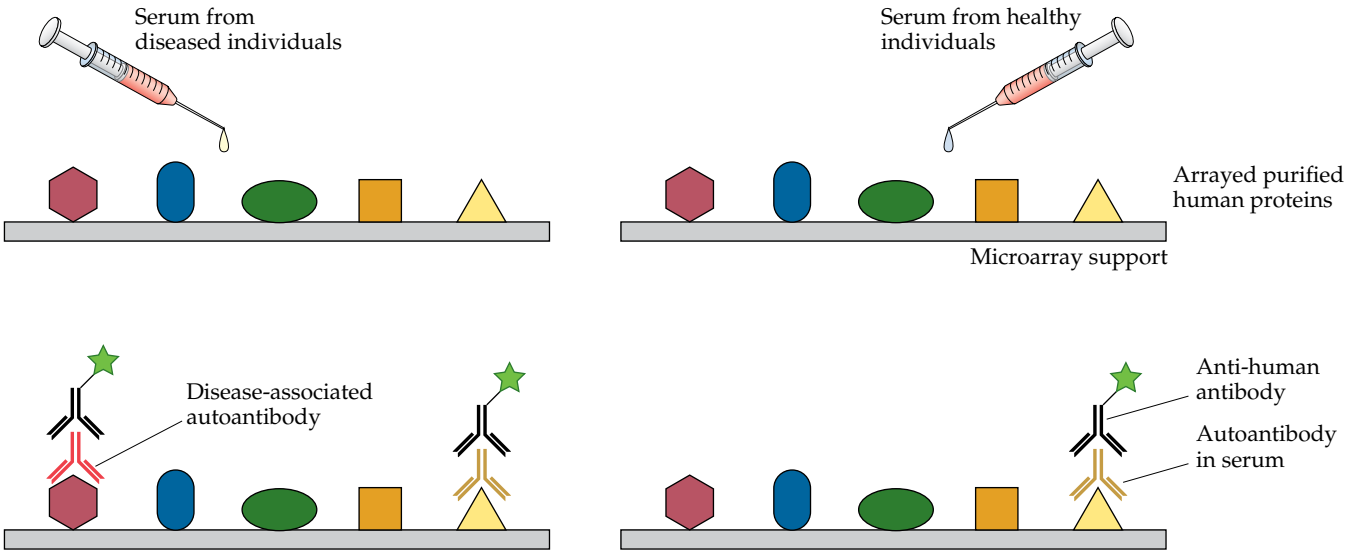


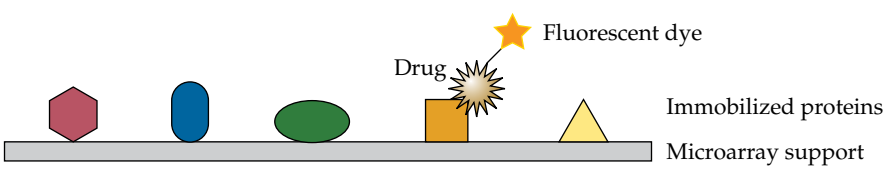
Figure 2.48 Identification of disease biomarkers with a human protein microarray. Serum samples are collected from diseased and healthy individuals and incubated with microarrays of purified human proteins. Serum autoantibodies bind to specific proteins on the microarray and are detected by applying a fluorescently labeled secondary antibody directed against human antibodies. Autoantibodies present in the serum from diseased individuals but not in serum from healthy individuals are potential biomarkers that can be used in diagnosis of the disease.

Protein–Protein Interactions

Proteins typically function as complexes comprising different interacting protein subunits. Important cellular processes such as DNA replication, energy metabolism, and signal transduction are carried out by large multiprotein complexes. Thousands of protein–protein interactions occur in a cell. Some of these are short-lived, while others form stable multicomponent complexes that may interact with other complexes. Determining the functional interconnections among the members of a proteome is not an easy task. Several strategies have been developed to examine protein interactions, including protein microarrays, two-hybrid systems, and affinity purification methods.

The two-hybrid method is generally used to test hypotheses regarding the pair-wise interactions between proteins. The underlying principle of

Figure 2.49 Protein microarrays to detect protein-drug interactions. Therapeutic drugs or other small molecules tagged with a fluorescent dye are applied to purified proteins arrayed on a solid support.



this assay is that the physical connection between two proteins reconstitutes an active transcription factor that initiates the expression of a reporter gene. The transcription factors employed for this purpose have two domains. One domain (DNA-binding domain) binds to a specific DNA site, and the other domain (activation domain) activates transcription (Fig. 2.50A). The two domains are not required to be part of the same protein to function as an effective transcription factor. However, the activation domain alone will not bind to RNA polymerase to activate transcription. Connection with the DNA-binding domain is necessary to place the activation domain in the correct orientation and location to initiate transcription of the reporter gene by RNA polymerase.

For a two-hybrid assay, the coding sequences of the DNA-binding and activation domains of a specific transcription factor are cloned into separate vectors. Often, the Gal4 transcriptional factor from *Saccharomyces cerevisiae* or the bacterial LexA transcription factor is used. A DNA (or cDNA generated from a eukaryotic mRNA) sequence that is cloned in frame with the DNA-binding domain sequence produces a fusion (hybrid) protein and is referred to as the "bait." This is the target protein for which interacting proteins are to be identified. Another DNA sequence is cloned into another vector in frame with the activation domain coding sequence. A protein attached to the activation domain is called the "prey" and potentially interacts with the bait protein. Host yeast cells are transformed with both bait and prey DNA constructs. After expression of the fusion proteins, if the bait and prey do not interact, then there is no transcription of the reporter gene (Fig. 2.50B). However, if the bait and prey proteins interact, then the DNA-binding and activation domains are also brought together. This enables the activation domain to make contact with RNA polymerase and activate transcription of the reporter gene (Fig. 2.50C). The product of an active reporter gene may produce a colorimetric response or may allow a host cell to proliferate in a specific medium.

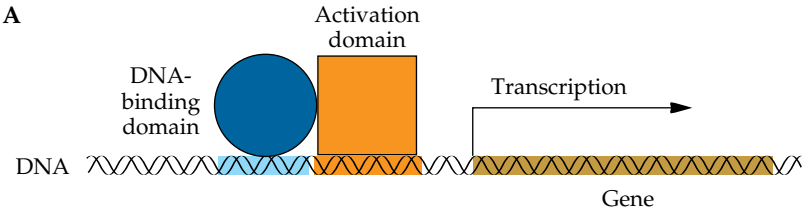
For a whole-proteome protein interaction study, two libraries are prepared, each containing thousands of cDNAs generated from total cellular mRNA (or genomic DNA fragments in a study of proteins from a prokaryote). To construct the bait library, cDNAs are cloned into the vector adjacent to the DNA sequence for the DNA-binding domain of the transcription factor Gal4 and then introduced into yeast cells. To construct the prey library, the cDNAs are cloned into the vector containing the sequence for the activation domain, and the constructs are transferred to yeast cells. The libraries are typically screened for bait-prey protein interactions in one of two ways. In one method, a prey library of yeast cells is arrayed on a grid. The prey library is then screened for the production of proteins that interact with a bait protein by introducing individual bait constructs to the arrayed clones by mating (Fig. 2.51A). Alternatively, each yeast clone in a bait library is mated *en masse* with a mixture of strains in the prey library, and then positive interactions are identified by screening colonies on plates for activation of the reporter gene (Fig. 2.51B). Challenges with using the two-hybrid system for large-scale determination of protein–protein interactions include the inability to clone all possible protein-coding genes in frame with the activation and DNA-binding domains, which leads to missed interactions (false negatives), and the detection

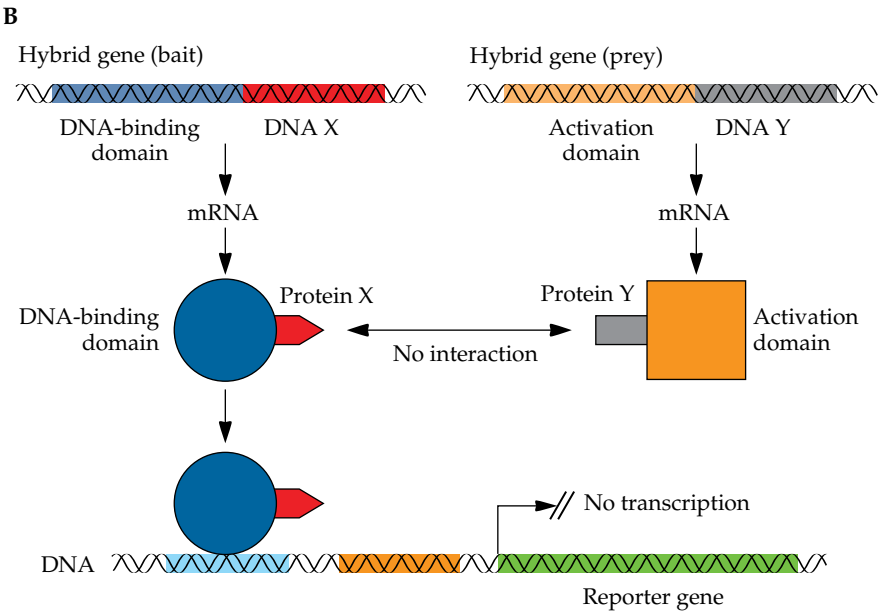
of interactions that do not normally occur in their natural environments within the original cells and therefore are not biologically relevant (false positives). Nonetheless, this approach has been used to successfully identify interacting proteins in a wide range of organisms from bacteria to humans.

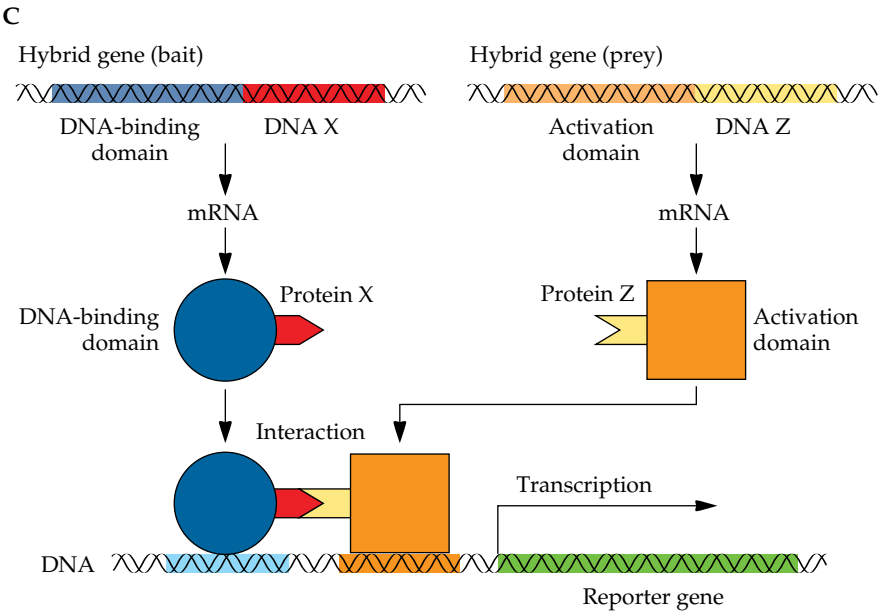
While the two-hybrid assay detects pairwise protein interactions, multiprotein complexes can be captured using affinity purification procedures. A bait protein for which interacting proteins are to be identified is expressed as a fusion protein with a small peptide (affinity tag). The peptide binds with a high affinity to a specific ligand molecule and facilitates purification of the bait protein together with interacting proteins that are subsequently identified via MS. The affinity tag may be an antigen that is recognized by a specific antibody to immunoprecipitate (pull down) the fused bait protein with partner proteins, or another peptide such as polyhistidine or calmodulin-binding peptide that captures the bait protein on a solid support via nickel cations or calmodulin-Ca^{2+}, respectively, during the purification step. The tags are tested to ensure that they do not alter the function of the bait protein. The underlying assumption is that the cellular proteins that normally interact with the native protein *in vivo* will also combine with the tagged protein.

A tandem affinity tag system allows two successive rounds of affinity binding to ensure that the target and its associated proteins are free of any nonspecific proteins. A DNA (or cDNA) sequence that encodes the bait protein is fused to a DNA sequence that encodes two small peptide tags separated by a protease cleavage site. The construct is introduced into a host cell, where it is expressed, and a tagged protein is synthesized (Fig. 2.52). After the cells are lysed, the tagged protein and any interacting proteins are purified using the affinity tags. The proteins of the complex are separated according to their molecular weight by PAGE and identified with MS. Computer programs are available for generating maps of complexes with common proteins, assigning proteins with shared interrelationships to specific cellular activities, and establishing the links between multiprotein complexes.

Figure 2.50 Two-hybrid assay for detecting pairwise protein interactions. **(A)** The DNA-binding domain of a transcription factor binds to a specific sequence in the regulatory region of a gene which orients and localizes the activation domain that is required for the initiation of transcription of the gene by RNA polymerase. **(B)** The coding sequences for the DNA-binding domain and the activation domain are fused to DNA X and DNA Y, respectively, and both constructs (hybrid genes) are introduced into a cell. After translation, the DNA-binding domain–protein X fusion protein binds to the regulatory sequence of a reporter gene. However, protein Y (prey) does not interact with protein X (bait), and the reporter gene is not transcribed because the activation domain does not, on its own, associate with RNA polymerase. **(C)** The coding sequence for the activation domain is fused to the DNA for protein Z (DNA Z) and transformed into a cell containing the DNA-binding domain–DNA X fusion construct. The proteins encoded by the hybrid genes interact, and the activation domain is properly oriented to initiate transcription of the reporter gene demonstrating a specific protein-protein interaction.


A
Activation domain
DNA-binding domain
Transcription
DNA
Gene


B
Hybrid gene (bait)
Hybrid gene (prey)
DNA-binding domain
DNA X
Activation domain
DNA Y
mRNA
mRNA
DNA-binding domain
Protein X
Protein Y
Activation domain
No interaction
No transcription
DNA
Reporter gene


C
Hybrid gene (bait)
Hybrid gene (prey)
DNA-binding domain
DNA X
Activation domain
DNA Z
mRNA
mRNA
DNA-binding domain
Protein X
Protein Z
Activation domain
Interaction
Transcription
DNA
Reporter gene

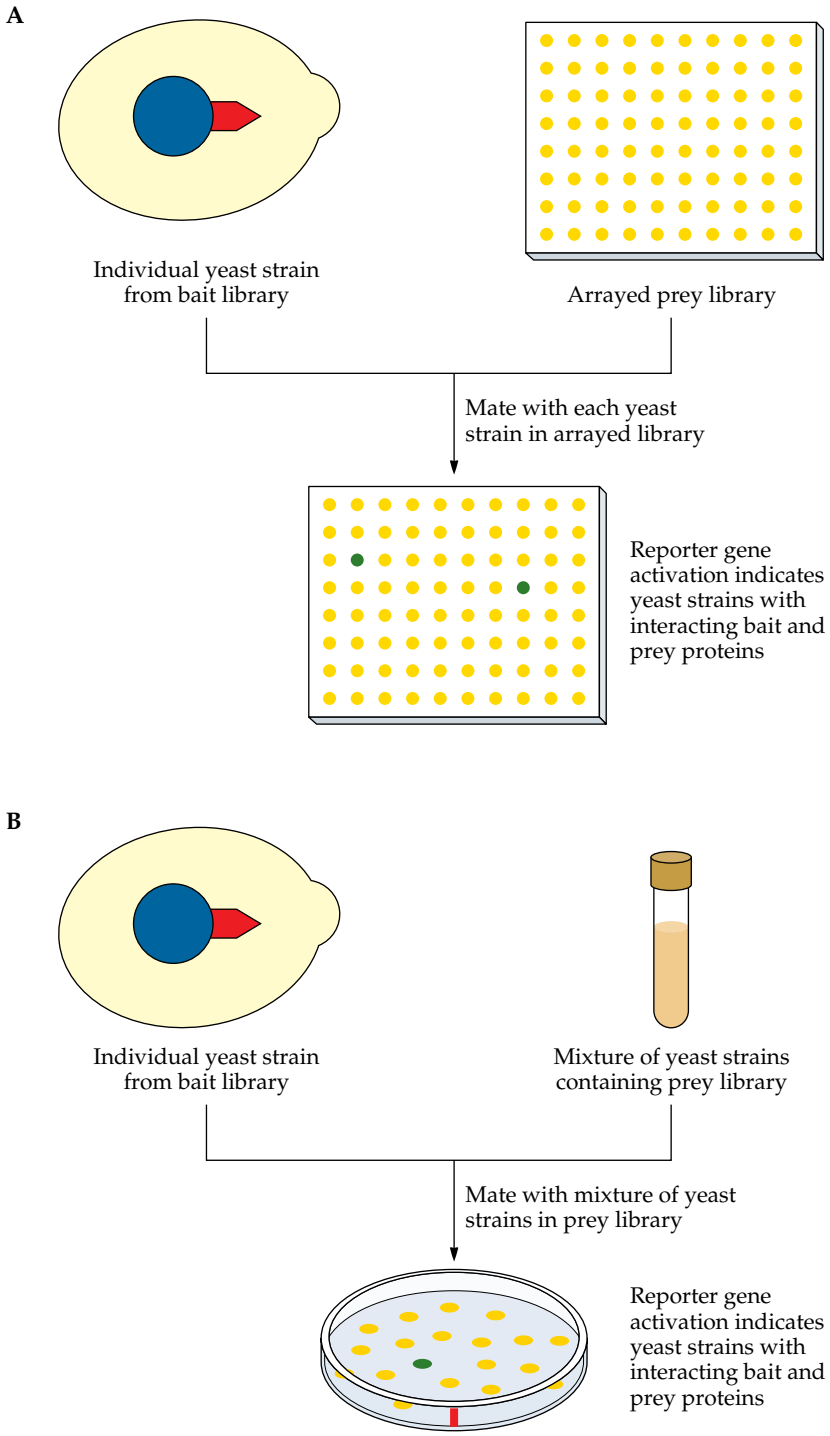


Figure 2.51 Large-scale screens for protein interactions using the yeast two-hybrid system. Two libraries are prepared, one containing genomic DNA fragments fused to the coding sequence for the DNA-binding domain of a transcription factor (bait library) and another containing genomic DNA fragments fused to the coding sequence for the activation domain of the transcription factor (prey library). Two methods are commonly used to screen for pairwise protein interactions. **(A)** Individual yeast strains in the bait library are mated with each yeast strain in an arrayed prey library. Resulting strains in the array that produce bait and prey proteins that interact are detected by assaying for reporter gene activation (cells growing in a multiwell plate that express the reporter gene are indicated in green). **(B)** Yeast strains in the prey library are mated *en masse* with individual strains in the bait library. The mixture of strains is screened for reporter gene activity that identifies strains with interacting bait and prey proteins (green).

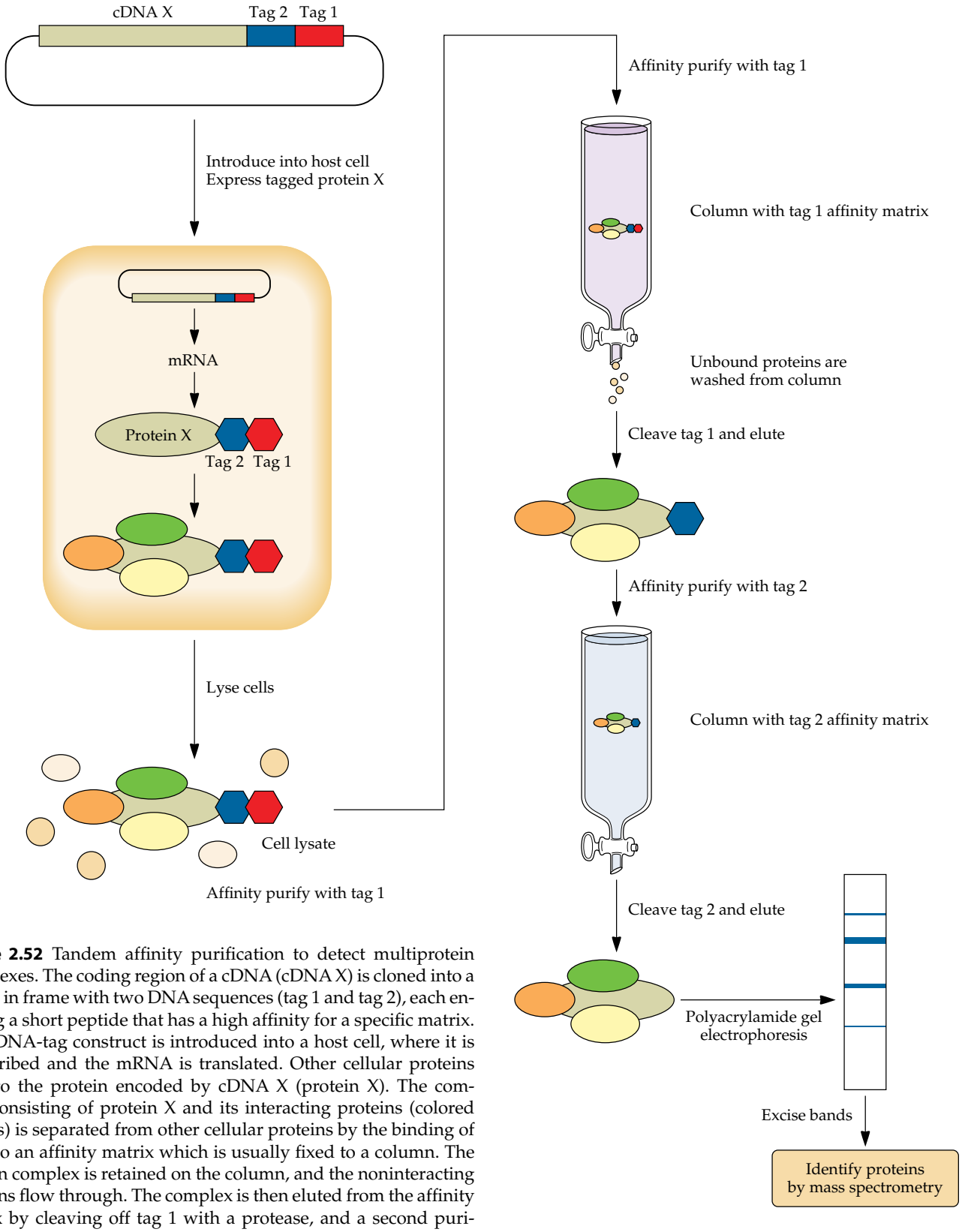


Figure 2.52 Tandem affinity purification to detect multiprotein complexes. The coding region of a cDNA (cDNA X) is cloned into a vector in frame with two DNA sequences (tag 1 and tag 2), each encoding a short peptide that has a high affinity for a specific matrix. The cDNA-tag construct is introduced into a host cell, where it is transcribed and the mRNA is translated. Other cellular proteins bind to the protein encoded by cDNA X (protein X). The complex consisting of protein X and its interacting proteins (colored shapes) is separated from other cellular proteins by the binding of tag 1 to an affinity matrix which is usually fixed to a column. The protein complex is retained on the column, and the noninteracting proteins flow through. The complex is then eluted from the affinity matrix by cleaving off tag 1 with a protease, and a second purification step is carried out with tag 2 and its affinity matrix. The proteins of the complex are separated by one-dimensional PAGE. Single bands are excised from the gel and identified by MS.

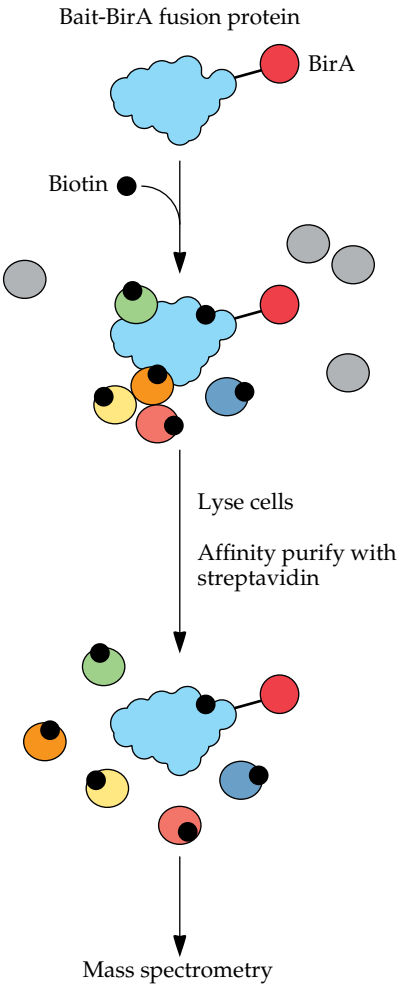


Figure 2.53 Proximity-dependent labeling to identify interacting proteins. A bait protein is expressed as a fusion protein with an enzyme that adds a tag to nearby proteins. One such enzyme, biotin ligase (BirA), has been engineered to catalyze the covalent binding of a biotin molecule to proteins that interact with the bait protein, including those that interact weakly or transiently. Following incubation with biotin, cells expressing the fusion protein are lysed, and biotinylated proteins are captured with beads conjugated to streptavidin, which has a high affinity for biotin. Captured proteins are identified by mass spectrometry.

Although many proteins form stable complexes, others interact only transiently *in vivo*. Examples of transient protein interactions include those involved in intracellular signal transduction and protein trafficking. Weak and transient interactions are not identified by affinity purification methods that require a stringent washing step to remove nonspecific proteins. However, these interactions may be revealed by fusing a bait protein to biotin ligase that catalyzes the addition of biotin to proteins that are in immediate proximity (~10 nm) *in vivo* and then identifying the biotinylated proteins (Fig. 2.53). Typically, a mutant form of the *E. coli* biotin ligase BirA is employed that promiscuously biotinylates proximate proteins. Cells that express the bait-biotin ligase fusion protein are incubated with biotin, and then biotinylated proteins are purified from the cell lysate using magnetic beads conjugated to streptavidin, which binds to biotin with high affinity. Purified proteins are identified via MS. Often, the protocol is tested with biotin ligase fused to the C- or N-terminal ends of the bait protein to ensure that the bait protein is functional. Control experiments with cells that express biotin ligase without the bait protein are conducted to identify the small number of endogenous substrates that are biotinylated. False-positive interactions may be identified due to random protein collisions, although these can be reduced by fusing biotin ligase to other proteins that may interact in the pathway and repeating the experiment.

Metabolomics

Metabolomics is a technique that provides a snapshot of the small molecules present in a complex biological sample. The metabolites present in cells and cell secretions are influenced by genotype, which determines the metabolic capabilities of an organism, and by environmental conditions such as the availability of nutrients and the presence of toxins or other stressors. Metabolite composition varies depending on the developmental and health status of an organism, and therefore, a comprehensive metabolite profile can identify molecules that reflect a particular physiological state. For example, metabolites present in diseased cells but not in healthy cells are useful biomarkers for diagnosing and monitoring disease. Metabolic profiles can also aid in understanding drug metabolism, which may reduce the efficacy of a treatment, or in understanding drug toxicity, which can help to reduce adverse drug reactions. Metabolomic analysis can be used to determine the catalytic activity of proteins, for example, by quantifying changes in metabolite profiles in response to mutations in enzyme-coding genes, and to connect metabolic pathways that share common intermediates.

Biological samples for metabolite analysis may be cell or tissue lysates, body fluids such as urine or blood, or cell culture media that contain a great diversity of metabolites. These include building blocks for biosynthesis of cellular components such as amino acids, nucleotides, and lipids. Also present are various substrates, cofactors, regulators, intermediates, and end products of metabolic pathways such as carbohydrates, vitamins, organic acids, amines and alcohols, and inorganic molecules. These molecules have very different properties, and therefore comprehensive detection and quantification using a single method based on chemical characteristics presents a challenge.

Metabolomics employs spectroscopic techniques such as MS and nuclear magnetic resonance (NMR) spectroscopy to identify and quantify the metabolites in complex samples. Often, multiple methods are used in

parallel to obtain a comprehensive view of a metabolome. In a manner similar to protein identification, described above, MS measures the *m/z* ratio of charged metabolites. The molecules may be ionized by various methods before separation of different ions in an electromagnetic field. MS is typically coupled with chromatographic techniques that first separate metabolites based on their properties. For example, MS may be coupled with gas chromatography to separate volatile metabolites. Some nonvolatile metabolites, such as amino acids, may be chemically modified (derivatized) to increase their volatility. Liquid chromatography separates metabolites dissolved in a liquid solvent based on their characteristic retention times as they move through an immobilized matrix.

NMR spectroscopy is based on the principle that in an applied magnetic field, molecules (more precisely, atomic nuclei with an odd mass number) absorb and emit electromagnetic energy at a characteristic resonance frequency that is determined by their structure. Thus, the resonance frequencies provide detailed information about the structure of a molecule and enable differentiation among molecules with different structures, even when the difference is very small, such as between structural isomers. In contrast to MS, an initial metabolite separation step is not required, and NMR measures different types of molecules. In addition, NMR is not destructive, and in fact, it has been adapted to visualize molecules in living human cells in the diagnostic procedure magnetic resonance imaging (MRI). A drawback of NMR is low sensitivity, which means that it does not detect low-abundance molecules.

An illustration of the application of metabolome analysis is the identification of metabolites that are associated with pulmonary arterial hypertension. This vascular disease is characterized by narrowing or blocking of blood vessels in the lungs that leads to increased blood pressure in lung arteries, and in many cases heart failure and death. Researchers compared 376 metabolites in lung tissue samples from 8 patients with severe pulmonary arterial hypertension and 8 individuals without the disease using MS combined with liquid and gas chromatography. The levels of 93 metabolites were significantly higher or lower in the diseased lung tissue than in normal lung tissue (Fig. 2.54). Metabolite profiles revealed that severe disease was associated with a decrease in metabolites of the glycolytic pathway, an increase in intermediates of the tricarboxylic acid cycle, and altered fatty acid profiles. For example, levels of the glycolytic precursor glucose were higher, and those of the glycolytic intermediates fructose 1,6-bisphosphate, 3-phosphoglycerate, and phosphoenolpyruvate were lower in diseased lung tissue samples, indicating a disruption in glycolysis. Dicarboxylic fatty acids accumulated in diseased lungs, and this was supported by transcriptome analysis using a DNA microarray which showed higher expression levels of the gene encoding an aldehyde dehydrogenase, a key enzyme in ω-oxidation of fatty acids to produce dicarboxylic fatty acids. Consistent with elevated levels of tricarboxylic acid cycle intermediates, several genes encoding enzymes in this pathway were upregulated in diseased lung tissue compared to normal lung tissue. Metabolomic and transcriptomic analyses indicate that key pathways in carbon and energy metabolism are altered in patients with severe pulmonary arterial hypertension and, in particular, suggest that fatty acids, rather than carbohydrates, may supply energy via the tricarboxylic acid cycle for vascular remodeling. Elevated metabolites may serve as biomarkers for disease diagnosis or as targets for metabolic modulation therapy, for example, using fatty acid oxidation inhibitors.

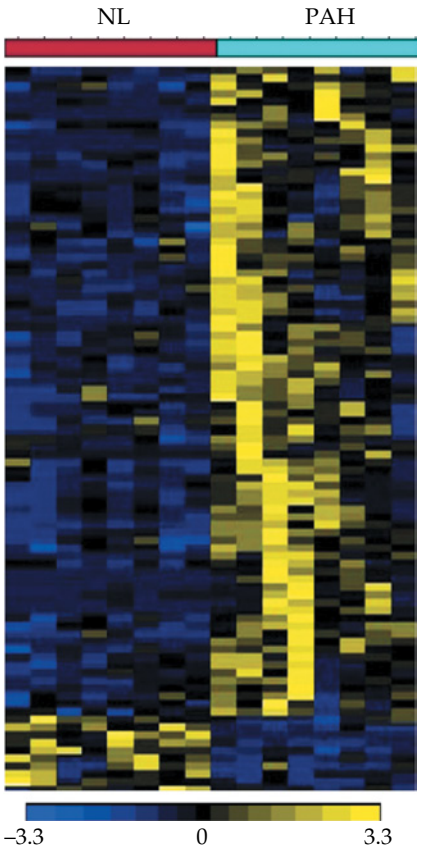


Figure 2.54 Metabolite profiles associated with severe pulmonary arterial hypertension (PAH). The relative levels of 93 metabolites in lung tissue samples from eight patients with PAH and eight individuals without the disease (NL) are represented as a heat map. Levels of a metabolite (rows) in each tissue sample (columns) were compared to the median metabolite level (black); shades of yellow represent increased levels, and shades of blue indicate decreased levels. In tissue samples from patients with PAH, 83 metabolites, including glucose, tricarboxylic acid cycle intermediates, and dicarboxylic fatty acids, were more abundant, and 10 metabolites, including intermediates in glycolysis, were less abundant than in normal lung tissue. Reprinted from Zhao et al., *PLoS ONE* **9:**e88727, 2014, © 2014 Zhao et al.

summary

Molecular biotechnology comprises a large number of fundamental techniques to identify, isolate, transfer, and express specific genes in a variety of host organisms. The tools for these processes were developed from an understanding of the biochemistry, genetics, and molecular biology of cells, especially prokaryotic cells, and viruses. Molecular cloning is the process of inserting a gene or other DNA sequence isolated from one organism into a vector and introducing it into a host cell. The discovery of restriction endonucleases was essential for this process, as it enabled predictable and reproducible cleavage of both target (insert) and vector DNAs in preparation for joining the two molecules. Digestion of target and vector DNA with the same restriction endonuclease generates compatible single-stranded extensions that can be joined by complementary base-pairing and the activity of the enzyme DNA ligase that catalyzes the formation of phosphodiester bonds. Another cloning method, known as recombinational cloning, does not utilize restriction endonucleases or DNA ligase for insertion of target DNA into a vector but, rather, exploits a system used by some viruses to integrate into the host genome via recombination at specific attachment sequences.

Cloned DNA is introduced into host cells, often bacterial cells that are competent to take up exogenous DNA, a process known as transformation. Vectors that carry the target DNA into the host cell are usually derived from natural bacterial plasmids that have been genetically engineered with several unique endonuclease recognition sequences (multiple-cloning sites) to facilitate cloning a DNA sequence (origin of replication) that enables replication in the host. Transformation is generally inefficient; however, transformed cells may be distinguished from nontransformed cells by testing for the activity of genes that are present on the vector, including genes for resistance to antibiotics or synthesis of colored products.

To clone and express genes that encode eukaryotic proteins in a bacterial host, the introns must first be removed. Purified mRNA, which does not contain introns, is used as a template for the synthesis of cDNA by the enzyme reverse transcriptase. Oligonucleotide primers can be designed to target a specific mRNA for cDNA synthesis or to anneal to the poly(A) tails present on most eukaryotic mRNAs to generate a cDNA library that contains all of the protein-coding sequences expressed by a source eukaryote under a given set of conditions. Construction of a genomic DNA library from a prokaryote is more straightforward and entails cleaving the DNA to obtain overlapping fragments for cloning. Libraries are screened by a variety of methods to identify clones that have a particular sequence or that produce a target protein.

The CRISPR-Cas system is used to alter (edit) intact genomes *in vivo*. In this system, an sgRNA that contains a 20-nucleotide sequence complementary to the target site in the genome is introduced into the host cell together with the endonuclease Cas9. The sgRNA guides Cas9 to the genomic target site, adjacent to a specific PAM sequence, and the target sequence is cleaved by the endonuclease. Double-stranded DNA cleavage activates cellular DNA repair systems, which results in nucleotide deletions or insertions that disrupt the target gene. If specific donor DNA is also introduced into the cell, it may be incorporated into the genome by recombination between the target sequence and homologous sequences flanking the donor DNA. The CRISPR-Cas technology can be used to disrupt, insert, tag, or alter the regulation of a gene in the genomes of microorganisms or multicellular organisms.

Amplification, synthesis, and sequencing of DNA are also fundamental tools of molecular biotechnology. PCR is a powerful method for generating millions of copies of a specific sequence of DNA from very small amounts of starting material. Amplification is achieved in 30 or more successive cycles of template DNA denaturation, annealing of two oligonucleotide primers to complementary sequences flanking a target gene in the single-stranded DNA, and DNA synthesis extending from the primer by a thermostable DNA polymerase. Among innumerable applications, PCR can be used to detect or quantify a specific nucleotide sequence in a complex biological sample or to obtain large amounts of a particular DNA sequence either for cloning or for sequencing.

Genes can be assembled from chemically synthesized oligonucleotides or from PCR fragments. Oligonucleotides with overlapping regions of complementarity are annealed, and gaps are filled in with DNA polymerase and DNA ligase to generate double-stranded DNA molecules. Alternatively, PCR fragments that span an entire gene and have overlapping homologous ends are combined with an exonuclease that digests several nucleotides from the 5′ ends of the fragments to generate single-stranded overhangs. The fragments are annealed via complementary overhangs, and gaps are closed with DNA polymerase and DNA ligase. The synthesized DNA molecules may be cloned into a vector or assembled into larger segments or an entire chromosome by ligation of restriction enzyme-digested fragments or homologous recombination.

The nucleotide sequence of a gene can reveal useful information about the function, regulation, and evolution

of the gene. Most of the sequencing technologies currently used generally involve (i) addition of nucleotides by DNA polymerase to a primer based on complementarity to a template DNA fragment and (ii) detection and identification of the nucleotide(s) added. The dideoxynucleotide (Sanger) sequencing method relies on the incorporation of a synthetic dideoxynucleotide that lacks a 3′ hydroxyl group into a growing DNA strand, which terminates DNA synthesis. Conditions are optimized so that the dideoxynucleotides are incorporated randomly, producing DNA fragments of different lengths that terminate with one of the four dideoxynucleotides, each tagged with a different fluorescent dye. The fragments are separated according to their size by electrophoresis, and the sequence of fluorescent signals is determined and converted into a nucleotide sequence. Sequencing using reversible chain terminators (Illumina sequencing) reveals the sequence of a DNA fragment by detecting single-nucleotide extensions. The four nucleotides are added to the reaction together in each cycle, and after the unincorporated nucleotides are washed away, the nucleotide incorporated by DNA polymerase is distinguished by its fluorescent signal. The fluorescent dye and a blocking group that prevents addition of more than one nucleotide during each cycle are chemically cleaved, and the cycle is repeated. The Sanger and Illumina sequencing methods generate sequence reads of about 800 and 200 bp, respectively. To generate longer reads of tens to hundreds of kbp, single-molecule sequencing technologies have been developed. In single-molecule real-time (PacBio) sequencing, a double-stranded DNA molecule is captured by a highly processive DNA polymerase attached to the bottom of a very low-volume well. Emissions from fluorescently tagged nucleotides held in the polymerase's active site are detected in real time by a narrowly focused laser, before the cleaved pyrophosphate and attached fluorophore diffuse away. Nanopore sequencing does not require DNA synthesis. Rather, a double-stranded DNA molecule is unwound by DNA polymerase, and one of the strands is threaded through a protein nanopore embedded in a synthetic membrane. An electric current is applied across the membrane, and as the DNA strand passes through the pore, it disrupts the current to generate a signal that is characteristic of the nucleotides in the pore. The error rate in the reads obtained by the single-molecule sequencing technologies is much higher than those attained by the Sanger and Illumina sequencing methods.

New sequencing technologies have enabled rapid and inexpensive acquisition of genome sequences. Hundreds of millions of short nucleotide reads can be acquired simultaneously (massive parallelization) and assembled into contigs. Using these approaches, the genome sequences of thousands of organisms from all domains of life have been completed or are in progress. The next steps are to annotate the sequence features and to determine the functions of the genes encoded in the genomes by investigating patterns of transcription (transcriptomics), protein synthesis (proteomics), and small-molecule production (metabolomics) using a variety of techniques such as RNA sequencing, DNA microarrays, 2D PAGE, mass spectrometry, protein microarray analysis, and NMR spectroscopy. Comparison of genome sequences can reveal the genetic basis of a disease, the mechanism of pathogenicity of a microbe, or the evolutionary relationships among organisms, while transcript, protein, and metabolite profiles can identify biomarkers for diagnosis and treatment of disease.

REFERENCES

Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, et al. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456:**53–59.

Bondy-Denomy J, **Davidson AR.** 2014. To acquire or resist: the complex biological effects of CRISPR-Cas systems. *Trends Microbiol* **22:**218–225.

Cello J, **Paul AV**, **Wimmer E.** 2002. Chemical synthesis of poliovirus cDNA: generation of infectious virus in the absence of natural template. *Science* **297:**1016–1018.

Chen CS, **Korobkova E**, **Chen H**, **Zhu J**, **Jian X**, **Tao SC**, **He C**, **Zhu H.** 2008. A proteome chip approach reveals new DNA damage recognition activities in *Escherichia coli*. *Nat Methods* **5:**69–74.

Cohen SN, **Chang ACY**, **Boyer HW**, **Helling RB.** 1973. Construction of biologically functional bacterial plasmids *in vitro*. *Proc Natl Acad Sci USA* **70:**3240–3244.

Cong L, **Ran FA**, **Cox D**, **Lin S**, **Barretto R**, **Habib N**, **Hsu PD**, **Wu X**, **Jiang W**, **Marraffini LA**, **Zhang F.** 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339:**819–823.

Czar MJ, **Anderson JC**, **Bader JS**, **Peccoud J.** 2009. Gene synthesis demystified. *Trends Biotechnol* **27:**63–72.

Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, et al. 2009. Real-time DNA sequencing from single polymerase molecules. *Science* **323:**133–138.

Fodor SP, **Rava RP**, **Huang XC**, **Pease AC**, **Holmes CP**, **Adams CL.** 1993. Multiplexed biochemical assays with biological chips. *Nature* **364:**555–556.

Formstecher E, **Aresta S**, **Collura V**, **Hamburger A**, **Meil A**, **Trehin A**, **Reverdy C**, **Betin V**, **Maire S**, **Brun C**, **Jacq B**, **Arpin M**, **Bellaiche Y**, **Bellusci S**, **Benaroch P**, **Bornens M**, **Chanet R**, **Chavrier P**, **Delattre O**, **Doye V**, **Fehon R**, **Faye G**, **Galli T**, **Girault JA**, **Goud B**, **de Gunzburg J**, **Johannes L**, **Junier MP**, **Mirouse V**, **Mukherjee A**, **Papadopoulo D**, **Perez F**, **Plessis A**, **Rossé C**, **Saule S**, **Stoppa-Lyonnet D**, **Vincent A**, **White M**, **Legrain P**, **Wojcik J**, **Camonis J**, **Daviet L.** 2005. Protein interaction mapping: a *Drosophila* case study. *Genome Res* **15:**376–384.

Gibson DG. 2014. Programming biological operating systems: genome design, assembly and activation. *Nat Methods* **11:**521–526.

Gibson DG, **Benders GA**, **Andrews-Pfannkoch C**, **Denisova EA**, **Baden-Tillson H**, **Zaveri J**, **Stockwell TB**, **Brownley A**, **Thomas DW**, **Algire MA**, **Merryman C**, **Young L**, **Noskov VN**, **Glass JI**, **Venter JC**, **Hutchison CA III**, **Smith HO.** 2008. Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science* **319:**1215–1220.

Gibson DG, **Glass JI, Lartigue C**, **Noskov VN**, **Chuang RY**, **Algire MA**, **Benders GA**, **Montague MG**, **Ma L**, **Moodie MM**, **Merryman C**, **Vashee S**, **Krishnakumar R**, **Assad-Garcia N**, **Andrews-Pfannkoch C**, **Denisova EA**, **Young L**, **Qi ZQ**, **Segall-Shapiro TH**, **Calvey CH**, **Parmar PP**, **Hutchison CA III**, **Smith HO**, **Venter JC.** 2010. Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* **329:**52–56.

Gibson DG, **Young L**, **Chuang R-Y**, **Venter JC**, **Hutchison CA III**, **Smith HO.** 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **6:**343–345.

Haab BB. 2006. Applications of antibody array platforms. *Curr Opin Biotechnol* **17:**415–421.

Handelsman J. 2004. Metagenomics: application of genomics to uncultured microorganisms. *Microbiol Mol Biol Rev* **68:**669–685.

Hudson ME, **Pozdnyakova I**, **Haines K**, **Mor G**, **Snyder M.** 2007. Identification of differentially expressed proteins in ovarian cancer using high-density protein microarrays. *Proc Natl Acad Sci USA* **104:**17494–17499.

Iacobucci I, **Monaco V**, **Cozzolino F**, **Monti M.** 2021. From classical to new generation approaches: an excursus of -omics methods for investigation of protein-protein interaction networks. *J Proteomics* **230:**103990.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* **409:**860–921.

Isaacson W. 2021. *The Code Breaker*. Simon and Schuster, New York, NY.

Ito T, **Chiba T**, **Ozawa R**, **Yoshida M**, **Hattori M**, **Sakaki Y.** 2001. A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci USA* **98:**4569–4574.

Jain M, **Olsen HE**, **Paten B**, **Akeson M.** 2016. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol* **17:**239.

Jinek M, **Chylinski K**, **Fonfara I**, **Hauer M**, **Doudna JA**, **Charpentier E.** 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337:**816–821.

Ju J, **Kim DH**, **Bi L**, **Meng Q**, **Bai X**, **Li Z**, **Li X**, **Marma MS**, **Shi S**, **Wu J**, **Edwards JR**, **Romu A**, **Turro NJ.** 2006. Four-color DNA sequencing by synthesis using cleavable fluorescent nucleotide reversible terminators. *Proc Natl Acad Sci USA* **103:**19635–19640.

Köcher T, **Superti-Furga G.** 2007. Mass spectrometry-based functional proteomics: from molecular machines to protein networks. *Nat Methods* **4:**807–815.

Levene MJ, **Korlach J**, **Turner SW**, **Foquet M**, **Craighead HG**, **Webb WW.** 2003. Zero-mode waveguides for single-molecule analysis at high concentrations. *Science* **299:**682–686.

Li Y. 2011. The tandem affinity purification technology: an overview. *Biotechnol Lett* **33:**1487–1499.

Logsdon GA, **Vollger MR**, **Eichler EE.** 2020. Long-read human genome sequencing and its applications. *Nat Rev Genet* **21:**597–614.

Lowe R, **Shirley N**, **Bleackley M**, **Dolan S**, **Shafee T.** 2017. Transcriptomics technologies. *PLOS Comput Biol* **13:**e1005457.

Mali P, **Yang L**, **Esvelt KM**, **Aach J**, **Guell M**, **DiCarlo JE**, **Norville JE**, **Church GM.** 2013. RNA-guided human genome engineering via Cas9. *Science* **339:**823–826.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, et al. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437:**376–380.

Martin JA, **Wang Z.** 2011. Next-generation transcriptome assembly. *Nat Rev Genet* **12:**671–682.

Maxam AM, **Gilbert W.** 1977. A new method for sequencing DNA. *Proc Natl Acad Sci USA* **74:**560–564.

Mertz JE, **Davis RW.** 1972. Cleavage of DNA by R 1 restriction endonuclease generates cohesive ends. *Proc Natl Acad Sci USA* **69:**3370–3374.

Mirete S, **Mora-Ruiz MR**, **Lamprecht-Grandío M**, **de Figueras CG**, **Rosselló-Móra R**, **González-Pastor JE.** 2015. Salt resistance genes revealed by functional metagenomics from brines and moderate-salinity rhizosphere within a hypersaline environment. *Front Microbiol* **6:**1121.

Mullis K, **Faloona F**, **Scharf S**, **Saiki R**, **Horn G**, **Erlich H.** 1986. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb Symp Quant Biol* **51:**263–273.

Noonan JP, **Coop G**, **Kudaravalli S**, **Smith D**, **Krause J**, **Alessi J**, **Chen F**, **Platt D**, **Pääbo S**, **Pritchard JK**, **Rubin EM.** 2006. Sequencing and analysis of Neanderthal genomic DNA. *Science* **314:**1113–1118.

Parrish JR, **Gulyas KD**, **Finley RL Jr.** 2006. Yeast two-hybrid contributions to interactome mapping. *Curr Opin Biotechnol* **17:**387–393.

Roberts RJ. 2005. How restriction enzymes became the workhorses of molecular biology. *Proc Natl Acad Sci USA* **102:**5905–5908.

Roux KJ, **Kim DI**, **Raida M**, **Burke B.** 2012. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J Cell Biol* **196:**801–810.

Rust M, Helfrich EJN, Freeman MF, Nanudorn P, Field CM, Rückert C, Kündig T, Page MJ, Webb VL, Kalinowski J, Sunagawa S, Piel J. 2020. A multiproducer microbiome generates chemical diversity in the marine sponge *Mycale hentscheli*. *Proc Natl Acad Sci USA* **117:**9508–9518.

Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239:**487–491.

Sanger F, Nicklen S, Coulson AR. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* **74:**5463–5467.

Shao Y, Wang XB, Zhang JJ, Li ML, Wu SS, Ma XY, Wang X, Zhao HF, Li Y, Zhu HH, Irwin DM, Wang DP, Zhang GJ, Ruan J, Wu DD. 2020. Genome and single-cell RNA-sequencing of the earthworm *Eisenia andrei* identifies cellular mechanisms underlying regeneration. *Nat Commun* **11:**2656.

Sims D, Sudbery I, Ilott NE, Heger A, Ponting CP. 2014. Sequencing depth and coverage: key considerations in genomic analyses. *Nat Rev Genet* **15:**121–132.

Slatko BE, Gardner AF, Ausubel FM. 2018. Overview of next generation sequencing technologies. *Curr Protoc Mol Biol* **122:**e59.

Sreekumar A, Poisson LM, Rajendiran TM, Khan AP, Cao Q, Yu J, Laxman B, Mehra R, Lonigro RJ, Li Y, Nyati MK, Ahsan A, Kalyana-Sundaram S, Han B, Cao X, Byun J, Omenn GS, Ghosh D, Pennathur S, Alexander DC, Berger A, Shuster JR, Wei JT, Varambally S, Beecher C, Chinnaiyan AM. 2009. Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. *Nature* **457:**910–914.

The 1000 Genomes Project Consortium. 2015. A global reference for human genetic variation. *Nature* **526:**68–74.

Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM, Solovyev VV, Rubin EM, Rokhsar DS, Banfield JF. 2004. Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428:**37–43.

Venter JC, Adams MD, Myer EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, et al. 2001. The sequence of the human genome. *Science* **291:**1304–1351.

Walhout AJ, Temple GF, Brasch MA, Hartley JL, Lorson MA, van den Heuvel S, Vidal M. 2000. GATEWAY recombinational cloning: application to the cloning of large numbers of open reading frames or ORFeomes. *Methods Enzymol* **328:**575–592.

Wang Z, Gerstein M, Snyder M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* **10:**57–63.

Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, He W, Chen YJ, Makhijani V, Roth GT, Gomes X, Tartaro K, Niazi F, Turcotte CL, Irzyk GP, Lupski JR, Chinault C, Song XZ, Liu Y, Yuan Y, Nazareth L, Qin X, Muzny DM, Margulies M, Weinstock GM, Gibbs RA, Rothberg JM. 2008. The complete genome of an individual by massively parallel DNA sequencing. *Nature* **452:**872–876.

Williams RJ. 2003. Restriction endonucleases: classification, properties, and applications. *Mol Biotechnol* **23:**225–243.

Yosef I, Qimron U. 2015. How bacteria get spacers from invaders. *Nature* **519:**166–167.

Zhao Y, Peng J, Lu C, Hsin M, Mura M, Wu L, Chu L, Zamel R, Machuca T, Waddell T, Liu M, Keshavjee S, Granton J, de Perrot M. 2014. Metabolomic heterogeneity of pulmonary arterial hypertension. *PLoS One* **9:**e88727.

review questions

1. Describe a strategy using restriction endonucleases to clone a bacterial gene into a vector for propagation in *E. coli*. Assume that the target sequence is known. Describe the selection for *E. coli* cells that carry the cloned gene. Consider methods to minimize unwanted products.
2. Describe the features of a useful cloning vector.
3. Outline a strategy to clone a eukaryotic gene into a vector for expression in *E. coli*. Briefly describe the activity of the enzymes used in the process.
4. Describe how a library of open reading frames that represents a proteome is constructed by recombinational cloning.
5. Outline an experiment to construct and screen a genomic library to identify a gene that encodes a particular function.
6. How is the CRISPR-Cas system used to delete nucleotides at a specific site in a eukaryotic genome?
7. How is the CRISPR-Cas system used to insert a sequence at a specific site in a eukaryotic genome?
8. Outline the steps in a PCR cycle. What component of a PCR determines the specificity of the amplified product?
9. Describe how PCR is used to clone a specific gene.
10. Why is real-time PCR quantitative?
11. Suggest two different strategies for synthesizing a 0.5-kb gene.
12. What is a dideoxynucleotide? How is it used to determine the sequence of a DNA molecule?
13. How are incorporated nucleotides recognized after each cycle of sequencing using reversible chain terminators?
14. Describe a sequencing technology that generates long sequence reads.
15. List some advantages and disadvantages of single-molecule real-time sequencing.
16. Why are adaptors often ligated to DNA fragments prior to sequencing?

17. Describe how a cluster of thousands of copies of a DNA sequencing template can be generated on a solid support.

18. Define the following terms as they relate to genome sequencing: shotgun cloning, multiplexing, massive parallelization, barcode, Phred quality score, contig, scaffold, paired end reads, coverage, *de novo* sequencing, resequencing, reference genome, annotation, metagenome.

19. How are gene expression levels quantified using high-throughput RNA sequencing?

20. Describe the process by which RNA from individual cells in a tissue is sequenced.

21. Explain why random hexamers used for RNA sequencing result in overrepresentation of sequences from the 5′ end of a gene.

22. How can 2D PAGE be used to identify proteins that are differentially expressed in two samples?

23. How are proteins identified by mass spectrometry?

24. Describe some applications for protein microarrays.

25. What biological information may be provided by a proteome that is not provided by a transcriptome?

26. Describe two methods to determine protein-protein interactions.

27. What is an affinity tag?

28. What biological information may be provided by the tandem affinity purification tag system that is not provided by a two-hybrid assay?

29. What are some advantages and disadvantages of proximity-dependent labeling compared to tandem affinity purification to identify protein-protein interactions?

30. Explain how metabolomics may be used to identify biomarkers of disease.

3 Production of Recombinant Proteins

Protein Production in Prokaryotic Hosts
Regulating Transcription
Increasing Translation Efficiency
Increasing Protein Stability
Increasing Protein Secretion
Facilitating Protein Purification
Integrating DNA into the Host Chromosome
Heterologous Protein Production in Eukaryotic Cells
Posttranslational Modification of Eukaryotic Proteins
General Features of Eukaryotic Expression Systems
Yeast Expression Systems
Baculovirus–Insect Cell Expression Systems
Mammalian Cell Expression Systems
Protein Engineering
Directed Mutagenesis
Random Mutagenesis
Examples of Protein Engineering
SUMMARY
REFERENCES
REVIEW QUESTIONS

Recombinant proteins are produced from heterologous genes cloned in a selected host organism. They are widely used in industry, research, and medicine, the latter including vaccines, growth factors, and antibodies. The primary objectives of commercial production of a recombinant protein are high quality, high yields, and low cost. There is no single strategy for obtaining maximal expression of every cloned gene, and many biological parameters must be manipulated to obtain optimal levels of protein production. These include the genetic elements for controlling transcription, translation, protein stability, and secretion of the product of the cloned gene from the host cell. The level of foreign gene expression also depends on the host organism. Initially, many of the commercially important proteins produced by recombinant DNA technology were synthesized in *Escherichia coli*. Today many other host systems, such as other bacterial strains, yeasts, microalgae, and insect and mammalian cells, are employed to produce heterologous proteins. Each of these systems has advantages and disadvantages (Table 3.1). For example, while cloned genes may be expressed at high levels and low cost in *E. coli*, the proteins produced are not glycosylated. Posttranslational glycosylation is essential for the function of many human therapeutic proteins, and therefore these proteins are often produced in cultured mammalian cells even though the costs are higher and the yields lower than when these proteins are produced in *E. coli*.

Protein Production in Prokaryotic Hosts

There are several good reasons to employ prokaryotic cells for production of heterologous proteins, and many bacterial hosts are commercially available, such as Gram-negative *E. coli* and *Pseudomonas fluorescens* and

Molecular Biotechnology: Principles and Applications of Recombinant DNA,
Sixth Edition. Bernard R. Glick and Cheryl L. Patten.

© 2022 American Society for Microbiology.

Table 3.1 Production of recombinant human proteins in various biological hosts

Parameter	Bacteria	Yeast	Mammalian cell culture	Transgenic plants
Glycosylation	None	Incorrect	Correct	Generally correct; small differences
Multimeric proteins assembled	Limited	Limited	Limited	Yes
Production costs	Low-medium	Medium	High	Very low
Protein folding accuracy	Low	Medium	High	High
Protein yield	High	Medium-high	Low-medium	Medium
Scale-up capacity	High	High	Low	Very high
Scale-up costs	High	High	High	Low
Time required	Low	Low-medium	Medium-high	High
Skilled workers required	Medium	Medium	High	Low
Acceptable to regulators	Yes	Yes	Yes	Not yet

Gram-positive *Bacillus subtilis*, *Lactococcus lactis*, and *Corynebacterium glutamicum*. The genetics, molecular biology, biochemistry, and physiology of these bacteria are well understood, they can often be grown to high cell densities in large-scale bioreactors, the growth medium is relatively inexpensive, protocols for their genetic manipulation have been optimized, and vectors are available that carry signals for high levels of gene expression. The latter include the promoter and other transcription regulatory sequences and sequences that control translation efficiency, such as the strength of the ribosome-binding site. Production of a foreign protein in a bacterium may require manipulation of the coding sequence to increase protein stability or direct it to be secreted. To ensure that the cloned gene is stably maintained in the host cell, it may be necessary to integrate it into the chromosome of the host cell.

Regulating Transcription

The minimal requirement for an effective gene expression system is the presence of a strong and regulatable promoter sequence upstream from a cloned gene. A strong promoter is one that has high affinity for RNA polymerase, with the consequence that the adjacent downstream region is frequently transcribed (Fig. 3.1). However, a high level of transcription is not always desirable, and the presence of regulatory sequences in the promoter

Figure 3.1 A strong *E. coli* promoter resembles the consensus sequences for the –35 and –10 boxes that bind to RNA polymerase. The consensus sequence was determined by aligning many *E. coli* promoters and identifying conserved nucleotides in the sequences centered –35 and –10 bp upstream of the transcription start site (+1). The distance, but not the nucleotide sequence, between the two boxes is also conserved.

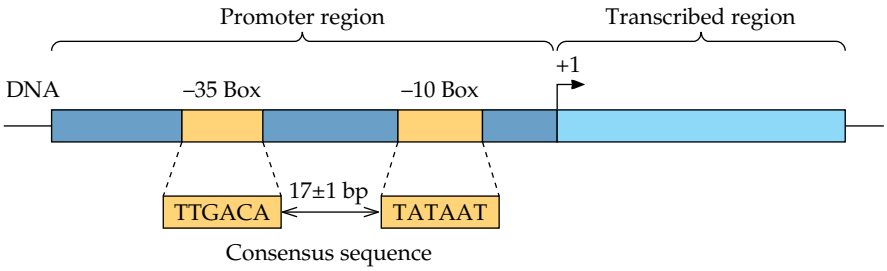


Table 3.2 Promoters commonly used for expression of cloned genes in prokaryotic hosts

Promoter	Source	Regulator(s)	Features
lac/*lacUV5*	*E. coli* lactose operon	Lactose repressor LacI, CAP activator	Induced by lactose (allolactose) or IPTG
Gene 10	Bacteriophage T7	T7 RNA polymerase	Strong promoter
trp	*E. coli* trypto-phan operon	Tryptophan repressor TrpR	Repressed by tryptophan
tet	*E. coli* Tn10 tetracycline resistance gene	Tetracycline repressor TetR	Induced by tetracycline
tac	Synthetic hybrid of *lacUV5* and *trp* promoters	LacI repressor	Induced by IPTG; approximately 10 times stronger than the *lac* promoter
araBAD	*E. coli* arabinose operon	Arabinose activator/ repressor AraC	Induced by arabinose
pL	Bacteriophage lambda	λcI repressor/ mutant λcI^{857} repressor	Induced by nalidixic acid, which elicits the SOS response; temperature sensitive λcI^{857} repressor is unstable above 37°C
cspA	*E. coli* cold shock gene		Induced at 15°C
nisA	*Lactococcus lactis* nisin operon	Two component regulators NisK and NisR	Induced by nisin
luxI	*Vibrio fischeri* luciferase operon	Activator LuxR	Acylhomoserine lactone synthesized by LuxI at high cell densities

region enables the cell (and the researcher or manufacturer) to control the extent of transcription in a precise manner. Many different promoters with distinctive properties that make them useful for controlling expression have been isolated from a range of organisms (Table 3.2).

A high, constitutive (continuous) level of expression of a cloned gene from a strong, unregulated promoter is often detrimental to the host cell because it creates an energy drain, thereby impairing essential host cell functions and growth (Table 3.3). In addition, all or a portion of the plasmid carrying a constitutively expressed cloned gene may be lost after several division cycles, since cells without a plasmid grow faster and eventually take over the culture. Such plasmid instability is a major problem that may prevent the efficient production of a plasmid-borne gene product on a large scale. To overcome these drawbacks, it is desirable to use a strong regulatable promoter to control transcription in such a way that a cloned

Table 3.3 Factors that may increase the metabolic burden on a prokaryotic host cell that is expressing high levels of a cloned gene

Condition	Effect(s)
High plasmid copy number/size	Requires high amounts of cellular energy and nucleotides for plasmid replication and maintenance; decreases growth rate
Overproduction of target and marker proteins	Depletes the pools of certain aminoacyl-tRNAs and amino acids; decreases cellular ATP or GTP levels
Secretion of high levels of target proteins	Jams export sites; prevents proper localization of essential host proteins
Production of foreign enzymes	May catabolize essential metabolites or produce toxic products
Depletion of dissolved oxygen in medium	Decreases metabolic activity and expression of cloned gene

The *tac* Promoter: a Functional Hybrid Derived from the *trp* and *lac* Promoters

milestone

De Boer and his colleagues began their efforts to construct the *tac* promoter with the idea of combining portions of two different strong and regulatable promoters to create an even stronger promoter that would direct very high levels of foreign-gene expression. When they undertook their studies, although the DNA sequences of a number of prokaryotic promoters, mostly from *E. coli*, were known, the precise features that enabled a promoter to be efficient at directing transcription were not well understood. It was known that almost all mutations that affected the strength of a prokaryotic promoter were found in either the –10 region or the –35 region, which is approximately 10 or 35 bp upstream of the mRNA transcription start site, respectively. Moreover, only mutations that made an existing promoter more like the consensus sequences for each of these regions, i.e., 5′-TATAAT-3′ for the –10 region and 5′-TTGACA-3′ for the –35 region, increased the strength of the promoter. The consensus sequences had been deduced by comparing the DNA sequences of all known promoters and determining which nucleotides occurred most often. De Boer and his colleagues also knew that the *lacUV5* promoter, which is a stronger variant of the *lac* promoter, had a consensus sequence at its –10 but not its –35 region, while the *trp* promoter, which normally controls the transcription of genes involved in the biosynthesis of tryptophan, has a consensus sequence at its –35 but not its –10 region. They decided to create a fusion promoter that included the –10 region from the *lac* promoter and the –35 region from the *trp* promoter. They tested this new "*tac*" promoter, as they called it, for its ability to direct the synthesis of the enzyme galactose kinase in *E. coli* and compared it in the same assay system with the *trp* and *lac* promoters. In agreement with their initial idea, the *tac* promoter was found to be approximately 5 times stronger than the *trp* promoter and 10 times stronger than the *lac* promoter (de Boer et al., *Proc. Natl. Acad. Sci. USA* **80:**21–25, 1983). In addition, like the *lac* promoter, the *tac* promoter was repressed by the *lac* repressor and derepressed by IPTG. Thus, this new promoter was not only strong, but also regulatable.

gene is expressed only at a specific stage in the host cell growth cycle and only for a specified duration. The production process is performed in two stages. During the first, or growth, stage, the promoter controlling the transcription of the target gene is turned off, while during the second, or induction, stage, this promoter is turned on.

Widely used, strong, regulatable promoters are those from the *E. coli* lactose (*lac*) and tryptophan (*trp*) operons and from bacteriophage genes such as the gene 10 promoter from *E. coli* bacteriophage T7. Each of these promoters interacts with regulatory proteins (repressors or activators), which provide a controllable switch for either turning on or turning off transcription of adjacent (downstream) cloned genes. The *lac* and *trp* promoters are recognized by the major form of the *E. coli* RNA polymerase holoenzyme. This holoenzyme is formed when a sigma factor protein, in this case the sigma factor RpoD, combines with the core proteins of RNA polymerase. The sigma factor directs the binding of the holoenzyme to promoter regions on the DNA.

Lac Promoter

The *E. coli lac* promoter is negatively regulated (turned off) by the LacI repressor protein, which binds to the operator of the *lac* operon and prevents it from being transcribed in the absence of lactose in the growth medium (Fig. 3.2). Induction (turning on) of the *lac* promoter is achieved by the addition of either lactose or isopropyl-β-D-thiogalactopyranoside (IPTG), a synthetic inducer, to the medium (Fig. 3.3). In the cell, lactose is converted to allolactose by low levels of β-galactosidase that are synthesized when the system is repressed, before it can act as an inducer. The enzyme β-galactosidase is encoded by the *lacZ* gene of the *lac* operon, and it

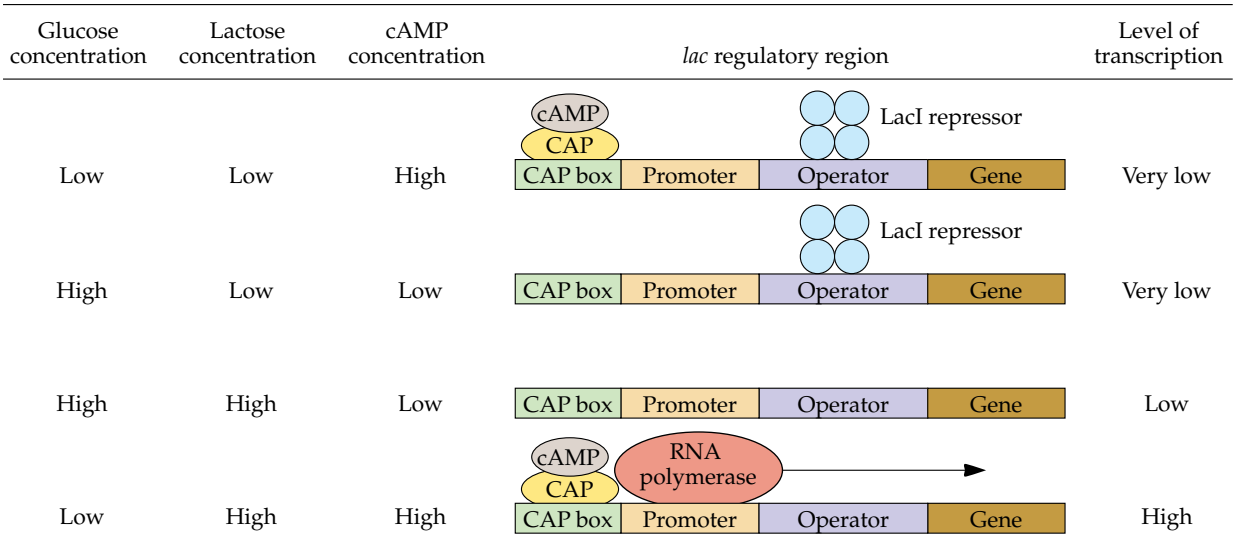


Figure 3.2 The effects of the concentrations of glucose and lactose in the growth medium on the level of transcription from the *E. coli lac* promoter. The cAMP concentration refers to intracellular levels that are regulated by glucose. The LacI repressor is a tetramer that binds to the *lac* operator in the absence of lactose and prevents transcription. In the presence of cAMP, positive regulator CAP binds to a CAP recognition site (CAP box) on the DNA and activates high levels of transcription. The arrow indicates the direction of transcription.

is primarily involved in the cleavage of lactose into glucose and galactose. Both allolactose and IPTG can bind to LacI and prevent it from binding to the *lac* operator, thereby enabling transcription to occur.

Transcription from the *lac* promoter is also positively regulated by the binding of the catabolite activator protein (CAP) (also sometimes referred to as the cyclic AMP [cAMP] repressor protein, or CRP) to a region of the DNA (the CAP box) just upstream of the promoter region (Fig. 3.2). When

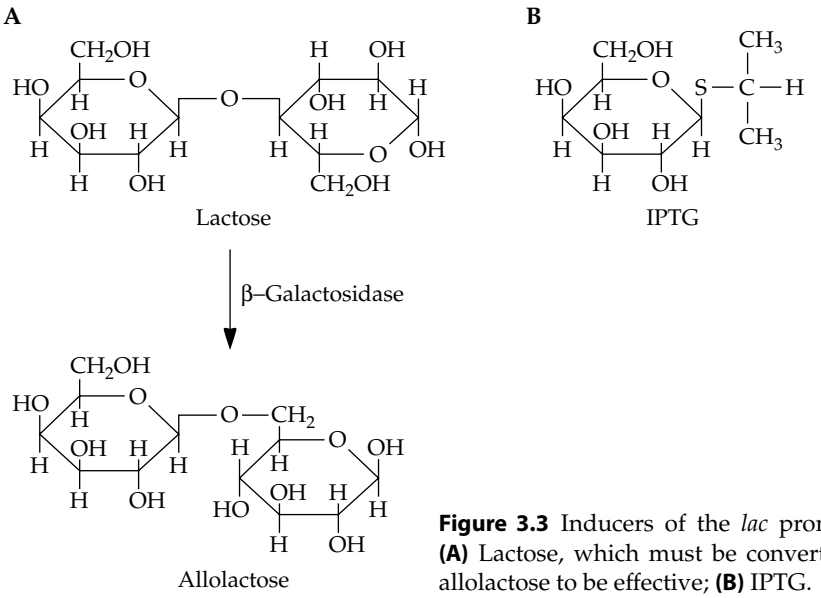


Figure 3.3 Inducers of the *lac* promoter. **(A)** Lactose, which must be converted to allolactose to be effective; **(B)** IPTG.

CAP binds to the CAP box, it increases the affinity of the promoter for RNA polymerase, thereby increasing transcription of the genes downstream from the promoter. The affinity of CAP for its binding site on the DNA is enhanced by its association with cAMP, which is produced at high levels by the cell when the amount of glucose in the medium is low. Thus, when inducer (lactose or IPTG) is present and there is no repressor bound to the operator, a high intracellular concentration of cAMP can lead to a high level of transcription of the genes downstream of the *lac* promoter. In practice, *lacUV5*, a variant of the *lac* promoter that contains an altered nucleotide sequence in the –10 region and is a stronger promoter than the wild-type *lac* promoter, is usually used in plasmid expression vectors. In addition, a mutant form of the *lacI* gene ($lacI^q$) that produces much higher levels of the *lac* repressor is often employed to decrease basal (background) levels of transcription (transcriptional leakiness) under noninduced conditions (i.e., transcription of a cloned gene in the absence of inducer).

T7 Expression System

The promoters of bacteriophage genes drive the production of high levels of viral proteins in a host bacterial cell and are often used in biotechnology applications that require a strong promoter. One such promoter from gene 10 of the *E. coli* bacteriophage T7 activates very high levels of gene expression. However, this promoter is not recognized by the *E. coli* RNA polymerase, but rather, it requires the T7 RNA polymerase for transcription, which transcribes eight times faster than the *E. coli* RNA polymerase. Therefore, to utilize this promoter for transcription of cloned genes, the T7 RNA polymerase gene is inserted in the *E. coli* chromosome (Fig. 3.4). To regulate expression of the T7 RNA polymerase gene, it is placed under the control of the *E. coli lac* promoter and operator. During the growth phase, LacI prevents the synthesis of T7 RNA polymerase, and the cloned gene product is not produced. When the culture has reached a suitable cell density, IPTG is added to the medium to induce expression of the T7 RNA polymerase gene. The T7 RNA polymerase binds to the T7 gene 10 promoter and transcribes the cloned gene.

Promoters must be chosen carefully, especially for the large-scale production of foreign proteins. Chemical inducers can be costly, toxic, or difficult to remove; thermal induction of promoters may induce the production of heat shock proteins, including proteases; nutrient-responsive promoters limit the types of media that can be used for cell growth and induction; and oxygen-regulated promoters often have significant basal levels of activity as a consequence of the inherent difficulty in precisely controlling dissolved oxygen levels in the growth medium. Promoters that are induced when cells enter the stationary phase of the growth cycle may be useful for large-scale applications. In addition, *E. coli* is not necessarily the bacterium of choice for the expression of some foreign proteins. While *E. coli* promoters can regulate the expression of cloned genes in some other bacteria, different promoters may be required in some host strains.

Increasing Translation Efficiency

Expressing a cloned gene from a strong, regulatable promoter, although essential, may not be sufficient to maximize the yield of the cloned gene product. Other factors, such as the efficiency of translation and the stability

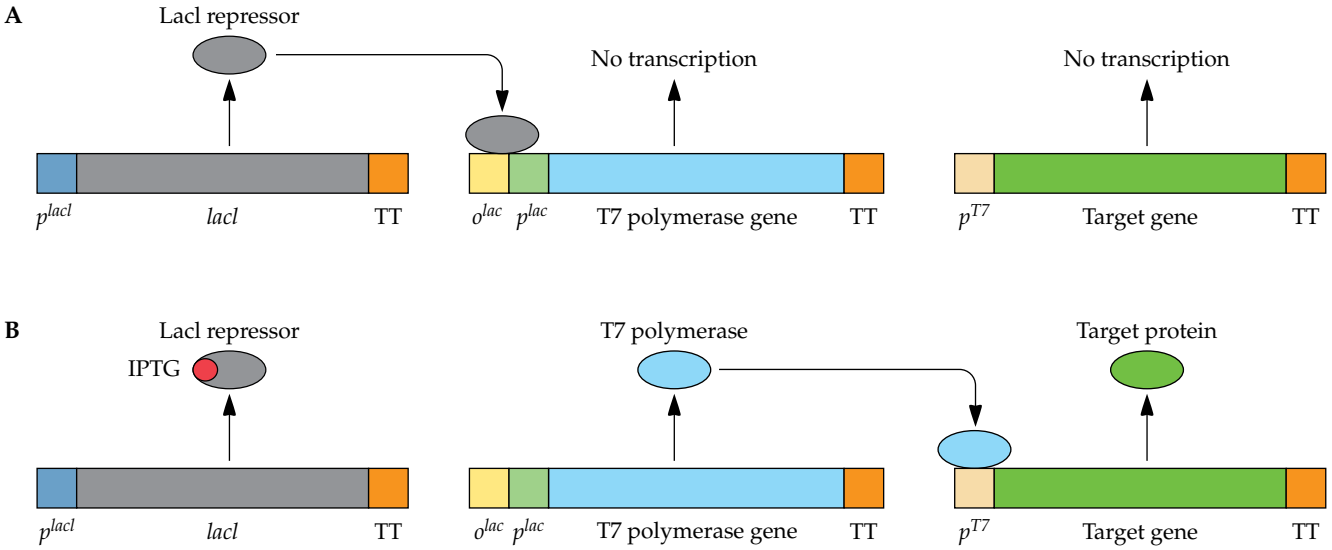


Figure 3.4 Regulation of target gene expression controlled by the promoter for gene 10 from bacteriophage T7 (p^{T7}). **(A)** In the absence of the inducer IPTG, the constitutively produced LacI repressor, the product of the *lacI* gene, which is under the control of the *lacI* promoter (p^{lacI}), represses the synthesis of the T7 RNA polymerase that is transcriptionally controlled by the *lac* operator (o^{lac}) and *lac* promoter (p^{lac}). In the absence of T7 RNA polymerase, the target gene, which is under the transcriptional control of the T7 gene 10 promoter (p^{T7}), is not transcribed. **(B)** When IPTG (or lactose) is added to the medium, it binds to the LacI repressor, thereby preventing it from repressing the transcription of T7 RNA polymerase. In the presence of T7 RNA polymerase, the target gene is transcribed. TT, transcription termination sequence.

of the newly synthesized target protein, may also affect the amount of product. In prokaryotic cells, various proteins are not necessarily synthesized with the same efficiency. In fact, they may be produced at very different levels (up to several hundredfold) even if they are encoded within the same polycistronic messenger RNA (mRNA). Differences in translational efficiency, as well as differences in transcriptional regulation, enable the cell to have hundreds or even thousands of copies of some proteins and only a few copies of others.

Ribosome-Binding Site

The molecular basis for differential translation is, in part, the presence of a translational initiation signal called a ribosome-binding site in the transcribed mRNA. A ribosome-binding site is a sequence of 6 to 8 nucleotides (e.g., UAAGGAGG) in mRNA that can base pair with a complementary sequence (AUUCCUCC for *E. coli*) on the 16S RNA component of the small ribosomal subunit. Generally, the stronger the binding of the mRNA to the ribosomal RNA, the greater the efficiency of translational initiation.

Many *E. coli* expression vectors have been designed to ensure that the mRNA of a cloned gene contains a strong ribosome-binding site. However, for proper translation of heterologous genes in *E. coli*, certain other conditions must also be satisfied. First, the ribosome-binding sequence must be located within a short distance (usually 2 to 20 nucleotides) from the translational start codon of the cloned gene. Second, after transcription, the 5′

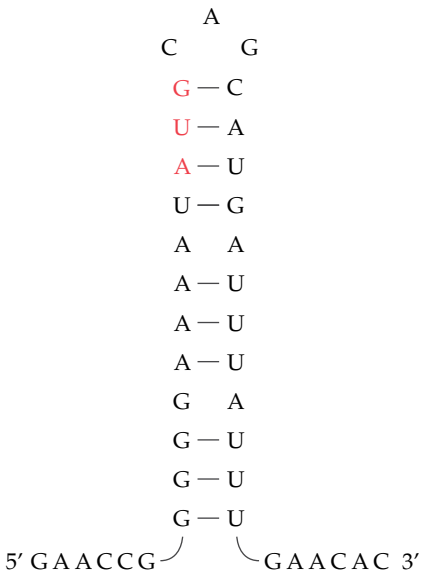


Figure 3.5 Example of the secondary structure of the 5′ end of an mRNA that would prevent efficient translation. The ribosome-binding site is GGGGG, the start codon is AUG (shown in red), and the first few codons are CAG-CAU-GAU-UUA-UUU. Note that in addition to the traditional A·U and G·C base pairs in mRNA, G can also base pair to some extent with U.

end of the mRNA sequence that includes the ribosome-binding site through the first few codons of the heterologous gene must not contain nucleotide sequences that have regions of complementarity. Intrastrand base pairing that leads to the formation of secondary structure in this region may shield the ribosome-binding site (Fig. 3.5) and therefore affect the extent to which the mRNA can bind to the appropriate sequence on the ribosome and initiate translation. Thus, for each cloned gene, it is important to establish that the ribosome-binding site is properly placed and that the secondary structure of the mRNA does not prevent its access to the ribosome.

A number of convenient vectors that incorporate both transcriptional and translational signals for the expression of cloned genes in *E. coli* have been developed. The commonly used pET series of expression vectors employ the T7 gene 10 promoter and transcription terminator to drive high levels of expression of a cloned gene in an *E. coli* host cell that produces T7 RNA polymerase (Fig. 3.4). In one version of the pET vector, a target gene is inserted into the multiple cloning site so that it is in the same reading frame as the ATG start codon provided on the vector. The strong ribosome-binding site from the T7 major capsid protein gene is present 8 nucleotides upstream from the start codon (Fig. 3.6). The *lac* operator is present between the promoter and the ribosome-binding site, and the gene encoding the LacI repressor is carried on the vector. After induction with IPTG and transcription, the mRNA of a cloned gene is efficiently translated. However, since the nucleotide sequences that encode the amino acids in the N-terminal region of the target protein vary from one gene to another, it is not possible to design a vector that will eliminate the possibility of mRNA secondary structure in all instances. Therefore, no single optimized translational initiation region can guarantee a high rate of translation initiation for all cloned genes. Consequently, the expression vectors described above are merely starting points for the optimization of translation initiation.

Optimizing Codon Usage

While the genetic code is universal among organisms, the codons that specify amino acids are used to different extents in various organisms (Table 3.4). Cells produce transfer RNAs (tRNAs) corresponding to a specific codon in approximately the same relative amount as that particular codon is used in the production of proteins. When a cloned gene has codons that are rarely used by the host cell, a cellular incompatibility

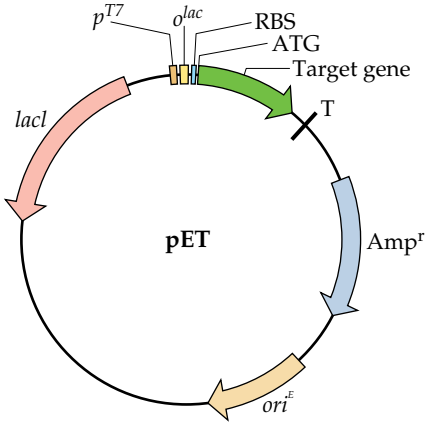


Figure 3.6 The expression vector pET. A target gene is inserted into the vector such that it is in the same reading frame as a start codon (ATG) present on the vector. Transcription of the target gene is controlled by the strong T7 gene 10 promoter (p^{T7}) and terminator (T), and translation is initiated from the high-affinity ribosome-binding site (RBS) from the T7 major capsid protein gene. In the absence of IPTG, the LacI repressor encoded by *lacI* binds to the *lac* operator sequence (o^{lac}) and blocks transcription of the target gene (see Fig. 3.4). The LacI repressor also prevents transcription of the T7 RNA polymerase gene that is integrated into the chromosome of host *E. coli* cells (not shown). The ampicillin resistance gene (Amp^r) is a selectable marker gene that confers resistance to the antibiotic ampicillin, and the origin of replication (ori^E) enables replication of the vector in *E. coli*.

Table 3.4 Genetic code and codon usage in *E. coli* and humans

Codon	Amino acid	Frequency of use in:		Codon	Amino acid	Frequency of use in:	
		E. coli	Humans			*E. coli*	Humans
GGG	Glycine	0.13	0.23	UAG	Stop	0.09	0.17
GGA	Glycine	0.09	0.26	UAA	Stop	0.62	0.22
GGU	Glycine	0.38	0.18	UAU	Tyrosine	0.53	0.42
GGC	Glycine	0.40	0.33	UAC	Tyrosine	0.47	0.58
GAG	Glutamic acid	0.30	0.59	UUU	Phenylalanine	0.51	0.43
GAA	Glutamic acid	0.70	0.41	UUC	Phenylalanine	0.49	0.57
GAU	Aspartic acid	0.59	0.44	UCG	Serine	0.13	0.06
GAC	Aspartic acid	0.41	0.56	UCA	Serine	0.12	0.15
GUG	Valine	0.34	0.48	UCU	Serine	0.19	0.17
GUA	Valine	0.17	0.10	UCC	Serine	0.17	0.23
GUU	Valine	0.29	0.17	AGU	Serine	0.13	0.14
GUC	Valine	0.20	0.25	AGC	Serine	0.27	0.25
GCG	Alanine	0.34	0.10	CGG	Arginine	0.08	0.19
GCA	Alanine	0.22	0.22	CGA	Arginine	0.05	0.10
GCU	Alanine	0.19	0.28	CGU	Arginine	0.42	0.09
GCC	Alanine	0.25	0.40	CGC	Arginine	0.37	0.19
AAG	Lysine	0.24	0.60	AGG	Arginine	0.03	0.22
AAA	Lysine	0.76	0.40	AGA	Arginine	0.04	0.21
AAU	Asparagine	0.39	0.44	CAG	Glutamine	0.69	0.73
AAC	Asparagine	0.61	0.56	CAA	Glutamine	0.31	0.27
AUG	Methionine	1.00	1.00	CAU	Histidine	0.52	0.41
AUA	Isoleucine	0.07	0.14	CAC	Histidine	0.48	0.59
AUU	Isoleucine	0.47	0.35	CUG	Leucine	0.55	0.43
AUC	Isoleucine	0.46	0.51	CUA	Leucine	0.03	0.07
ACG	Threonine	0.23	0.12	CUU	Leucine	0.10	0.12
ACA	Threonine	0.12	0.27	CUC	Leucine	0.10	0.20
ACU	Threonine	0.21	0.23	UUG	Leucine	0.11	0.12
ACC	Threonine	0.43	0.38	UUA	Leucine	0.11	0.06
UGG	Tryptophan	1.00	1.00	CCG	Proline	0.55	0.11
UGU	Cysteine	0.43	0.42	CCA	Proline	0.20	0.27
UGC	Cysteine	0.57	0.58	CCU	Proline	0.16	0.29
UGA	Stop	0.30	0.61	CCC	Proline	0.10	0.33

occurs that decreases translation efficiency. For example, AGG, AGA, AUA, CUA, and CGA are the least-used codons in *E. coli*. The host cell may not produce enough of the tRNAs that recognize these rarely used codons, and consequently the yield of the protein produced from the cloned gene will be much lower than expected. Any codon that is used less than 5% to 10% of the time by the host organism may cause problems. An insufficient supply of tRNAs may also lead to the incorporation of incorrect amino acids into the protein. Errors in the amino acid sequence of a protein may diminish its usefulness if the specific activity and stability are reduced.

To alleviate this problem, the target gene may be expressed in a different host or be chemically synthesized (chapter 2) or engineered by directed mutagenesis (described later in this chapter) to contain codons that are more commonly used by the host cell. Codon optimization has

Table 3.5 Increases in gene expression that result from altering the codon usage of the wild-type gene (or cDNA) to more closely correspond to the host *E. coli* cell[a]

Protein	Improvement (fold)
Human interleukin-2	16
Clostridium tetani tetanus toxin fragment C	4
Human cardiac troponin T	10–40
Mouse c-FOS protein	>200
Spinach plastocyanin	1.2
Human neurofibromin	3
Human glutathione transferase M2-2	140
Human phosphatidylcholine transfer protein	>100
Human interleukin-6	3
Human interleukin-18	5
Plasmodium vaccine candidate antigen	4

[a]In some cases, only a small number of codons were altered, while in others, the entire gene was synthesized with the optimal codon usage for expression in *E. coli*.

enabled the production of large quantities of a variety of heterologous proteins that are otherwise difficult to express at high levels in *E. coli* (Table 3.5). Alternatively, host cells that have been engineered to overexpress rare tRNAs are commercially available. Some contain plasmids such as pRARE that carry several rare tRNA genes (Fig. 3.7A). An *E. coli* strain (named SixPack) has additional copies of genes for the six least abundant tRNAs integrated into a ribosomal RNA operon in its chromosome (Fig. 3.7B). This strain expressed the six rare tRNAs at levels 1.8- to 3.8-fold higher than the parental strain. Expression of recombinant proteins from genes with a high ratio of rare codons (greater than 8%) was up to 20-fold higher in the engineered host.

Increasing Protein Stability

High levels of expression of some foreign proteins in bacterial hosts often result in the formation of insoluble, inactive protein aggregates known as inclusion bodies. Foreign proteins may misfold in a cellular environment in which the pH, osmolarity, and redox status are different from that of the natural host. Misfolding exposes hydrophobic amino acids that lead to aggregation of proteins, especially at high concentrations. Recovery of functional recombinant proteins from inclusion bodies is cumbersome, requiring optimization of solubilization and refolding steps, and yields are reduced. There are also large differences in the intrinsic stabilities of different proteins. Under normal growing conditions, the half-lives of different proteins range from a few minutes to hours. The basis for this differential stability is the extent of disulfide bond formation and the susceptibility to cleavage by proteases. With these features in mind, host strains and/or target genes can be engineered to increase the solubility and stability of recombinant proteins.

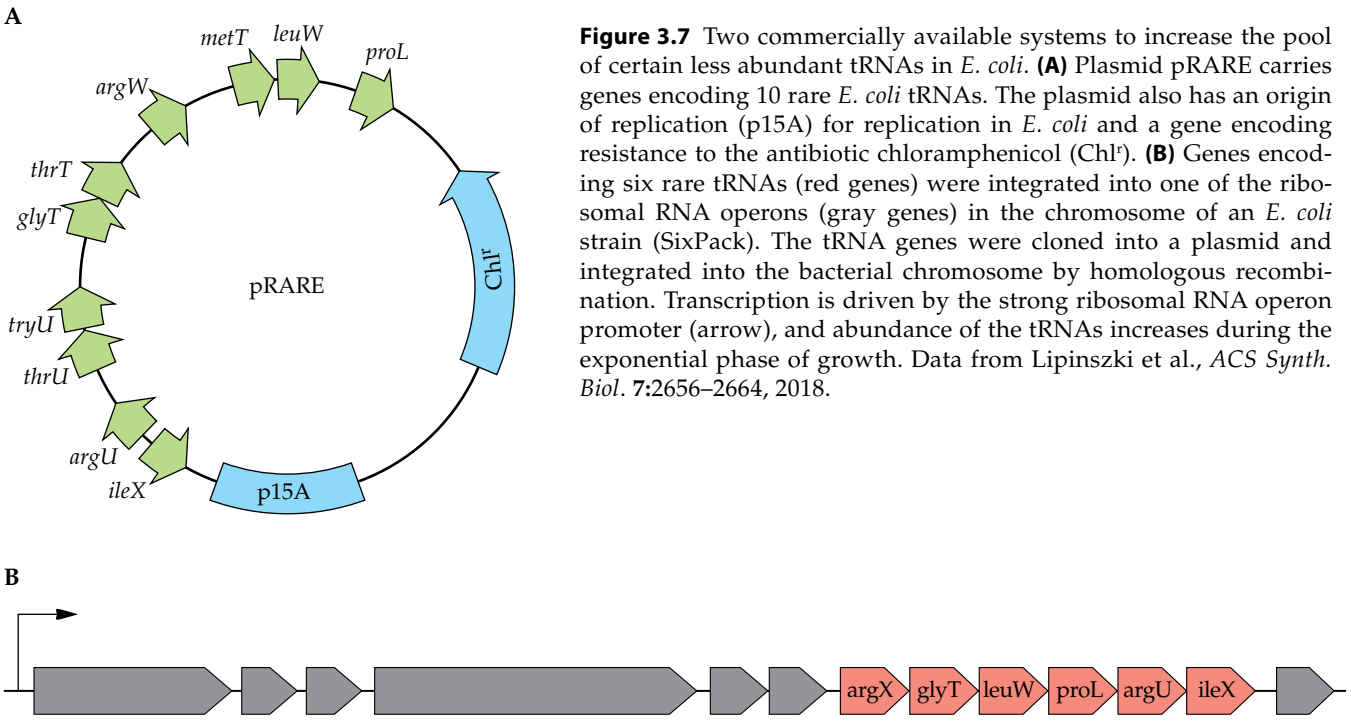


Figure 3.7 Two commercially available systems to increase the pool of certain less abundant tRNAs in *E. coli*. **(A)** Plasmid pRARE carries genes encoding 10 rare *E. coli* tRNAs. The plasmid also has an origin of replication (p15A) for replication in *E. coli* and a gene encoding resistance to the antibiotic chloramphenicol (Chl^r). **(B)** Genes encoding six rare tRNAs (red genes) were integrated into one of the ribosomal RNA operons (gray genes) in the chromosome of an *E. coli* strain (SixPack). The tRNA genes were cloned into a plasmid and integrated into the bacterial chromosome by homologous recombination. Transcription is driven by the strong ribosomal RNA operon promoter (arrow), and abundance of the tRNAs increases during the exponential phase of growth. Data from Lipinszki et al., *ACS Synth. Biol*. **7**:2656–2664, 2018.

Facilitating Protein Folding

Molecular chaperones. One simple strategy to increase the amount of soluble active protein is to cultivate recombinant strains at low temperatures, which facilitates proper protein folding. However, mesophilic bacteria such as *E. coli* grow extremely slowly at low temperatures. In one study, genes encoding the molecular chaperones chaperonin 60 and cochaperonin 10 from the psychrophilic bacterium *Oleispira antarctica* were introduced into a host strain of *E. coli* with the result that the *E. coli* strain gained the ability to grow at a high rate at low temperatures (4 to 10°C). This strain was subsequently transformed with a plasmid encoding a target protein, a temperature-sensitive esterase. The expression of the esterase in the *E. coli* strain carrying the two chaperone genes at 4 to 10°C yielded esterase-specific activity that was 180-fold higher than the activity from the native *E. coli* strain (without the heterologous chaperonins) grown at 37°C. Although very high levels of expression of the cloned esterase were not attained, this work illustrates an expression system for proteins that are sensitive to high temperature and might otherwise be difficult to produce.

While the psychrophile chaperones enhanced the growth of *E. coli* at low temperatures, they did not directly participate in proper folding of the foreign protein. It is also possible to coexpress the target gene with one or more molecular chaperones that interact with and mediate correct folding of proteins. *E. coli* produces several chaperones that function in protein folding. The "folding chaperones" utilize ATP cleavage to promote conformational changes that enable refolding of their substrates (Table 3.6). The "holding chaperones" bind to partially folded proteins until the folding chaperones have done their job. The "disaggregating chaperone" promotes the solubilization of proteins that have become aggregated. Protein folding also involves the "trigger factor," which binds to nascent polypeptide chains, acting as a holding chaperone. The proper folding of proteins has been facilitated by coexpression with some of these chaperones. In a study in which the chaperones DnaK and GroEL (and their cochaperonin protein molecules) were overexpressed, the yields of several target proteins expressed at the same time were increased up to 5-fold.

Table 3.6 *E. coli* proteins that facilitate the correct folding of recombinant proteins[a]

Localization	Function	Name
Cytoplasm	Holding chaperone	Hsp31 Hsp33 IbpA IbpB Trigger factor
	Folding chaperone	GroEL (Hsp60) DnaK (Hsp70) HscA HscC HtpG (Hsp90)
	Disaggregase	ClpB
	Secretory chaperone	SecB
Periplasm	Generic chaperones	Skp (OmpH) FkpA
	Specialized chaperones	SurA LolA PapD FimC
	Peptidyl-prolyl isomerases	SurA PpiD FkpA PpiA (RotA)
	Disulfide bond formation	DsbA DsbB DsbC DsbD DsbE DsbG CcmH

[a]Data from Baneyx and Mujacic, *Nat. Biotechnol.* **22:**1399–1408, 2004.

Fusion proteins. Another strategy to reduce the formation of inclusion bodies is to express the target protein as a fusion protein. Fusion proteins are constructed at the DNA level by ligating the coding regions of the target gene and a fusion partner (Table 3.7), such that a single polypeptide is

Table 3.7 Some fusion partners to increase the solubility of a recombinant protein

Fusion partner	Protein	Source organism
MBP	Maltose-binding protein	*E. coli*
Trx	Thioredoxin	*E. coli*
NusA	N-utilization substance	*E. coli*
SUMO	Small ubiquitin modifier protein	*Homo sapiens*
GST	Glutathione-*S*-transferase	*Schistosoma japonicum*
Fh8	8-kDa antigen	*Fasciola hepatica*
ZZ	IgG repeat domain ZZ of protein A	*Staphylococcus aureus*
EspA	Secreted protein A	*E. coli*

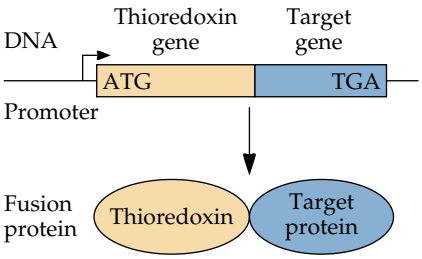
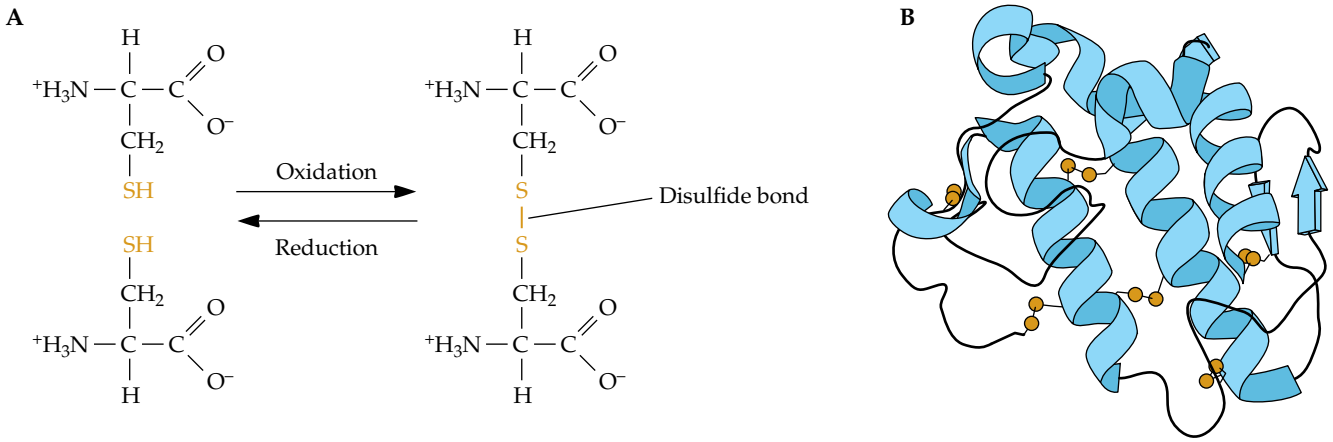


Figure 3.8 Expression of a soluble fusion protein. To avoid formation of insoluble inclusion bodies, the gene encoding the target protein is cloned adjacent to, and in the same reading frame as, the thioredoxin gene and expressed as a single polypeptide. ATG encodes the start codon; TGA encodes a stop codon; the arrow indicates the transcription start site.

synthesized. The fusion partner may have intrinsic chaperone activity, attract chaperones, or be highly acidic, thereby inhibiting protein aggregation through electrostatic repulsions. It is essential that the combined coding sequences have the correct reading frame; otherwise, an incomplete or an incorrect translation product will result, and the protein will not have the desired function. The target protein may be joined to the fusion partner at either the C- or N-terminal end, although in practice both are typically tested to determine the best orientation for production of a functional target protein, and it is sometimes necessary to try several different fusion partners to optimize solubility. Fusion proteins that contain thioredoxin, a small (12-kilodalton [kDa]), highly soluble protein, as the fusion partner remain soluble even when up to 40% of the cellular protein consists of the fusion protein. With this system, the target gene is cloned just downstream from the thioredoxin gene and both genes are transcribed from a single promoter (Fig. 3.8). Fusion proteins containing thioredoxin accumulate preferentially at the cytoplasmic face of the host *E. coli* inner membrane at sites known as adhesion zones (regions where the inner and outer membranes adhere). This facilitates selective release of the soluble fusion protein by osmotic shock from *E. coli* cells into the growth medium. The presence of the fusion partner makes most fusion proteins unsuitable for clinical use and may affect the biological functioning of the target protein. In addition, fusion proteins require more extensive testing before being approved by regulatory agencies, such as the U.S. Food and Drug Administration. Thus, strategies have been developed to remove the fusion partner from the target protein following purification (described later in this chapter).

Disulfide bonds. Correct disulfide bond formation is essential for many proteins to fold properly and achieve an active configuration (Fig. 3.9). Covalent disulfide bonds form by oxidation of sulfhydryl groups on cysteine amino acids that are adjacent in the three-dimensional structure of the folded protein, a reaction that is catalyzed by periplasmic (DsbA and DsbC) and membrane-bound (DsbB and DsbD) enzymes in *E. coli*. Disulfide bond formation in the reducing environment of the cytoplasm is

Figure 3.9 Disulfide bond in a protein. **(A)** A covalent, disulfide bond forms by oxidation of sulfhydryl (SH) groups on cysteines. **(B)** Disulfide bonds between cysteines (represented by brown circles) within a polypeptide (ribbon diagram) contribute to the structural stability of the protein.



rare. Thus, foreign proteins that tend to form inclusion bodies may be directed to the oxidizing environment of the periplasm (described later in this chapter). Overexpression of the disulfide bond isomerase DsbC also promotes disulfide bond formation. This was demonstrated for the human therapeutic protein tissue plasminogen activator, a 527-amino acid serine protease that requires formation of 17 disulfide bonds to attain an active state. The cDNA for this protein was cloned downstream of a DNA sequence that encodes a short N-terminal sequence (signal peptide) to facilitate secretion to the periplasm in *E. coli*. However, only trace amounts of the protein were produced. Coexpression of high levels of periplasmic DsbC resulted in more than a 100-fold increase in the production of functional human tissue plasminogen activator. To realize the maximum benefit from DsbC overproduction, it was necessary to induce the synthesis of this protein approximately 30 minutes prior to the induction of human tissue plasminogen activator. Alterations in the levels of the other Dsb proteins did not affect the amount of active human tissue plasminogen activator that could be recovered; however, overproduction of all four Dsb proteins yielded the greatest amount of properly folded and active recombinant horseradish peroxidase.

A host *E. coli* strain was engineered to enable formation of disulfide bonds in recombinant proteins that are produced in the cytoplasm (Fig. 3.10). To create suitable oxidizing conditions in the cytoplasm, the genes encoding glutathione reductase (*gor*) and thioredoxin reductase (*trxB*) that normally maintain cysteines in cytoplasmic proteins in their reduced state were deleted (note that the viability of the bacterium required a compensating mutation in the peroxidase AhpC to reduce some essential proteins). In addition, from a chromosomally integrated *dsbC* gene, DsbC was overexpressed without a signal peptide, so it remained in the cytoplasm. When produced in wild-type *E. coli*, the activity of three recombinant enzymes, luciferase, urokinase, and tissue plasminogen activator, which have 10, 18, and 12 disulfide bond-forming cysteines, respectively, was very low or undetectable. The activity of the enzymes increased substantially when expressed in the oxidizing cytoplasm of the DsbC-overproducing host strain. Numerous studies have used this *E. coli* strain to produce high yields of recombinant proteins that require disulfide bonds for folding and activity.

Decreasing Protein Degradation

Bacterial cells produce several proteases that regulate protein levels by degrading those that are damaged or are no longer needed. The amino acids at the N terminus of a protein can influence its stability, and therefore engineering a protein to include specific N-terminal amino acids may render it less susceptible to degradation by cellular proteases. For example, alteration of the N-terminal amino acids increased the *in vitro* survival time of β-galactosidase from approximately 2 minutes to more than 20 hours (Table 3.8). Amino acid additions that extend the intrinsic survival of a protein can be readily incorporated into cloned genes. Often the presence of a single extra amino acid at the N-terminal end is sufficient to stabilize a target protein. Long-lived proteins can accumulate in cells and thereby increase the yield of the product. This phenomenon occurs in both prokaryotes and eukaryotes.

Specific internal amino acid sequences can also make a protein more susceptible to proteolytic degradation. For example, proteins that contain

Table 3.8 Stability of β-galactosidase with certain amino acids added to its N terminus[a]

Amino acid added	Half-life
Met, Ser, Ala	>20 h
Thr, Val, Gly	>20 h
Ile, Glu	>30 min
Tyr, Gln	~10 min
Pro	~7 min
Phe, Leu, Asp, Lys	~3 min
Arg	~2 min

[a]Data from Bachmair et al., *Science* **234:** 179–186, 1986.

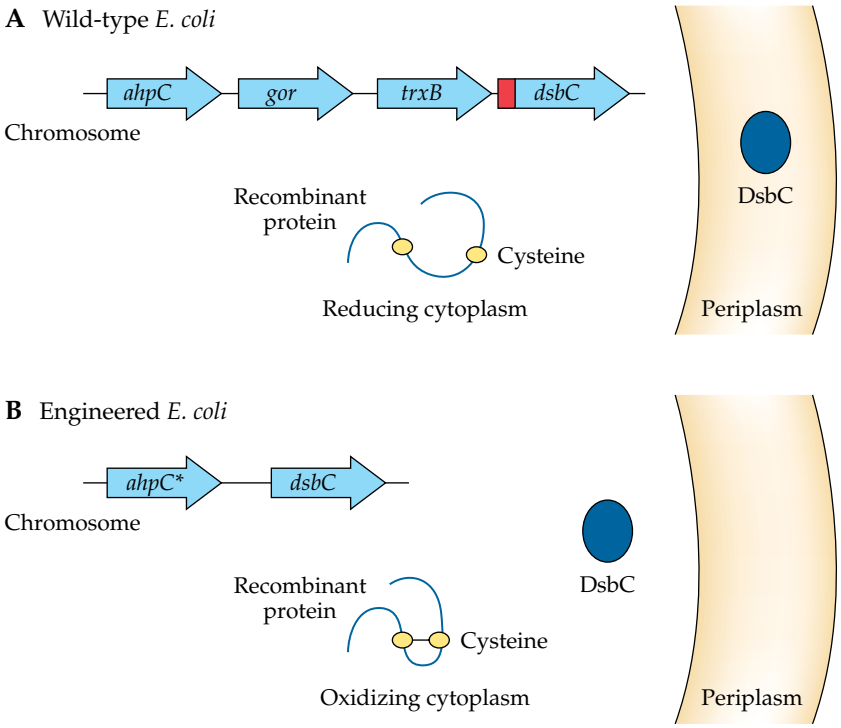


Figure 3.10 An engineered *E. coli* host strain catalyzes the formation of disulfide bonds in cytoplasmic recombinant proteins. **(A)** Wild-type *E. coli* produces reductases (from genes *gor* and *trxB*) that maintain cysteines in cytoplasmic proteins in their reduced state, in which they do not form disulfide bonds (Fig. 3.9). Disulfide bond isomerase DsbC is expressed with a signal peptide (encoded by the DNA sequence represented in red) that directs it to the inner membrane for translocation to the periplasm (Fig. 3.11). **(B)** In the genetically modified *E. coli* strain, the *gor* and *trxB* genes were deleted from the chromosome to create oxidizing conditions in the cytoplasm that are suitable for disulfide bond formation. The lethality of this condition was overcome by altering AhpC (encoded by *ahpC**) to reduce essential proteins. DsbC is overexpressed without a signal peptide and is therefore retained in the cytoplasm to catalyze disulfide bond formation between some cysteines in a protein. Note that the position of the genes in the chromosome is represented schematically.

PEST sequences, which are rich in proline (P), glutamic acid (E), serine (S), and threonine (T) and are often flanked by clusters of positively charged amino acids such as lysine and arginine, are marked for degradation within the cell. In some instances, it is possible to enhance the stability of a protein by altering its PEST regions by genetic manipulation. Such changes, of course, must not alter the function of the target protein.

In a number of studies, proteins synthesized from cloned genes have been found to be more resistant to degradation by host cell proteases when they are part of a fusion protein, in which the fusion partner is not especially susceptible to proteolysis. Alternatively, stable foreign proteins can be produced in bacterial host strains that are deficient in the production of proteases. However, this is not as simple as it might appear, as *E. coli* has at least 25 different proteases. These proteases are important for the degradation of abnormal or defective proteins, which is a housekeeping function that is necessary for the continued viability of the cells. In one study, strains

with mutations in one or more protease genes were constructed. Generally, the strains that were most deficient in overall protease activity grew most slowly. Thus, decreasing protease activity caused cells to be debilitated. However, an *E. coli* strain with mutations in both the gene for the RNA polymerase sigma factor that is responsible for heat shock protein synthesis (*rpoH*) and the gene for a protease that is required for cell growth at high temperatures (*degP*) secreted target proteins that had a 36-fold-greater specific activity than when they were produced in wild-type host cells. This increase in activity reflects a decrease in the proteolytic degradation of these secreted proteins.

Increasing Protein Secretion

E. coli and other Gram-negative bacteria have multiple pathways for the secretion of various proteins to the periplasmic space between the inner (cytoplasmic) and outer membranes or outside the cell into the surrounding environment. Directing a foreign protein to the periplasm or the growth medium makes its purification easier and less costly (see chapter 9), as many fewer proteins are present outside the cytoplasmic membrane than in the cytoplasm. Moreover, the stability of a recombinant protein depends on its cellular location in *E. coli*. For example, recombinant proinsulin is approximately 10 times more stable if it is exported into the periplasm than if it is localized in the cytoplasm. In addition, secretion of proteins to the periplasm facilitates the correct formation of disulfide bonds because the periplasm provides an oxidative environment, in contrast to the more reducing environment of the cytoplasm. Table 3.9 shows the amounts of secreted recombinant pharmaceutical protein attainable in various bacterial systems.

Secretion into the Periplasm

Secretion via the Sec pathway. Many proteins destined for the periplasm are secreted via the general secretion pathway, which consists of a secretion complex (Sec translocon) in the cytoplasmic membrane and trafficking proteins that selectively deliver proteins to the complex (Fig. 3.11). An amino acid sequence called the signal peptide (also called

Table 3.9 Yields of several secreted recombinant proteins produced in different bacteria

Protein	Yield (per liter)	Host bacterium
Hirudin	>3 g	*E. coli*
Human antibody fragment	1–2 g	*E. coli*
Human insulin-like growth factor	8.5 g	*E. coli*
Monoclonal antibody 5T4	700 mg	*E. coli*
Humanized anti-CD18 F(ab′)2	2.5 g	*E. coli*
Human epidermal growth factor	325 mg	*E. coli*
Alkaline phosphatase	5.2 g	*E. coli*
Staphylokinase	340 mg	*B. subtilis*
Human proinsulin	1 g	*B. subtilis*
Human calcitonin precursor	2 g	*Staphylococcus carnosus*
Organophosphohydrolase	1.2 g	*Ralstonia eutropha*
Human CD4 receptor	200 mg	*Streptomyces lividans*
Human insulin	100 mg	*Streptomyces lividans*

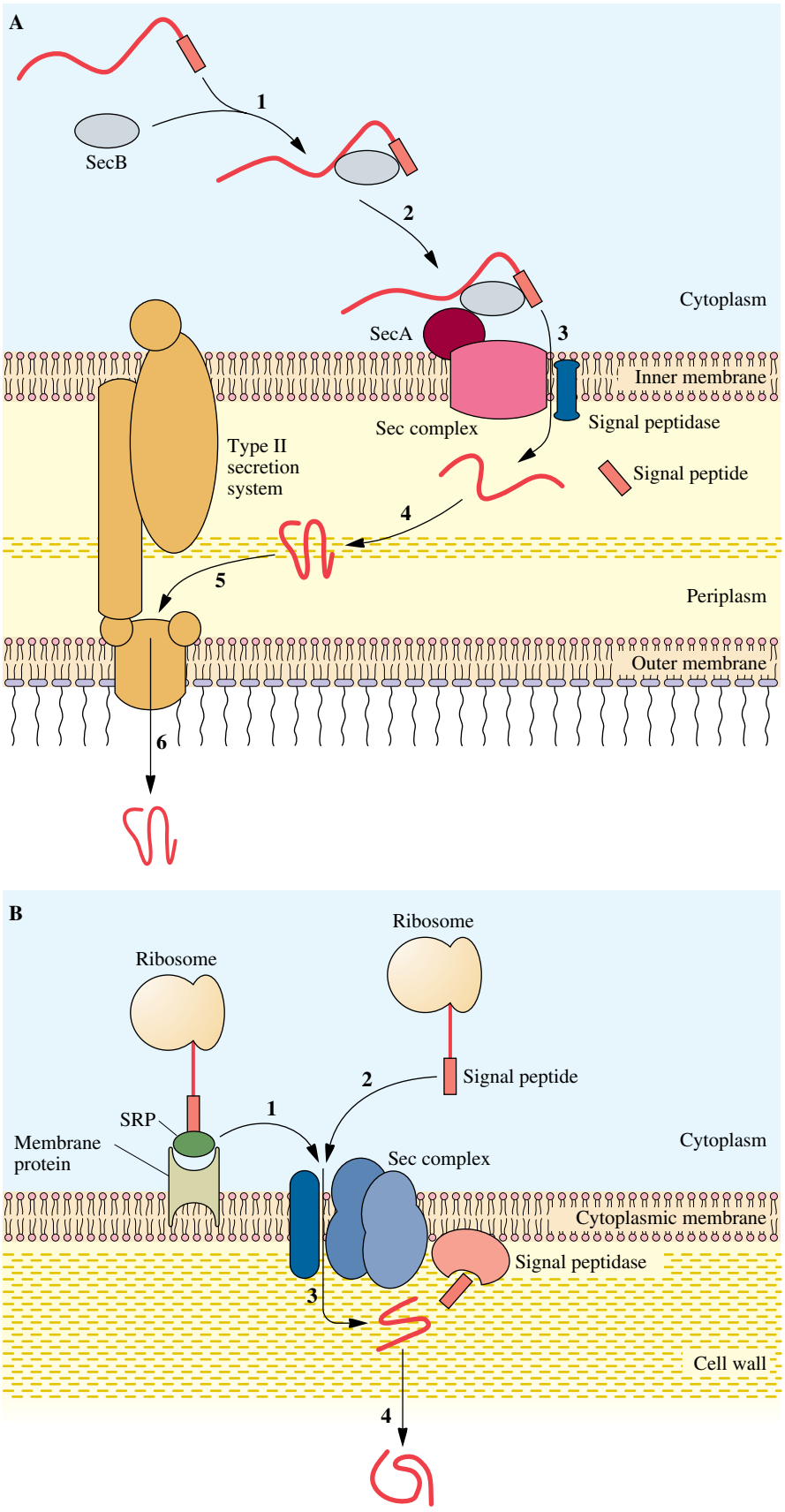
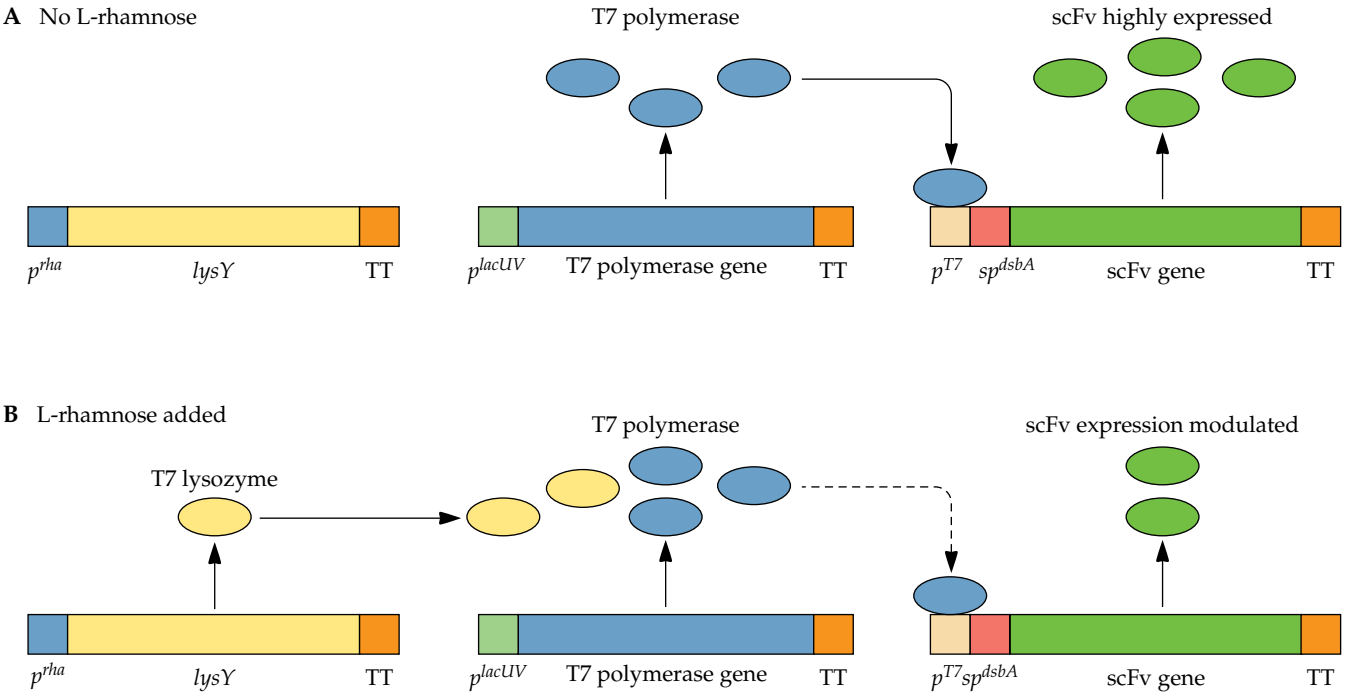


Figure 3.11 Protein secretion in bacteria. **(A)** General secretion pathway in Gram-negative bacteria. The SecB protein binds to the N-terminal signal peptide (red rectangle) of a secretory protein in the cytoplasm (1). SecB delivers the protein to the SecA protein that is associated with the Sec complex (SecYEG translocon) in the inner membrane (2), and the secretory protein is translocated across the inner membrane (3). A signal peptidase removes the signal peptide, and the secretory protein is properly folded in the periplasm (4). Some secretory proteins are directed to a protein complex (type II secretion system) in the outer membrane (5) and are translocated to the external environment (6). **(B)** Protein secretion in Gram-positive bacteria. A signal recognition particle (SRP) binds to the signal peptide of a secretory protein, and this complex binds to a membrane protein that directs the secretory protein (1) to the Sec complex. There is also an SRP-independent pathway (2), where a signal peptide alone makes contact with the Sec complex. The secretory protein is translocated through a channel within the Sec complex (3), and the signal peptide is removed by a signal peptidase. Proper folding of the secretory protein occurs as it passes through the cell wall (4).

the signal sequence or leader peptide), located at the N-terminal end of a secretory protein, facilitates its export through the secretion complex. It is sometimes possible to engineer a protein for secretion to the periplasm by adding the DNA sequence encoding a signal peptide to the cloned gene. When the recombinant protein is secreted into the periplasm, the signal peptide is precisely removed by a signal peptidase that is a component of the secretion complex so that the N-terminal end of the recombinant protein is identical to that of the natural protein.

A recombinant single-chain antibody fragment (scFv; see chapter 5) was targeted to the periplasm via the Sec translocon by fusing it to the signal peptide from DsbA (Fig. 3.12A). Transcription of the gene encoding the fusion protein was driven by T7 RNA polymerase, which is expressed from the IPTG-inducible *lacUV* promoter. Despite strong expression, periplasmic levels of the recombinant protein were low, possibly due to saturation of the Sec translocon. When the capacity of the Sec translocon to secrete proteins is exceeded, the proteins tend to aggregate and form inclusion bodies

Figure 3.12 Strategy to increase secretion of a recombinant protein to the periplasm. Recombinant single-chain antibody fragments (scFv) were directed to the *E. coli* periplasm by the N-terminal signal peptide from periplasmic protein DsbA (encoded by sp^{dsbA}). **(A)** High levels of the recombinant fusion protein expressed from the strong T7 promoter (p^{T7}) overwhelmed the protein secretion apparatus (Fig. 3.11A), reducing periplasmic levels of scFv. **(B)** Addition of L-rhamnose to the culture medium induced expression of T7 lysozyme encoded by the *lysY* gene, which is under the control of the promoter from the L-rhamnose catabolic operon (p^{rha}). T7 lysozyme is a natural inhibitor of T7 RNA polymerase. The levels of T7 lysozyme, and therefore the levels of scFv, were controlled by the amount of L-rhamnose added, which can be adjusted to optimize secretion of the recombinant protein to the periplasm. TT, transcription termination sequence.



in the cytoplasm. Saturation of the secretion apparatus can be avoided by modulating expression of the secretory protein. To this end, the activity of the T7 RNA polymerase, and hence expression of the recombinant scFv protein, was regulated with T7 lysozyme, which inhibits T7 RNA polymerase (Fig. 3.12B). The T7 lysozyme gene was expressed from an L-rhamnose-inducible promoter such that levels of T7 lysozyme were precisely controlled by the amount of L-rhamnose added to the culture medium. This resulted in a substantial increase in the amount of scFv secreted by host cells following induction of T7 lysozyme with 500 μM L-rhamnose compared to uninduced cells. Moreover, in contrast to uninduced cultures, cell biomass was not reduced in the cultures in which expression of scFv was controlled.

In many instances, when foreign proteins engineered for secretion are overproduced, the precursor form is only partially processed, with about half of the secreted proteins retaining the signal peptide and the other half being fully processed to the mature form. This is probably the result of overloading some of the components involved in the secretion process. If this is the case, then it might be possible to increase the ratio of processed to unprocessed proteins by increasing the level of expression of some of the limiting components of the protein secretion pathway. This hypothesis was tested in a series of experiments in *E. coli* carrying the plasmid-borne genes *secY* and *secE* that encode components of the SecYEG translocon, the integral transmembrane complex of the general secretion pathway (Fig. 3.11A). Following overexpression of these proteins, the fraction of the target protein (in this case, the cytokine interleukin-6) that was secreted to the periplasm as the processed form with the signal peptide removed increased from about 50% to more than 90%.

Secretion via the Tat pathway. Recombinant proteins may also be directed to the periplasm via a second *E. coli* secretion system, the twin-arginine translocase (Tat). This inner membrane protein complex exports proteins with the N-terminal signal peptide sequence arginine-arginine-X-phenylalanine-X-lysine (Arg-Arg-X-Phe-X-Lys, where X is any amino acid). In one engineered *E. coli* host strain, the chromosomal operon (*tatABCD*) encoding the translocase was expressed from the strong, IPTG-inducible *tac* promoter (Fig. 3.13A). Recombinant proteins are directed to the Tat system by fusing the signal sequence from the Tat-dependent protein TorA to the N terminus of the target protein (Fig. 3.13B). High levels (more than 30 g/liter) of human growth hormone and a single-chain antibody fragment have been produced using this system.

Secretion into the Medium

Recovery of recombinant proteins on a large scale is easier and more cost-effective if they are present in the culture medium; however, the outer membrane of Gram-negative bacteria restricts the secretion of proteins into the extracellular environment. One solution is to produce the recombinant protein in a Gram-positive bacterium such as *Bacillus subtilis*, which lacks an outer membrane and therefore can secrete proteins directly into the medium (Fig. 3.11B). Alternatively, the host cell can be engineered to increase membrane permeability, or the recombinant protein can be fused to a protein that is naturally secreted to the extracellular medium by a Gram-negative bacterium.

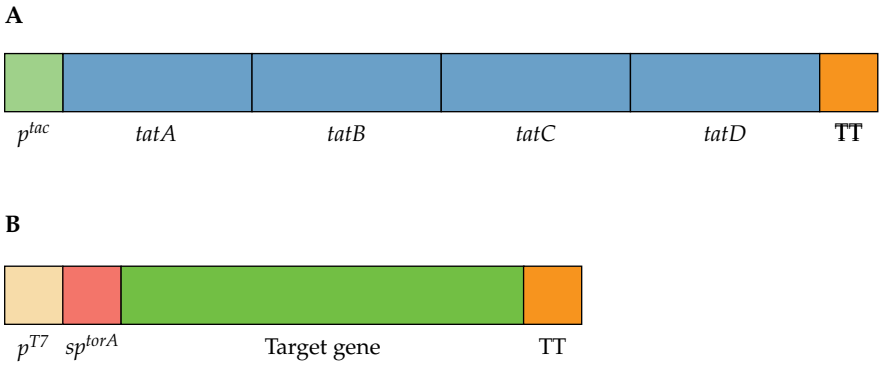


Figure 3.13 Twin-arginine translocation (Tat) pathway for periplasmic secretion of recombinant proteins. **(A)** Production of the Tat translocase was increased in an *E. coli* host strain by inserting the strong, regulatable *tac* promoter (p^{tac}) upstream of the *tatABCD* operon in the chromosome. **(B)** To direct a recombinant protein to the Tat translocase, the target gene is fused to the sequence for the N-terminal twin-arginine signal peptide from TorA (sp^{torA}). p^{T7}, T7 promoter; TT, transcription termination sequence.

Membrane permeabilization. Proteins that naturally permeabilize membranes may be employed to release recombinant proteins into the extracellular medium. One such protein, bacteriocin release protein, is normally used by some Gram-negative bacteria to secrete a bacteriocidal protein called a bacteriocin. Bacteriocin release protein activates phospholipase A, which is present in the bacterial inner membrane and cleaves membrane phosopholipids so that both the inner and outer membranes are permeabilized. Some cytoplasmic and periplasmic proteins are also released into the culture medium. Thus, by expressing the bacteriocin release protein gene from a strong regulatable promoter, *E. coli* cells may be permeabilized at will. The cells that carry the bacteriocin release protein gene are transformed with a plasmid carrying a cloned target gene that has been fused to a secretion signal peptide sequence for transport to the periplasm. The cloned gene is placed under the same transcriptional-regulatory control as the bacteriocin release protein gene so that the two genes can be induced simultaneously. Once in the periplasm, the target protein is released to the external medium via the permeabilized outer membrane (Fig. 3.14).

Extracellular protein secretion systems. Relatively few proteins are secreted to the growth medium by Gram-negative bacteria. In some strains, proteins destined for the extracellular environment may first be secreted into the periplasm by the Sec or Tat pathway and then across the outer membrane by the last step of the general secretion pathway (known as the type II secretion system) (Fig. 3.11A), or they may be secreted directly from the cytoplasm via specialized secretion systems such as the type I and III secretion systems found in many pathogens. The export signals of proteins naturally translocated by these systems may be exploited to carry recombinant proteins to the culture medium.

In the type III secretion system, an N-terminal signal sequence directs a protein, usually via a specific chaperone, to the secretion apparatus

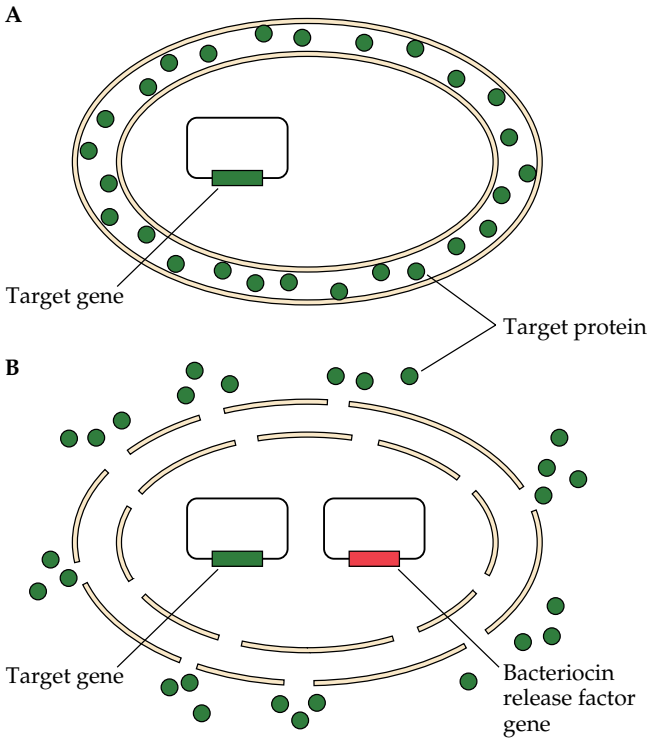


Figure 3.14 Membrane permeabilization to release recombinant proteins. **(A)** *E. coli* cells are engineered to secrete a foreign protein to the periplasm by fusing the target gene (green) to a sequence encoding a secretion signal peptide. **(B)** The periplasmic proteins are then released to the growth medium by permeabilizing cell membranes with a bacteriocin release protein. The gene encoding the bacteriocin release protein (red) is introduced into the host cell on a second plasmid and expressed from an inducible promoter.

for secretion though a needle-like structure that traverses the inner and outer membranes (Fig. 3.15A). To export recombinant proteins via the well-characterized *Salmonella enterica* SPI-1 type III secretion system, the target protein is fused to the N-terminal sequence from the SptP protein that is naturally secreted through this system, which consists of a 15-amino-acid secretion signal and a 152-amino-acid chaperone-binding domain (Fig. 3.15B). The secretion signal is not cleaved during secretion, and therefore a protease recognition site is included between the signal and target protein sequences so that the signal can be removed after secretion. The DNA sequence encoding the fusion protein is cloned into a plasmid vector with the gene encoding the chaperone (SicP) that delivers the SptP-tagged protein to the secretion apparatus, and both are expressed from a native type III secretion system promoter. Using this system, up to 14 mg/liter of spider silk protein monomers were secreted to the extracellular medium. Similar levels were attained for recombinant resilin and elastin, biopolymers with industrial applications.

Type III secretion systems are found in pathogenic bacteria, which are not ideal hosts for the production of recombinant proteins. However, a complete type III secretion system comprises approximately 30 structural and regulatory proteins and therefore is difficult to transfer to a nonpathogenic host strain. In addition, expression of the proteins is strictly regulated such that they are produced only when required for pathogenesis. To illustrate how these drawbacks can be overcome, the type III secretion system of *S. enterica* was modified in two major ways. First, genes encoding nonessential components of the secretion system were removed, which reduced the size of the locus to half. Codon usage was optimized in the remaining genes, and these were reorganized in

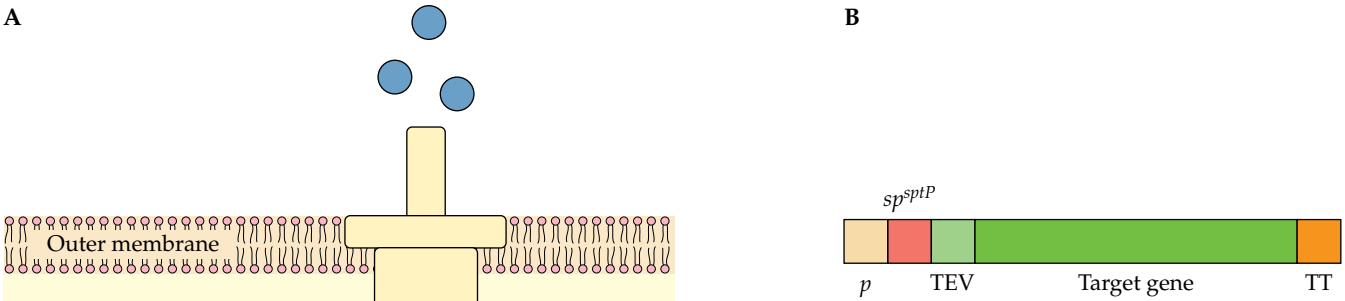


Figure 3.15 Extracellular secretion of recombinant proteins via the type III secretion system. **(A)** The type III secretion system is a protein complex that spans the inner and outer membranes of a Gram-negative bacterium. Proteins with a specific N-terminal signal peptide are delivered to the secretion apparatus by a specific chaperone and are exported to the culture medium. **(B)** Recombinant proteins expressed from genes that possess a sequence encoding the N-terminal signal peptide from SptP (sp^{sptP}), a natural type III secretion system substrate, are bound by the chaperone SicP. Following secretion, the signal peptide may be cleaved by the *Tobacco etch virus* (TEV) protease if the DNA sequence for the protease recognition site is placed between the signal peptide coding sequence and the target gene. *p*, promoter; TT, transcription termination sequence.

two operons, with genes encoding proteins required at high levels placed at the 5′ ends of the operon. Second, expression of the genes was uncoupled from natural, complex expression conditions by replacing native transcription and translation signals with synthetic promoters and ribosome-binding sites. The two engineered operons are controlled independently by two variants of T7 RNA polymerase that bind to different promoter sequences and are each controlled by different inducers (Fig. 3.16). One operon is transcribed by the T3 variant of T7 RNA polymerase that is expressed from the IPTG-inducible *tac* promoter, and the other is transcribed by the K1F variant of T7 RNA polymerase that is expressed from the acyl homoserine lactone-inducible *lux* promoter. This enables control of the timing and relative expression levels of the proteins encoded in the two operons. Upon induction with IPTG and acyl homoserine lactone, the modified system successfully secreted to the extracellular medium a protein that is a natural substrate of this type III secretion system. This demonstrates the functionality of a minimal type III secretion system that could be transferred to a nonpathogenic host and controlled to express and secrete recombinant proteins that can be harvested from the culture medium.

Facilitating Protein Purification

Affinity Purification Systems

A number of purification tags have been developed to simplify the purification of recombinant proteins (Table 3.10). The basis for this approach is the expression of a target protein as a fusion protein with a

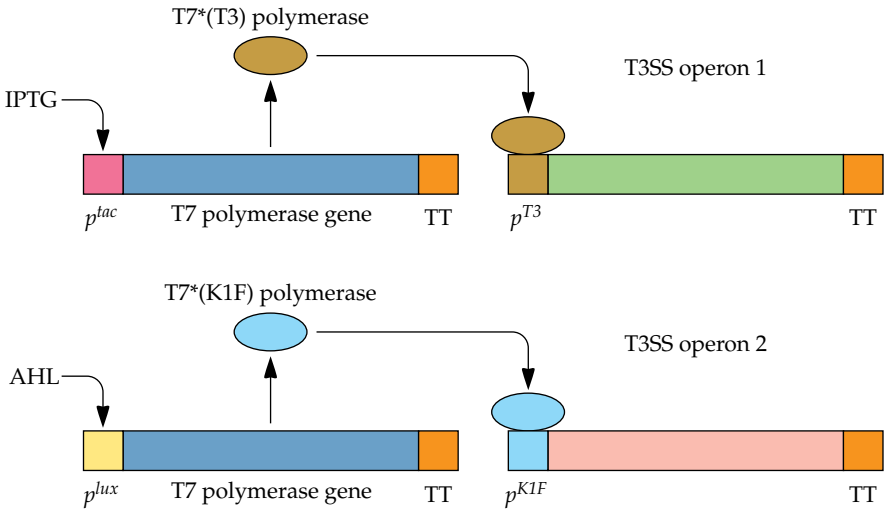


Figure 3.16 Control of a minimal type III secretion system (T3SS). Genes encoding only the proteins that are essential for protein secretion are organized in two operons, each transcribed from promoters (p^{T3} and p^{K1F}) that are recognized by different variants (T3 and K1F) of T7 RNA polymerase. Expression of T7 RNA polymerase variant T3 is induced from the *tac* promoter (p^{tac}) by IPTG, and expression of T7 RNA polymerase variant K1F is induced from the *lux* promoter (p^{lux}) by an acyl homoserine lactone (specifically, *N*-(β-ketocaproyl)-L-homoserine lactone; AHL). TT, transcription termination sequence. Data from Song et al., *Nat. Commun*. **8:**14737, 2017.

short peptide sequence that has a high affinity for a protein, antibody, carbohydrate, or other ligand; for this reason, the peptide tags are often referred to as affinity tags. For example, the coding sequence for a target protein such as the human interleukin-2 (an immune signaling protein with therapeutic applications) is joined to DNA encoding the affinity tag sequence Asp-Tyr-Lys-Asp-Asp-Asp-Asp-Lys (where Asp, Tyr, and Lys indicate aspartic acid, tyrosine, and lysine, respectively) that has the dual function of reducing the degradation of the expressed interleukin-2 and enabling the product to be purified. Following expression, the target protein can be purified by immunoaffinity chromatography. Monoclonal antibodies that bind to the affinity tag are immobilized on a polypropylene (or other solid) support and specifically capture the fusion protein (Fig. 3.17). In this case, the fusion protein was found to have the same biological activity as native interleukin-2.

In many instances antigen–antibody complexes that form during the immunoaffinity process are difficult to separate without the use of denaturing chemicals. As an alternative, it has become very popular to express a protein with an affinity tag that contains six or eight histidines attached to either the N- or C-terminal end of the target protein. Following expression and host cell lysis, the histidine-tagged protein, along with other cellular proteins, is then passed over an affinity column of nickel–nitrilotriacetic acid. The histidine-tagged protein, but not the other cellular proteins, binds tightly to the nickel ions that are immobilized on the column by nitrilotriacetic acid. The bound protein is subsequently eluted from the column by the addition of imidazole (the side chain of histidine). With this protocol,

Table 3.10 Some protein fusion systems used to facilitate the purification of foreign proteins in *E. coli* and other host organisms[a]

Fusion partner	Size	Ligand	Elution conditions
ZZ	14 kDa	Immunoglobulin G	Low pH
Histidine tail	6–10 amino acids	Ni^{2+}	Imidazole
Strep tag	10 amino acids	Streptavidin	Iminobiotin
Pinpoint	13 kDa	Streptavidin	Biotin
Maltose-binding protein	40 kDa	Amylose	Maltose
GST	26 kDa	Glutathione	Reduced glutathione
Flag	8 amino acids	Specific monoclonal antibody (MAb)	EDTA or low pH
Poly-arginine	5–6 amino acids	SP-Sephadex	High salt at pH of >8.0
c-myc	11 amino acids	Specific MAb	Low pH
S tag	15 amino acids	S fragment of RNase A	Low pH
Calmodulin-binding peptide	26 amino acids	Calmodulin	EGTA and high salt
Cellulose-binding domain	4–20 kDa	Cellulose	Urea or guanidine hydrochloride
Chitin-binding domain	51 amino acids	Chitin	SDS or guanidine hydrochloride
SBP tag	38 amino acids	Streptavidin	Biotin

[a]ZZ, a fragment of *Staphylococcus aureus* protein A; Strep tag, a peptide with affinity for streptavidin; Pinpoint, a protein fragment that is biotinylated and binds streptavidin; GST, glutathione *S*-transferase; Flag, a peptide recognized by enterokinase; c-myc, a peptide from a protein that is overexpressed in many cancers; S tag, a peptide fragment of ribonuclease (RNase) A; SBP (streptavidin-binding protein), a peptide with affinity for streptavidin; SP-Sephadex, a cation-exchange resin composed of sulfopropyl groups covalently attached to Sephadex beads.

some cloned and overexpressed proteins have been purified up to 100-fold with greater than 90% recovery in a single step.

Proteolytic Removal of Purification Tags

While purification tags and other fusion partners may not disrupt the function of a protein, to satisfy the government agencies that regulate the use of pharmaceuticals, it is necessary to remove these sequences if the product is to be used for human immunotherapy or other medical purposes. One way to do this is to join the gene for the target protein to the DNA sequence for the affinity tag/fusion partner with oligonucleotides that encode short stretches of amino acids that are recognized by a specific nonbacterial protease. For example, a sequence encoding the amino acid sequence isoleucine-glutamic acid-glycine-arginine (Ile-Glu-Gly-Arg) can be inserted between the target and fusion partner sequences. Following synthesis and purification of the fusion protein, the blood coagulation protein factor X_a can be used to release the target protein from the fusion partner, because factor X_a is a specific protease that cleaves peptide bonds uniquely on the C-terminal side of the Ile-Glu-Gly-Arg sequence (Fig. 3.18). Moreover, because this peptide sequence occurs rather infrequently in native proteins, this approach can be used to recover many different cloned gene products. The proteases most commonly used to cleave a fusion partner/affinity tag from a target protein are enterokinase, *Tobacco etch virus* protease, thrombin, and factor X_a. However, following this cleavage, it is necessary to perform additional purification steps to separate both the protease and the cleaved peptide from the target protein.

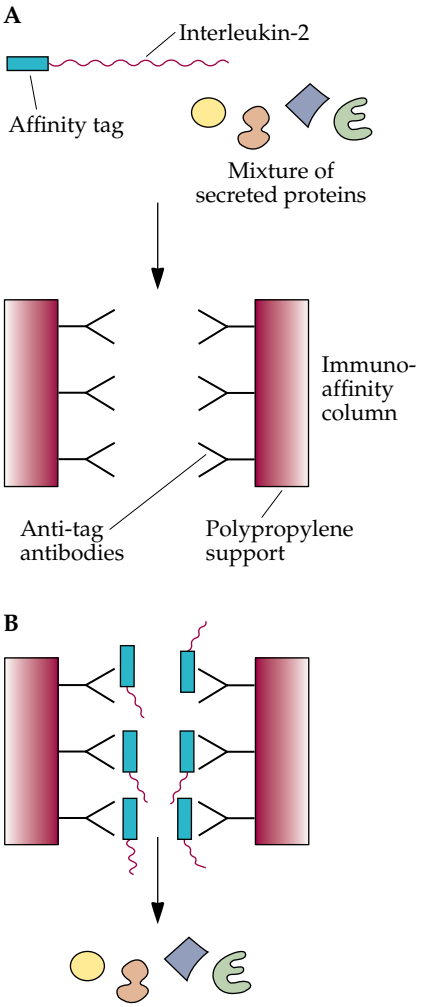


Figure 3.17 Purification of a fusion protein by immunoaffinity chromatography. **(A)** An antibody that binds to a short peptide sequence (affinity tag) of the fusion protein (anti-tag antibody) is attached to a solid polypropylene support. The mixture of secreted proteins is passed through the column containing the bound antibody. **(B)** The affinity tag of the fusion protein binds to the antibody, and the other proteins pass through. The immunopurified fusion protein can then be selectively eluted from the column by the addition of pure affinity tag peptide.

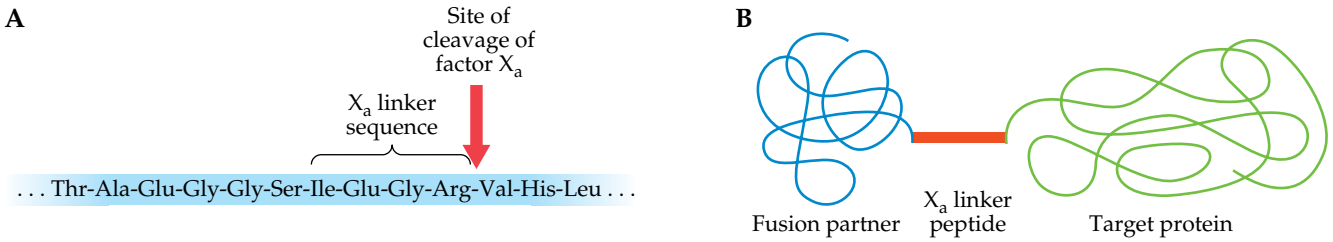


Figure 3.18 Proteolytic cleavage of a fusion protein by blood coagulation factor X_a. **(A)** The factor X_a recognition sequence (X_a linker sequence) lies between the amino acid sequences of two different proteins. A functional target protein (with valine at the N terminus) is released after cleavage. **(B)** A tripartite fusion protein including a stable fusion partner, a linker peptide, and the target protein.

Self-Cleaving Purification Tags

To avoid additional purification steps that are costly and may reduce yields of target protein, self-cleaving purification tags can be used. An intein (intervening protein) is an amino acid segment found in some natural proteins that can excise itself and join the flanking segments (exteins) with a peptide bond. Excision of the intein is mediated by a cysteine or serine amino acid at the N-terminal end of the intein sequence, an asparagine at the C-terminal end, and a cysteine, serine, or threonine immediately after the asparagine (Fig. 3.19). For affinity tag removal, the target protein can be expressed with an intein-affinity tag sequence at either the N or C terminus. For removal of an N-terminal tag, the codon for the intein N-terminal cysteine/serine is replaced with the codon for alanine to prevent rejoining of the affinity tag and target protein sequences that flank the intein (Fig. 3.20A). Similarly, for cleavage of a C-terminal tag, the codon for the intein C-terminal asparagine is replaced (Fig. 3.20B). Following purification of a target protein, cleavage of the intein-affinity tag is induced by increasing the pH of the solution or by treatment with a thiol reagent such as dithiothreitol or β-mercaptoethanol (Fig. 3.20).

Integrating DNA into the Host Chromosome

A plasmid vector imposes a metabolic load on the cell because of the energy that is used for its replication and for the transcription of RNA and translation of the proteins that it encodes. As a consequence, a fraction of the cell population often loses its plasmids during cell growth. Cells that lack plasmids generally grow faster than those that retain them, so plasmidless cells eventually dominate the culture. After a number of generations of cell growth, the loss of plasmid-containing cells diminishes the yield of the cloned gene product. On a laboratory scale, plasmid-containing cells are typically maintained by growing the cells in the presence of an antibiotic that enables only cells bearing a plasmid carrying antibiotic-resistant genes to thrive. But the addition of antibiotics to pilot plant- or industrial-scale fermentations can be extremely costly, and it is imperative that anything that is added to the fermentation be completely removed before the product is approved for human

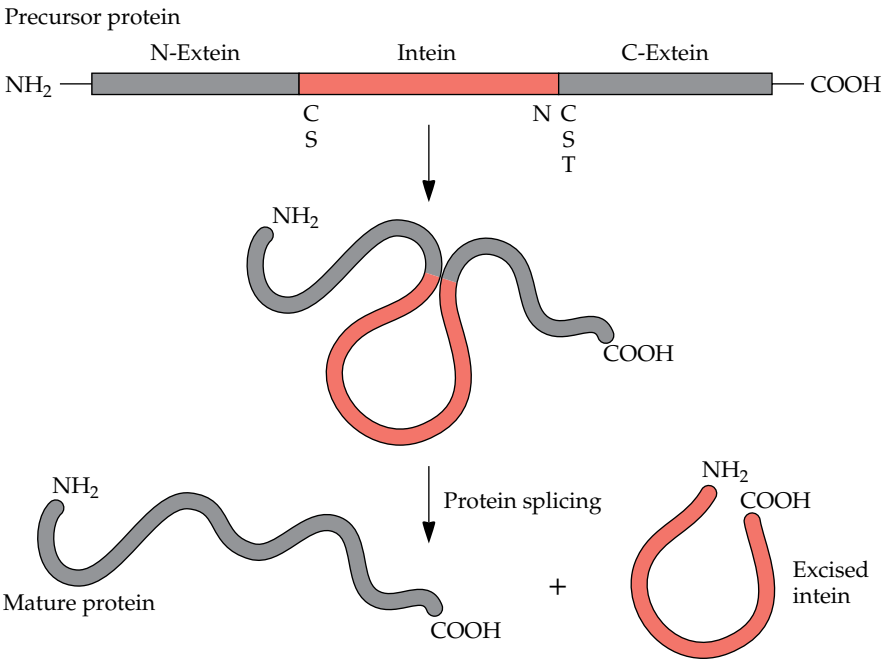
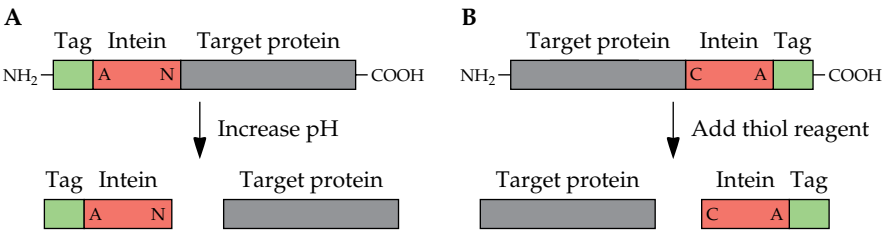


Figure 3.19 Self-splicing proteins. Excision of an intein sequence from a precursor protein is facilitated by specific amino acids at the N- and C-ends of the intein. Following excision, the N- and C-exteins are joined to produce the mature protein. C, cysteine; S, serine; N, asparagine; T, threonine. Data from Elleuche and Pöggler, *Appl. Microbiol. Biotechnol*. **87:**479–489, 2010.

use. For genetically engineered microorganisms that are designed to be released into the environment to remain both effective and environmentally safe, it is essential that the cloned DNA be retained and be neither easily lost nor transferred to other microorganisms. For these reasons, the introduction of cloned DNA directly into the chromosomal DNA of the host organism can overcome the problem of plasmid loss. When DNA is part of the host chromosomal DNA, it is relatively stable and consequently can be maintained for many generations in the absence of selective agents.

Figure 3.20 Removal of protein purification tags using inteins. Affinity tags can be removed by inserting the coding sequence for an intein between the coding sequences for the affinity tag and the target protein. Following translation and purification, self-excision of the intein is induced by increasing pH or addition of a thiol reagent such as dithiothreitol. **(A)** Cleavage of an N-terminal tag. **(B)** Cleavage of a C-terminal tag. A, alanine; N, asparagine; C, cysteine. Data from Fong et al., *Trends Biotechnol*. **28:**272–279, 2010.



The chromosomal integration site of a target gene must not be within an essential coding gene. Consequently, the gene must be targeted to a specific nonessential site within the chromosome. In addition, to ensure efficient production of the recombinant protein, the target gene should be under the control of a regulatable promoter.

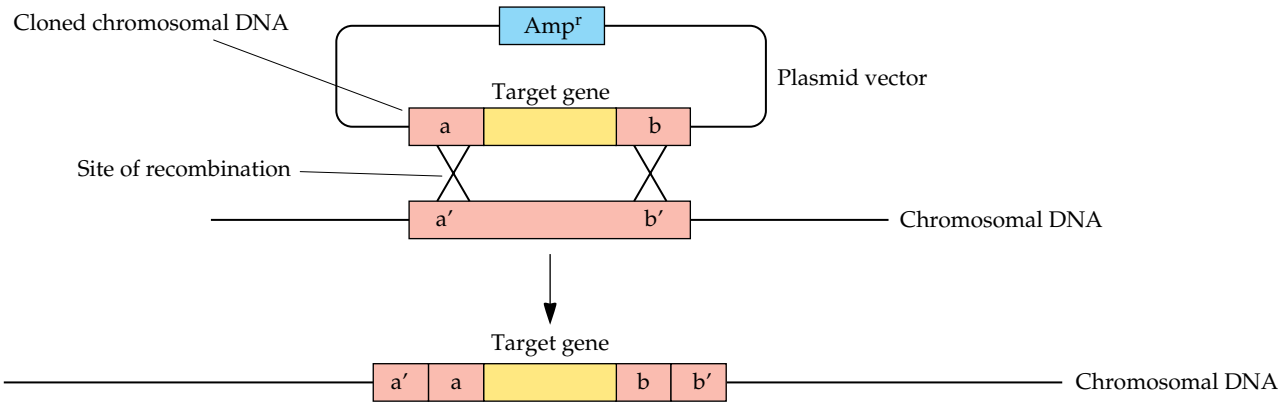
Chromosomal Integration by One-Step Homologous Recombination

Chromosomal integration of a target gene occurs by homologous recombination. For this, the target gene is flanked by sequences that share similarity, at least 50 nucleotides, with the sequence at the integration site. Recombination, mediated by specific host enzymes, results in physical exchange between the two DNA molecules. A generalized protocol for DNA integration includes the following steps.

1. Identify the desired chromosomal integration site, i.e., a segment of DNA on the host chromosome that can be disrupted without affecting the normal functions of the cell.
2. Clone the sequence from the chromosomal integration site into a plasmid vector. The chromosomal sequence may be obtained by PCR (described in chapter 2). The plasmid vector must not have an origin of replication that enables it to be replicated in the host bacterium.
3. Insert a target gene with a regulatable promoter into the cloned chromosomal integration site on the plasmid (Fig. 3.21).
4. Transfer the plasmid containing the target gene into the host cell, usually by conjugation (described in chapter 2).
5. Select and perpetuate host cells that express the target gene. Propagation of the target gene can occur only if it has been integrated into the chromosomal DNA of the host cell.

When a host cell is transformed with a nonreplicating plasmid that carries the target gene in the middle of the cloned chromosomal DNA fragment,

Figure 3.21 Integration of a target gene into a chromosomal site by one-step homologous recombination. The target gene is cloned into a plasmid vector between sequences (a and b) that are homologous to chromosomal DNA sequences (a′ and b′) at the site targeted for integration. Recombination occurs between homologous plasmid-borne DNA regions a and b and host chromosome DNA regions a′ and b′, respectively. A double-crossover event (×) results in the stable integration of the target gene into the chromosome.



homologous DNA on the plasmid and in the host chromosome can recombine (Fig. 3.21). Target gene integration into the chromosome occurs as a result of a host enzyme-catalyzed double crossover. Alternatively, a single crossover that incorporates the entire plasmid into the host chromosome may occur (not shown).

This protocol is exemplified by the integration of the α-amylase gene from *Bacillus amyloliquefaciens* into the chromosome of *B. subtilis*. The α-amylase gene, under the control of a *B. subtilis* promoter, was inserted into the middle of a chromosomal DNA fragment from *B. subtilis* that had been cloned in an *E. coli* plasmid. The plasmid carries the *bla* gene that confers resistance to ampicillin and cannot replicate in *B. subtilis*. The plasmid can, however, transform *B. subtilis*. Transformants expressing α-amylase, an enzyme involved in the hydrolysis of starch, were identified by a zone of clearance around colonies that grew on solid medium containing starch. This indicated that the α-amylase gene had been integrated into the *B. subtilis* chromosome and was expressed and secreted. The selected transformants were sensitive to ampicillin, indicating that a double crossover had occurred and that following the exchange of DNA, the plasmid was lost from the population. Resistance to ampicillin would indicate that a single recombination event had occurred and caused the integration of the entire plasmid into the *B. subtilis* chromosomal DNA. Integration at the correct chromosomal site may be confirmed by PCR using primers that anneal to sequences flanking the integration site.

Chromosomal Integration by Two-Step Homologous Recombination

For target genes whose protein products do not have an activity that can be easily detected in a screen for transformants, selection for chromosomal integration can be performed in two steps (Fig. 3.22). First, following introduction of the plasmid vector carrying the target gene into the host bacterium, transformants are selected by plating on solid medium containing an antibiotic. Because the plasmid cannot replicate in the host cell, antibiotic-resistant transformants have integrated the entire plasmid, including the gene that confers resistance to the antibiotic into the chromosome by recombination at only one region of sequence homology. The single crossover is confirmed by PCR using one primer that anneals to the vector sequence and another that anneals to the flanking chromosomal DNA. Next, cells from a single recombinant colony are grown in liquid medium lacking the antibiotic and are subsequently screened for loss of the antibiotic resistance marker gene. Cells are first plated on medium that lacks the antibiotic, and then colonies are replica-plated onto medium containing the antibiotic to identify those that have reverted to antibiotic sensitivity. These transformants have undergone recombination at the second region of sequence homology that resulted in excision of the plasmid sequence, but not the target gene, from the chromosome. Selection for loss of the plasmid sequence is facilitated if it contains a gene encoding a toxin (chapter 2, Table 2.2); for example, cells that retain an integrated plasmid carrying the *sacB* gene are killed in the presence of sucrose. The second recombination event is required when the presence of plasmid sequences, especially antibiotic resistance genes, is undesirable, for example, for organisms that are intended for deliberate release into the environment (e.g., to degrade pollutants).

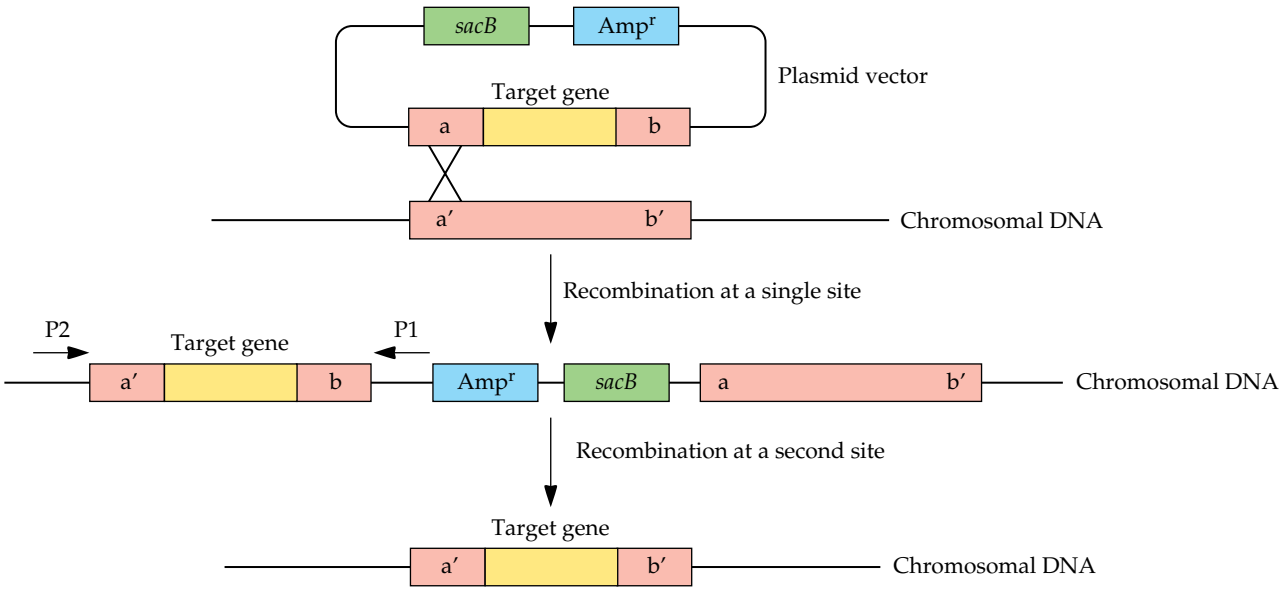


Figure 3.22 Integration of a target gene into a chromosomal site by two-step homologous recombination. The target gene is cloned into a plasmid vector between sequences (a and b) that are homologous to chromosomal DNA sequences (a′ and b′) at the site targeted for integration. In the first step, a single crossover (×) between homologous sequences a and a′ results in integration of the entire plasmid, including the ampicillin resistance gene (Amp^r) and the *sacB* gene, into the host chromosome (note that it is equally likely that a single crossover could occur between homologous sequences b and b′). Clones with integrated plasmids are selected on medium containing ampicillin, and plasmid integration may be confirmed by PCR with one primer that anneals to a plasmid sequence and the other primer that anneals to a flanking chromosomal DNA sequence (denoted by arrows labeled P1 and P2, respectively). In the second step, recombination at a second site of sequence homology (b and b′) results in excision of the plasmid sequence, and clones are screened for loss of the *sacB* (growth on medium containing sucrose) and Amp^r genes (lack of growth of medium containing ampicillin). The target gene is stably integrated into the chromosome.

Removal of Selectable Markers by Site-Specific Recombination

An antibiotic resistance gene or other undesirable vector sequence may also be removed following integration into the chromosome by a site-specific recombinase. In this approach, the sequences to be excised are flanked by short specific DNA sequences that are recognized and recombined by a site-specific recombinase, resulting in excision of the intervening sequence (Fig. 3.23). One combination of an enzyme and DNA sequence that is useful for this sort of manipulation is the Cre–*loxP* recombination system, which consists of the Cre recombinase enzyme and two 34-bp *loxP* recombination sites. The antibiotic resistance gene to be removed is flanked by *loxP* sites, and after integration of the plasmid into the chromosome, the gene is removed by the Cre enzyme. A gene encoding the Cre enzyme is located on its own plasmid, which can be introduced into the host cells carrying the chromosomally-integrated plasmid. Gene excision is triggered by the addition of IPTG to the growth medium, which turns on the *lac* promoter driving expression of the *cre* gene and causes the Cre enzyme to be

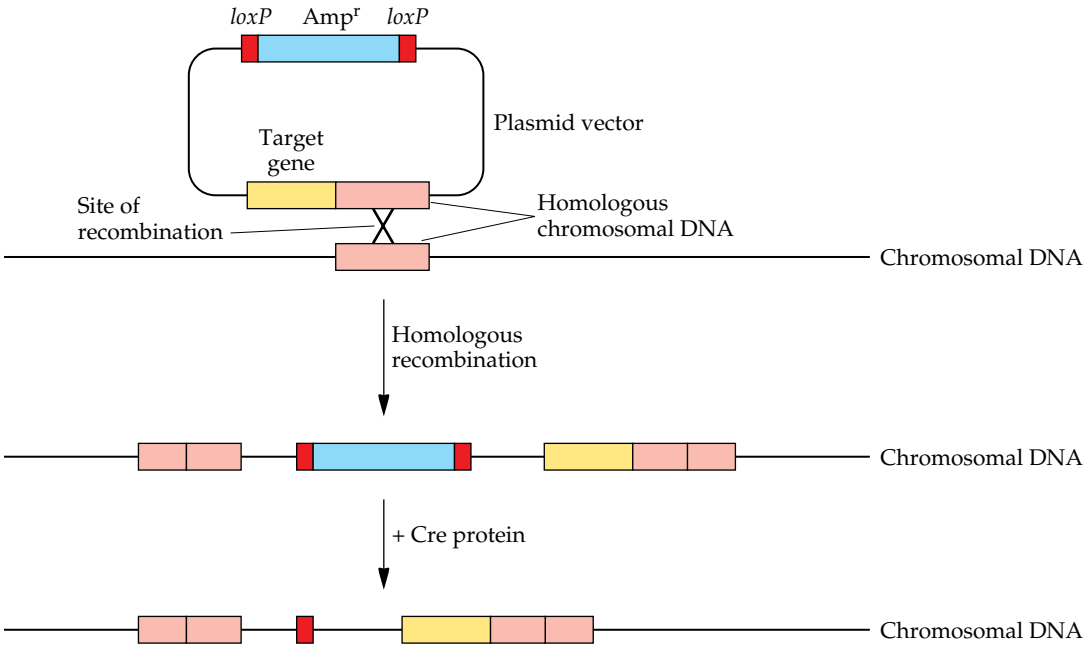


Figure 3.23 Removal of an antibiotic resistance gene following integration of plasmid DNA into a bacterial chromosome. The target gene is cloned into a plasmid vector adjacent to a sequence that is homologous to a chromosomal DNA sequence at the site targeted for integration. Recombination (×) occurs between homologous sequences in the chromosome and on the plasmid, resulting in the integration of the entire plasmid into the chromosome. The antibiotic resistance gene (Amp^r), which is flanked by *loxP* sites, is excised by the action of the Cre recombinase enzyme, leaving one *loxP* site on the integrated plasmid. The gene encoding the Cre enzyme is on a separate plasmid within the same cell under the transcriptional control of the *E. coli lac* promoter so that excision is induced when IPTG is added to the growth medium.

synthesized. Following excision, the plasmid containing the *cre* gene may be removed from the host cells by raising the temperature. This plasmid has a temperature-sensitive replicon that allows it to be maintained in the cell at 30°C but not above 37°C. Given the alarming increase in antibiotic-resistant strains of many bacteria, it is essential to avoid the introduction and spread of antibiotic-resistant genes in the environment. Removing antibiotic-resistant genes from genetically engineered bacteria is an important step in that direction.

Heterologous Protein Production in Eukaryotic Cells

Proteins from a wide variety of organisms have been successfully produced using prokaryotic expression systems. Although expression of any gene from any source organism in a prokaryotic host is theoretically possible, in practice, the eukaryotic proteins produced in bacteria do not always have the desired biological activity or stability. In addition, despite careful purification procedures, bacterial compounds that are toxic or that cause a rise in body temperature in humans and animals (pyrogens) may contaminate the final product. To avoid these problems, investigators

have developed eukaryotic expression systems in fungal, insect, and mammalian cells for the production of therapeutic agents for either humans or animals; large quantities of stable, biologically active proteins for biochemical, biophysical, and structural studies; and proteins for industrial processes. Moreover, any human protein intended for medical use must be identical to the natural protein in all its properties. The inability of prokaryotic organisms to produce authentic versions of eukaryotic proteins is, for the most part, due to improper posttranslational protein processing, including improper protein cleavage and folding, and to the absence of appropriate mechanisms that add chemical groups to specific amino acid acceptor sites.

Posttranslational Modification of Eukaryotic Proteins

In prokaryotes, all of the steps in protein synthesis take place in the cytoplasm, and therefore, translation of messenger RNA (mRNA) occurs concurrently with transcription; as soon as the nascent transcript emerges from RNA polymerase, it is accessible to the ribosome to begin translation. With the aid of folding proteins, known as chaperones, that bind to polypeptides as they are being synthesized, proteins are folded into their proper three-dimensional configuration during synthesis. In contrast, eukaryotes synthesize mRNA in the nucleus, from which it is exported to ribosomes in the cytoplasm or on the endoplasmic reticulum, where translation occurs. Proteins produced on ribosomes associated with the endoplasmic reticulum either are inserted in the membrane of the endoplasmic reticulum or are secreted into the lumen of the endoplasmic reticulum during synthesis, where they are processed further.

Many eukaryotic proteins, including most of those that are of interest as therapeutic agents for the treatment of human or animal diseases, undergo some type of posttranslational processing that is often required for protein activity and stability. Some proteins are produced as inactive precursor polypeptides that must be cleaved by proteases at specific sites to produce the active form of the protein. For example, the small peptide hormone insulin is produced in animal pancreatic cells as a single polypeptide, preproinsulin, that is cleaved to produce two shorter peptides that are joined by disulfide bonds (Fig. 3.24). Production of inactive preproinsulin ensures that the peptide is not active in the pancreatic cells that produce it, but upon secretion, cleaved mature insulin can act on other cells. Similarly, the digestive enzyme trypsinogen, which degrades proteins, is produced as an inactive polypeptide to avoid digestion of components of the producing cell. Upon secretion into the small intestines, trypsinogen is cleaved by an enteropeptidase to yield the active enzyme trypsin.

Similar to prokaryotes, proper folding of proteins in eukaryotic cells requires the assistance of chaperones. In the endoplasmic reticulum, the chaperones BiP and calnexin bind nascent polypeptides, and protein disulfide isomerases catalyze the formation of disulfide bonds between cysteines. Proper folding is important not only for the protein to attain a configuration for optimal activity but also to protect regions of the protein that would otherwise be recognized by proteases that can destroy the protein. Quality control systems ensure that only correctly folded proteins are released from the endoplasmic reticulum and transported within vesicles

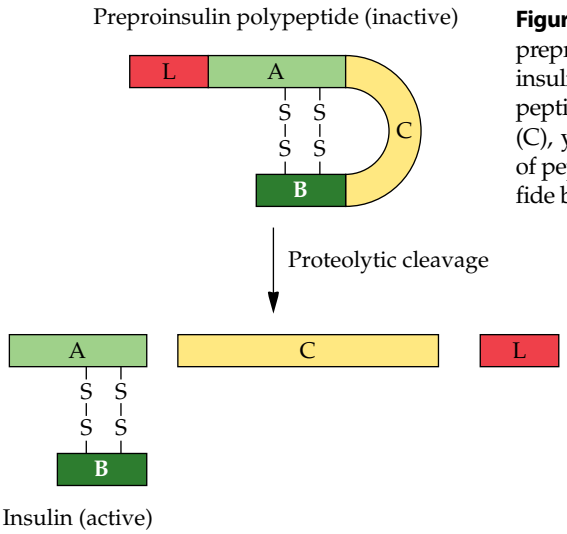


Figure 3.24 Cleavage of inactive preproinsulin to yield active mature insulin. Proteases remove the leader peptide (L) and an internal peptide (C), yielding a protein that consists of peptides A and B linked by disulfide bonds.

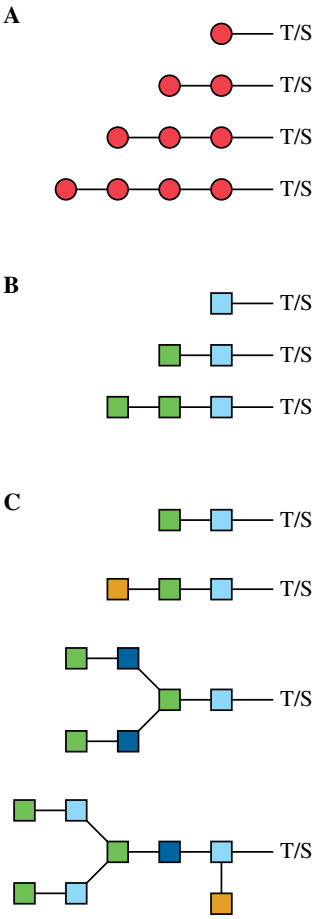
to the Golgi apparatus for further processing. Proteins intended for secretion from the cell are subsequently transported to the cell membrane within specific transport vesicles and released by exocytosis.

The addition of specific sugars (glycosylation) to certain amino acids is a major posttranslational modification that provides stability and distinctive binding properties to a protein. Proper protein glycosylation is important because it contributes to protein conformation by influencing protein folding, can target a protein to a particular location, for example, through interaction with a specific receptor molecule, or can increase protein stability by protecting it from proteases. In the cell, oligosaccharides are attached to newly synthesized proteins in the endoplasmic reticulum and in the Golgi apparatus by specific enzymes known as glycosylases and glycosyltransferases. Different tissues may differentially glycosylate the same protein, thereby increasing protein heterogeneity. Because different sugar modifications can alter the properties of a protein, this presents opportunities for protein engineering to improve the efficacy or to alter the activity of a protein. About 50% of all human proteins are glycosylated. Human therapeutic proteins that require glycosylation for optimal activity include antibodies, blood factors, some interferons, and some hormones.

The most common glycosylations entail the attachment of specific sugars to the hydroxyl group of either serine or threonine (O-linked glycosylation) (Fig. 3.25) and to the amide group of asparagine (N-linked glycosylation) (Fig. 3.26). The initial core sugar groups that are added to these amino acid acceptor sites tend to be similar among eukaryotes, although the subsequent elaborations among yeasts, insects, and mammals are quite diverse, especially for N-linked glycosylation. Other amino acid modifications include phosphorylation, acetylation, sulfation, acylation, γ-carboxylation, and the addition of C_{14} and C_{16} fatty acids, i.e., myristoylation (or myristylation) and palmitoylation (or palmitylation), respectively.

Unfortunately, there is no universally effective eukaryotic host cell that performs the correct modifications on every protein. In some cases, a host cell may add unusual sugars to either authentic or spurious amino acid

Figure 3.25 Examples of some O-linked oligosaccharides in yeasts **(A)**, insects **(B)**, and mammals **(C)**. O-linked oligosaccharides have a number of arrangements with different combinations of sugars. Some of the more prevalent forms are shown here. S, serine; T, threonine; red circles, mannose; light blue squares, *N*-acetylgalactosamine; green squares, galactose; orange squares, sialic acid; dark blue squares, *N*-acetylglucosamine.



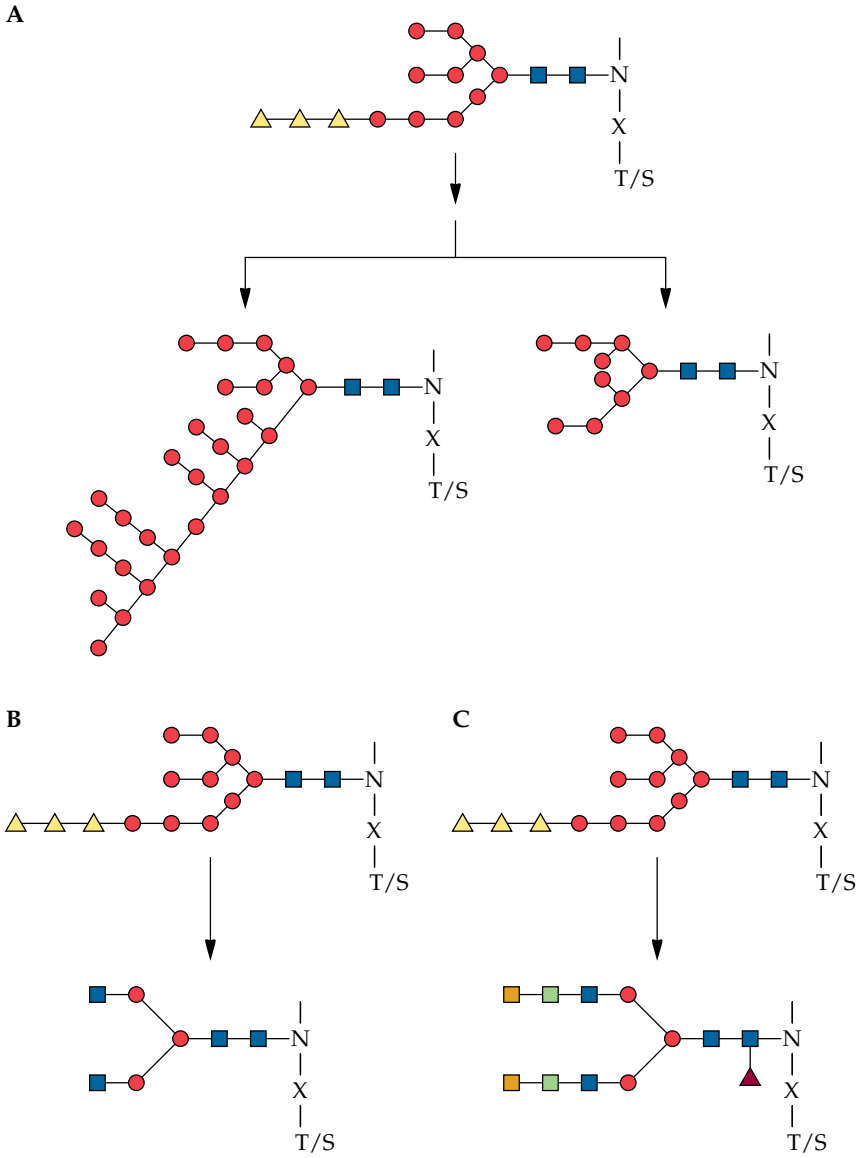


Figure 3.26 Examples of some N-linked oligosaccharides in yeasts **(A)**, insects **(B)**, and mammals **(C)**. All N-linked glycosylations in eukaryotes start with the same initial group, which is subsequently trimmed and then elaborated in diverse ways within and among species. Some yeast sites have 15 or fewer mannose units (core series), and others have more (outer-chain family). In *S. cerevisiae*, the oligosaccharides frequently have 50 or more mannose units. An asparagine (N) next to any amino acid (X) followed by either threonine (T) or serine (S) can be targeted for glycosylation. Red circles, mannose; dark blue squares, *N*-acetylglucosamine; yellow triangles, glucose; green squares, galactose; orange squares, sialic acid; maroon triangle, fucose.

sites and, consequently, create an extremely antigenic protein or possibly one that lacks its proper function. Even though a recombinant protein may fall short of the stringent properties that are required for a therapeutic agent, it may still be useful for either research or industrial purposes. Different eukaryotic expression systems must be tested to determine

which one synthesizes the largest amount of a functional recombinant protein. The choice of an expression system depends primarily on the quality of the recombinant protein that is produced, but the yield of product, ease of use, and cost of production and purification are also important considerations.

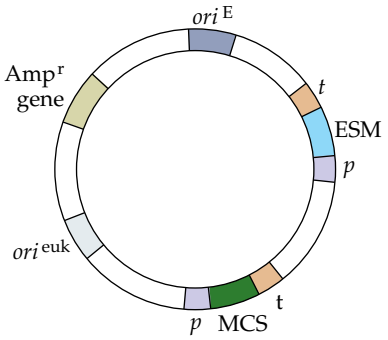


Figure 3.27 Generalized eukaryotic expression vector. The major features of a eukaryotic expression vector are a eukaryotic transcription unit with a promoter (*p*), a multiple cloning site (MCS) in which to insert a target gene, and a DNA segment with termination and polyadenylation signals (*t*). A eukaryotic origin of replication (ori^{euk}) enables replication of the vector in a eukaryotic host, and a eukaryotic selectable marker (ESM) gene enables selection of eukaryotic cells carrying the vector. An origin of replication that functions in *E. coli* (ori^{E}) and an *E. coli* selectable marker (Amp^{r}) gene are useful during the cloning of the heterologous gene.

General Features of Eukaryotic Expression Systems

The basic requirements for expression of a target protein in a eukaryotic host are similar to those required in prokaryotes. Vectors into which the target gene is cloned for delivery into the host cell can be specialized plasmids designed to be maintained in the eukaryotic host, such as the yeast 2-μm plasmid, host-specific viruses, such as the insect baculovirus, or artificial chromosomes, such as the yeast artificial chromosome (YAC). The vector must have a eukaryotic promoter that drives the transcription of the cloned target gene, eukaryotic transcriptional and translational stop signals, a sequence that enables polyadenylation of the mRNA, and a selectable eukaryotic marker gene (Fig. 3.27). Because recombinant DNA procedures are technically more difficult to carry out with eukaryotic cells, they are typically carried out in prokaryotic cells before transferring the vectors to the eukaryotic host for protein expression. Therefore, most eukaryotic vectors are shuttle vectors with two origins of replication and two separate selectable marker genes; one set of these genes functions in the bacterium, most often *E. coli*, and the other set functions in the eukaryotic host cell. If a eukaryotic expression vector is to be used as a plasmid, i.e., as extrachromosomal replicating DNA, then it must also have a eukaryotic origin of replication. Alternatively, if the vector is designed for stable integration into the host chromosomal DNA, then it must have a sequence that is homologous to a segment of the host chromosome for insertion by homologous recombination.

The introduction of DNA into bacterial or fungal cells is called transformation. In these systems, the term describes an inherited change due to the acquisition of exogenous (foreign) DNA. However, in animal cells, transformation refers to changes in the growth properties of cells in culture after they become cancerous. To avoid confusion, the term transfection is used to denote inherited changes in animal cells that are due to the addition of exogenous DNA.

Three techniques are commonly used to transform yeasts: electroporation, lithium acetate treatment, and cell wall removal (protoplast formation). Transfection of cultured animal cells is achieved by incubating cells with DNA that has been coprecipitated with either calcium phosphate or diethylaminoethyl (DEAE)–dextran or by electroporation. Electroporation entails subjecting cells to short pulses of electric current, thus creating transient pores through which DNA can enter the cell (chapter 2, Fig. 2.7). Viruses, lipid–DNA complexes, and protein–DNA aggregates are also used to transfer exogenous DNA into a recipient animal cell.

Yeast Expression Systems

Yeasts share many of the molecular, genetic, and biochemical features of other, "higher" eukaryotes and are therefore a good choice for heterologous protein production. They have growth advantages similar to those of prokaryotes, such as rapid growth in low-cost medium, generally do

not require growth factors to be added to the growth medium, can correctly process eukaryotic proteins, and can secrete large amounts of heterologous proteins. Initially, the yeast *Saccharomyces cerevisiae* was used extensively as a host cell for the expression of cloned eukaryotic genes. It has a long history of use in traditional biotechnologies in the brewing and baking industries. Today, a variety of yeast and other fungal expression systems are available, and they have been optimized for recombinant protein expression. Versatile expression vectors with broad host ranges have been constructed because the optimal host for production of a particular target protein must often be determined experimentally in a number of different systems.

High levels of recombinant protein production have been achieved using *S. cerevisiae*. There are advantages of using this single-celled yeast. First, a great deal is known about its biochemistry, genetics, and cell biology. The genome sequence of *S. cerevisiae* was completed in 1996, and it is used extensively in studies as a model organism for cell function. Second, it can be grown rapidly to high cell densities on relatively simple media in both small culture vessels and large-scale bioreactors. Third, several strong promoters have been isolated from the yeast and characterized, and a naturally occurring plasmid, called the 2-μm plasmid, can be used as part of an endogenous yeast expression vector system. Fourth, *S. cerevisiae* is capable of carrying out many posttranslational modifications. Fifth, *S. cerevisiae* normally secretes so few proteins that, when it is engineered for extracellular release of a recombinant protein, the product can be easily purified. Sixth, because of its many years of use in the baking and brewing industries, *S. cerevisiae* has been listed by the U.S. Food and Drug Administration as a "generally recognized as safe" (GRAS) organism. It does not harbor human pathogens or produce fever-stimulating pyrogens. Therefore, the use of this organism for the production of human therapeutic agents (drugs or pharmaceuticals) does not require the same extensive experimentation demanded for unapproved host cells. A number of proteins that have been produced in *S. cerevisiae* are currently being used commercially as vaccines, pharmaceuticals, and diagnostic agents (Table 3.11). For example, at present, more than 50% of the world supply of insulin is produced by *S. cerevisiae*. Engineered *S. cerevisiae* strains are also major producers of a hepatitis B vaccine.

Table 3.11 Recombinant proteins produced by *S. cerevisiae* expression systems[a]

Vaccines
Hepatitis B virus surface antigen
Malaria circumsporozoite protein
HIV-1 envelope protein
Diagnostics
Hepatitis C virus protein
HIV-1 antigens
Human therapeutic agents
Epidermal growth factor
Insulin
Insulin-like growth factor
Platelet-derived growth factor
Proinsulin
Fibroblast growth factor
Granulocyte-macrophage colony-stimulating factor
α1-Antitrypsin
Blood coagulation factor XIIIa
Hirudin
Human growth factor
Human serum albumin

[a]HIV-1, human immunodeficiency virus type 1.

Saccharomyces cerevisiae Expression Vectors

There are three main classes of *S. cerevisiae* expression vectors: episomal, or plasmid, vectors, integrating vectors, and YACs. Of these, episomal plasmid vectors have been used extensively for the production of either intra- or extracellular heterologous proteins. Typically, the vectors contain features that allow them to function in both bacteria and *S. cerevisiae*. An *E. coli* origin of replication and bacterial antibiotic resistance genes are usually included on the vector, enabling all manipulations to first be performed in *E. coli* before the vector is transferred to *S. cerevisiae* for expression.

Episomal vectors. The yeast episomal plasmid vectors are based on the high-copy-number 2-μm plasmid, a small, circular plasmid found in most natural strains of *S. cerevisiae*. The vector replicates independently of the host chromosome via a single origin of replication (autonomous replicating

sequence [ARS]) and is maintained in more than 30 copies per cell. Many *S. cerevisiae* selection schemes rely on mutant host strains that require a particular amino acid (histidine, tryptophan, or leucine) or nucleotide (uracil) for growth. Such strains are said to be auxotrophic because minimal growth medium must be supplemented with a specific nutrient. The vector is equipped with a functional (wild-type) version of a gene that complements the mutated gene in the host strain. For example, when a plasmid vector with a functional copy of the *LEU2* gene is transformed into a mutant *leu2* host cell and plated onto medium that lacks leucine, only cells that carry the plasmid will grow.

A number of promoters derived from *S. cerevisiae* genes are utilized for engineering efficient transcription of heterologous genes in yeast vectors (Table 3.12). Generally, tightly regulatable, inducible promoters are preferred for producing large amounts of recombinant protein at a specific time during large-scale growth. In this context, the galactose-regulated promoters respond rapidly to the addition of galactose with a 1,000-fold increase in transcription. Repressible, constitutive, and hybrid promoters that combine the features of different promoters are also available. Maximal expression also depends on efficient termination of transcription. Often, for plasmid vectors, the transcription terminator sequence is from the same gene as the promoter.

Many heterologous proteins are engineered to contain a secretory signal peptide that directs them to the endoplasmic reticulum for subsequent export to the cell membrane and release to the extracellular medium. The main reason for this modification is that it is much easier to purify a secreted protein than one from a cell lysate. The most commonly used signal sequence for *S. cerevisiae* is derived from the yeast mating factor α gene. The secretion efficiency of recombinant proteins fused to a signal peptide can vary widely, and therefore libraries of synthetic signal sequences are created and screened to identify those that increase the secretion of a target protein. Other sequences that stabilize the

Table 3.12 Promoters for *S. cerevisiae* expression vectors

Source gene	Gene product	Expression conditions	Status
ADHI	Alcohol dehydrogenase I	2–5% glucose	Constitutive
ADHII	Alcohol dehydrogenase II	0.1–0.2% glucose	Inducible
CUP1	Metallothionein	0.03–0.1 mM copper	Inducible
CYC1	Cytochrome c_1	Glucose	Repressible
ENO	Enolase		Constitutive
GAL1	Galactokinase	Galactose	Inducible
GAL10	UDP galactose epimerase	Galactose	Inducible
GAPD, GAPDH	Glyceraldehyde-3-phosphate dehydrogenase	2–5% glucose	Constitutive
MET17	Homocysteine synthase	Absence of methionine	Repressible
PDC1	Pyruvate decarboxylase	Glucose	Inducible
PGK	Phosphoglycerate kinase	2–5% glucose	Constitutive
PH05	Acid phosphatase	Absence of phosphate	Inducible
SNR52	RNA polymerase III		Constitutive
TEF1	Translation elongation factor 1		Constitutive
TPI1	Triose phosphate isomerase	2–5% glucose	Constitutive

recombinant protein, protect it from proteolytic degradation, and provide a purification tag can be fused to the coding sequence of the heterologous gene. These extra amino acid sequences are equipped with a protease cleavage site so that they can be removed from the recombinant protein after it is purified.

Integrating vectors. Plasmid-based yeast expression systems are often unstable under large-scale (≥10 liters) growth conditions even in the presence of selective pressure. To remedy this problem, a target gene is integrated into the host genome to provide a more reliable production system. To avoid disruption of important functions, the target gene is inserted into the *S. cerevisiae* chromosome within a nonessential gene such as *HO*, which encodes an endonuclease, or in a noncoding sequence such as the nontranscribed spacer of the ribosomal RNA locus; however, the site of integration can affect levels of heterologous gene expression. The target gene with yeast-specific promoter and transcriptional termination sequences and a yeast selectable marker are inserted between two DNA segments derived from the ends of the nonessential yeast sequence. These may be initially constructed in a plasmid and then excised by restriction enzyme digestion or assembled from PCR fragments (see chapter 2). The linear fragment is transformed into *S. cerevisiae*. A double recombination event between homologous sequences on the introduced fragment and the yeast chromosome results in integration of the target gene (Fig. 3.28). Linear fragments are more likely than circular plasmids to recombine with chromosomal DNA. Homologous recombination is enhanced during DNA repair, and therefore the CRISPR-Cas system may be employed to generate double-stranded breaks in the host chromosome at the site selected for integration (chapter 2, Fig. 2.19). After selecting for transformants, it may be desirable to excise the selectable marker gene. This can be accomplished using a site-specific recombinase, such as Cre recombinase, described above, if the marker gene is flanked by its recognition sites. A major drawback of chromosomal integration in yeast strains is the low yield of recombinant protein from a single gene copy.

Yeast artificial chromosomes. A YAC is designed to clone a large segment of DNA (~100 kilobase pairs [kb]), which is then maintained as a separate chromosome in the host yeast cell. The YAC system is highly stable and has been used for the physical mapping of human genomic DNA, the analysis of large transcription units, and the formation of genomic libraries containing DNA from individual human chromosomes. A YAC vector mimics a chromosome because it has a sequence that acts as an origin of DNA replication (ARS), a yeast centromere sequence to ensure that after cell division each daughter cell receives a copy of the YAC, and telomere sequences that are present at both ends after linearization of the YAC DNA for stability (Fig. 3.29). In some cases, the input DNA is cloned into a site that disrupts a yeast marker gene. In the absence of the product of the marker gene, a colorimetric response is observed when recipient cells are grown on a specialized medium. Alternatively, some YAC vectors contain a selectable marker gene that is independent of the cloning site.

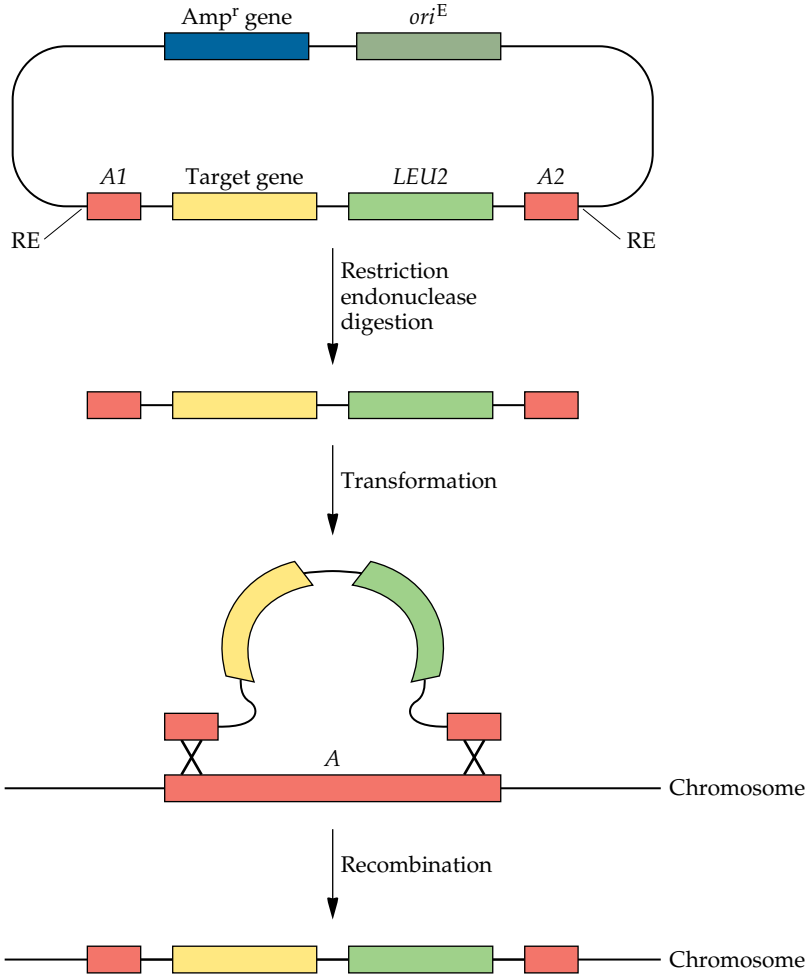


Figure 3.28 Integration of DNA into an *S. cerevisiae* chromosome. A target gene and selectable marker gene (*LEU2*), both with yeast transcription and translation control elements (not shown), are inserted into a YIp vector between two segments (*A1* and *A2*) from the ends of a nonessential yeast gene A. The ampicillin resistance (Amp^r) gene and the origin of replication (*ori*^{E}) function in *E. coli*. A leucine-requiring (*leu2*) yeast strain is transformed with restriction endonuclease-digested vector DNA because chromosomal DNA is more likely to recombine with linearized DNA than with circular DNA. The restriction endonuclease sites (RE) flank the segments from the nonessential yeast gene. Homologous DNA sequences of the nonessential yeast gene *A* undergo recombination (×) that leads to the incorporation of both the target gene and the *LEU2* gene into the corresponding chromosome site. Transformants grow on medium that is not supplemented with leucine. Nonrecombined DNA is degraded.

Secretion of Heterologous Proteins by *S. cerevisiae*

Signal peptides. All glycosylated proteins of *S. cerevisiae* are secreted. Consequently, the coding sequences of recombinant proteins that require either O-linked or N-linked sugars for biological activity must be equipped with a signal peptide to pass through the secretory system (Fig. 3.30). Usually, the signal sequence from the yeast mating type α-factor gene

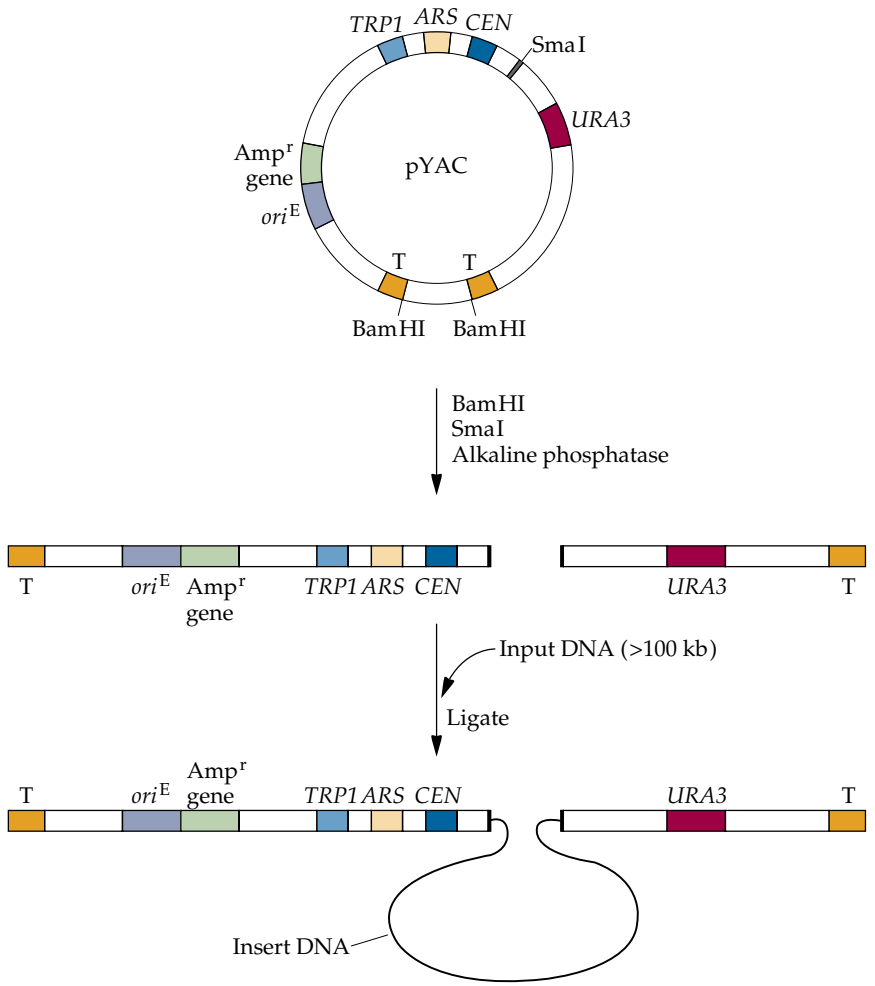


Figure 3.29 YAC cloning system. The YAC plasmid (pYAC) has an *E. coli* selectable marker (Amp^r) gene, an origin of replication that functions in *E. coli* (ori^E), and yeast DNA sequences, including *URA3*, *CEN*, *TRP1*, and *ARS*. *CEN* provides centromere function, *ARS* is a yeast autonomous replicating sequence that is equivalent to a yeast origin of replication, *URA3* is a functional gene of the uracil biosynthesis pathway, and *TRP1* is a functional gene of the tryptophan biosynthesis pathway. The T regions are yeast chromosome telomeric sequences. The SmaI site is the cloning insertion site. pYAC is first treated with SmaI, BamHI, and alkaline phosphatase and then ligated with size-fractionated (100-kb) input DNA. The final construct carries cloned DNA and can be stably maintained in double-mutant *ura3* and *trp1* cells.

(prepro-α-factor) is inserted immediately in front (upstream) of the cDNA of the target gene. Under these conditions, correct disulfide bond formation, proteolytic removal of the signal sequence, and appropriate posttranslational modifications often occur, and an active recombinant protein is secreted. During this process, the signal peptide is removed by an endoprotease that recognizes the dipeptide lysine-arginine (Lys-Arg). The Lys-Arg codons must be located adjacent to the cDNA sequence so that, following removal of the signal peptide, the recombinant protein will have the correct amino acid at its N terminus. For example, a properly processed and active form of the protein hirudin was synthesized and

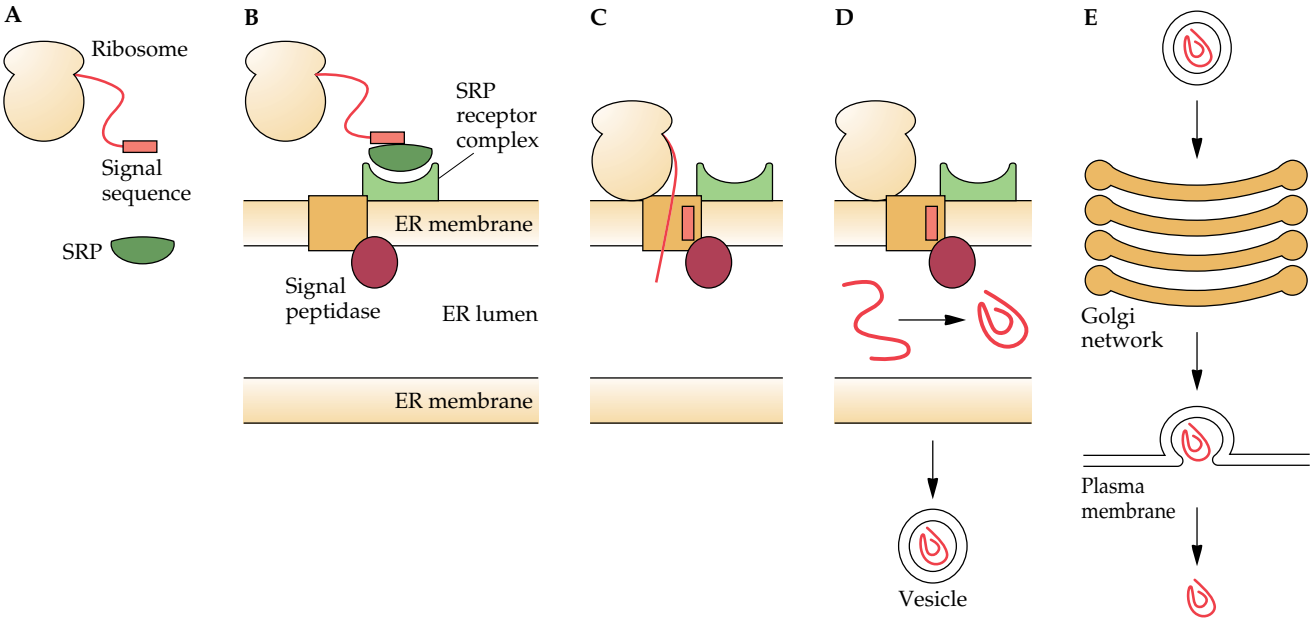


Figure 3.30 Protein secretion pathway in eukaryotes. **(A)** A signal recognition particle (SRP) binds to the signal sequence of a secretory protein. **(B)** The SRP attaches to an SRP receptor on the endoplasmic reticulum (ER) membrane. **(C)** The secretory protein is translocated into the lumen of the ER, and a signal peptidase removes the signal sequence. **(D)** The secretory protein is folded, partially modified, and packaged in a transport vesicle intended for the Golgi network. **(E)** The ER-released vesicle carrying the secretory protein enters the Golgi network at the *cis* face and passes through the Golgi stack, where it is further modified. After it is sorted, a plasma membrane-specific vesicle is formed at the *trans* face of the Golgi network. The secretory transport vesicle fuses with the plasma membrane and releases the secretory protein to the extracellular environment.

secreted by an *S. cerevisiae* strain containing a plasmid vector that had the prepro-α-factor sequence added to the hirudin coding sequence. The gene for hirudin is from an invertebrate, the leech *Hirudo medicinalis*. This protein is a powerful blood anticoagulant that is not immunogenic in humans.

Protein folding. Overexpression of proteins tends to result in the formation of undesirable intracellular aggregates of the proteins, rather than their secretion into the medium, which facilitates purification. Major problems that must be addressed to increase heterologous-protein secretion in yeast cells are the incorrect folding of the polypeptide, the activation of cellular mechanisms to cope with the stress of protein overproduction, and the aberrant processing and release of the protein of interest from the endoplasmic reticulum. Correct protein folding occurs in the endoplasmic reticulum in eukaryotes and is facilitated by a number of different proteins, including molecular chaperones, enzymes that promote disulfide bond formation, signal transduction proteins that monitor the demand and capacity of the protein-folding machinery, and proteases that clear away improperly folded or aggregated proteins (Fig. 3.31). The eukaryotic enzyme protein disulfide isomerase is instrumental in forming the correct

disulfide bonds within a protein. Aberrant disulfide bond formation changes a protein's configuration, which abolishes the protein's activity and causes instability. Poor yields of overexpressed proteins often occur because the capacity of the cell to properly fold and secrete proteins has been exceeded.

Several strategies have been implemented to increase the host cell's capacity to process higher than normal levels of proteins. The overproduction of molecular chaperones and protein disulfide isomerases may increase the yield of active recombinant proteins, especially those with disulfide bonds. To test this hypothesis, the yeast protein disulfide isomerase gene was cloned between the constitutive glyceraldehyde phosphate dehydrogenase promoter and a transcription terminator sequence in a yeast integrating vector, and the entire construct was integrated into a chromosomal site. The modified strain showed a 16-fold increase in protein disulfide isomerase production compared with the wild-type strain. When protein disulfide isomerase-overproducing cells were transformed with a plasmid vector carrying the gene for human platelet-derived growth factor B, there was a 10-fold increase in the secretion of the recombinant protein over that of transformed cells with normal levels of protein disulfide isomerase. The overproduction of protein disulfide isomerase specifically increases the secretion of proteins with disulfide bonds. Higher levels of secreted products were also obtained for the recombinant proteins human erythropoietin, bovine prochymosin, and leech hirudin in *S. cerevisiae* cells that overexpress the chaperone BiP.

Overexpression of the molecular chaperone BiP or protein disulfide isomerase increased the secretion of some heterologous protein; however, overexpression of a single chaperone may not have the desired outcome and, in some instances, may increase the degradation of the target protein. This is because proper protein folding requires the coordinated efforts of many interacting factors (Fig. 3.31). Even when levels of one chaperone are adequate, the levels of cochaperones or cofactors may be limiting. The unfolded-protein response of yeast cells coordinates the expression of several chaperones, as well as cochaperones. When the demand for protein folding exceeds the folding capacity of the endoplasmic reticulum, the unfolded-protein response increases the expression of chaperones, protein disulfide isomerase, and other proteins involved in protein secretion. Engineering the proteins of the unfolded-protein response may be a better approach to increase the overall capacity of the cell to fold proteins in a coordinated way by maintaining appropriate ratios of all factors required. Accumulation of unfolded proteins in the endoplasmic reticulum activates the *S. cerevisiae* transcription factor Hac1, which activates the expression of proteins of the unfolded-protein response, and therefore, expression of Hac1 was targeted for genetic manipulation. Overexpression of Hac1 in *S. cerevisiae* improved secretion of the important industrial enzyme α-amylase, which is used for starch hydrolysis in a wide range of processes.

Protein trafficking. Secretion of recombinant proteins may also be improved by increasing trafficking of proteins within membrane vesicles from the endoplasmic reticulum to the Golgi and from the Golgi to the plasma membrane. One strategy is to increase expression of SNARE proteins (soluble *N*-ethylmaleimide-sensitive factor attachment receptor proteins) that mediate fusion between membranes of protein transport

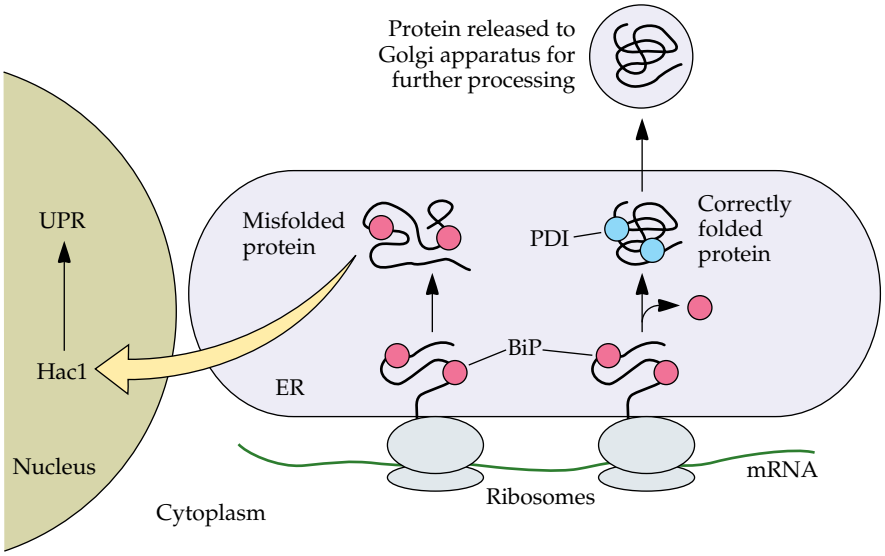


Figure 3.31 Summary of protein folding in the endoplasmic reticulum of yeast cells. During synthesis on ribosomes associated with the endoplasmic reticulum (ER), nascent proteins are bound by the chaperones BiP and calnexin, which aid in the correct folding of the protein. Protein disulfide isomerases (PDI) catalyze the formation of disulfide bonds between cysteine amino acids that are nearby in the folded protein. Quality control systems ensure that only correctly folded proteins are released from the ER. Proteins released from the ER are transported to the Golgi apparatus for further processing. Prolonged binding of BiP to misfolded proteins leads to activation of the *S. cerevisiae* transcription factor Hac1, which controls the expression of several proteins that mediate the unfolded-protein response (UPR). Data from Gasser et al., *Microb. Cell Fact*. **7:**11–29, 2008.

vesicles and membrane-enclosed organelles or the plasma membrane (Fig. 3.32). For example, overexpression of a SNARE protein (Sed5p) that facilitates fusion between endoplasmic reticulum-derived vesicles and the Golgi increased secretion of a recombinant cellulase by 22% compared to parental *S. cerevisiae* cells. Coexpression of another SNARE protein (Sso1p) for fusion of Golgi-derived vesicles with the plasma membrane increased secretion of the cellulase by 130%. Extracellular protein secretion in *S. cerevisiae*, and other eukaryotes, is complex, and the expression levels of many different proteins may require optimization to achieve high levels of recombinant protein secretion.

Other Fungal Expression Systems

Recombinant proteins have been produced successfully in *S. cerevisiae* from cloned genes from many sources. However, in many cases, expression levels are low and protein yields are modest. One of the major drawbacks of using *S. cerevisiae* is the tendency for the yeast to hyperglycosylate heterologous proteins by adding 50 to 150 mannose residues in N-linked oligosaccharide side chains that often alter protein function. Although the initial stages of protein glycosylation in the lumen of the endoplasmic reticulum are similar in yeast and humans, following transfer of the protein to the Golgi apparatus, further processing differs significantly. The outcome is the production of a sialylated protein in humans and a hypermannosylated

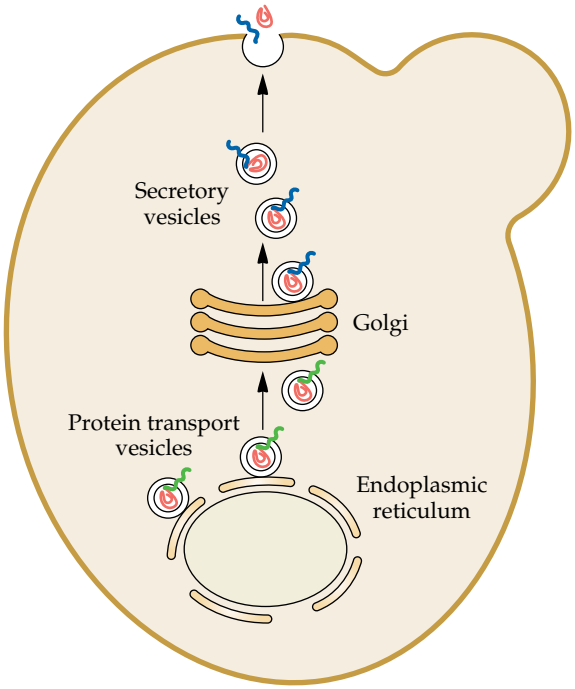


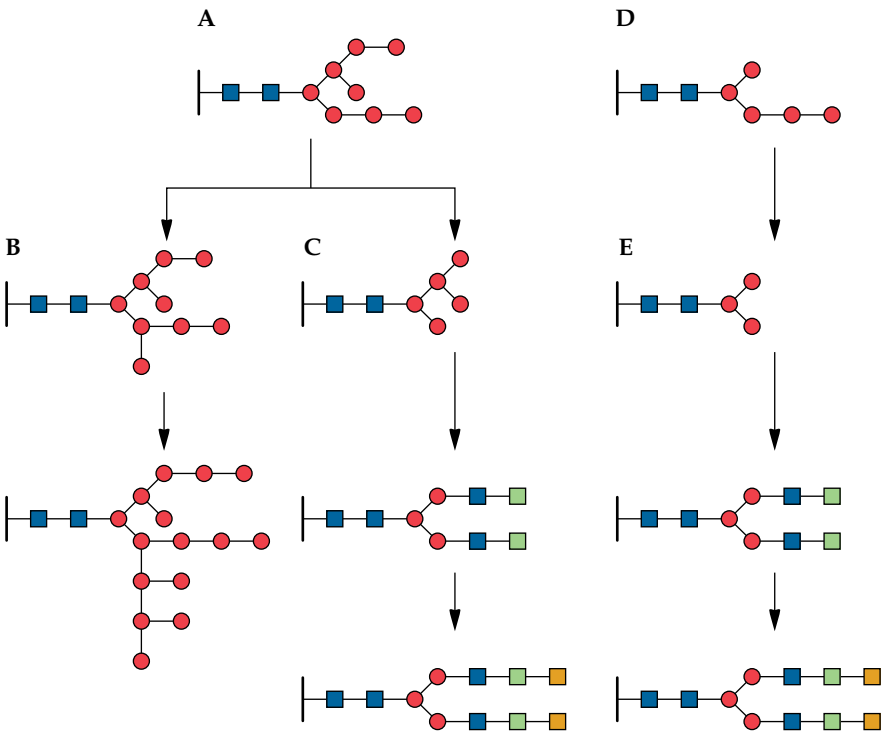
Figure 3.32 Transmembrane SNARE proteins mediate intracellular transport between organelles and the plasma membrane. Some SNARE proteins, such as *S. cerevisiae* Sed5p (green), deliver proteins in transport vesicles from the endoplasmic reticulum to the Golgi apparatus for further processing. Other SNARE proteins, such as Sso1p (blue), traffic secretory vesicles containing proteins from the Golgi to the plasma membrane for extracellular secretion. The SNARE proteins are required for interaction and fusion of the vesicular and target membranes that release the protein. Overexpression of these proteins increased secretion of recombinant proteins in *S. cerevisiae*.

protein in yeast, with α-1,3 bonds between the sugar residues that can make the heterologous protein antigenic (Fig. 3.26). Also, proteins that are designed for secretion are frequently retained in the periplasmic space between the yeast plasma membrane and the yeast cell wall, increasing the time and cost of purification. Finally, *S. cerevisiae* produces ethanol at high cell densities, which can be toxic to the cells and, as a consequence, lowers the quantity of secreted protein. For these reasons, researchers have examined other yeast species and eukaryotic cells that could act as effective host cells for recombinant protein production.

"Humanized" *Pichia pastoris*. The methylotrophic yeast *P. pastoris* can utilize methanol as a source of energy and carbon. It is an attractive host for recombinant protein production because glycosylation occurs to a lesser extent than in *S. cerevisiae* and the linkages between sugar residues are of the α-1,2 type, which are not allergenic to humans. With these natural characteristics as a starting point, a *P. pastoris* strain was extensively engineered with the aim of developing a "humanized" strain that glycosylates proteins in a manner identical to that of human cells. Both human and yeast cells add the same small (10-residue), branched oligosaccharide to nascent proteins in the endoplasmic reticulum

(Fig. 3.33). However, this is the last common precursor between the two cell types. In the Golgi apparatus, yeast cells add an α-1,6 mannose residue to the oligosaccharide, which subsequently leads to hypermannosylation. Mammalian cells, on the other hand, remove some mannose residues from the precursor (trimming) and then sequentially add specific sugars to yield a glycoprotein with an oligosaccharide that terminates in sialic acid. To create a "humanized" strain, the enzyme responsible for addition of the α-1,6 mannose was first eliminated from *P. pastoris* to prevent hypermannosylation. Next, the gene encoding a mannose-trimming enzyme (a mannosidase) from the filamentous fungus *Trichoderma reesei* was inserted into the yeast genome and was found to trim the oligosaccharide to a human-like precursor. Genes encoding enzymes for the sequential addition of sugar residues that terminate the oligosaccharide chains in galactose were also added. It should be noted that the coding sequences for all engineered genes contained a secretion signal for localization of the encoded protein to the Golgi apparatus. Finally, several

Figure 3.33 Differential processing of glycoproteins in *P. pastoris*, humans, and "humanized" *P. pastoris*. Initial additions of sugar residues to glycoproteins in the endoplasmic reticulum are similar in human and *P. pastoris* cells **(A)**. However, further *N*-glycosylation in the Golgi apparatus differs significantly between the two cell types. N-glycans are hypermannosylated in *P. pastoris* **(B)**, while in humans, mannose residues are trimmed and specific sugars are added, leading to termination of the oligosaccharide in sialic acid **(C)**. *P. pastoris* cells have been engineered to produce enzymes that process glycoproteins in a manner similar to that of human cells. In "humanized" *P. pastoris*, a recombinant glycoprotein produced in the endoplasmic reticulum **(D)** is transported to the Golgi apparatus, where it is further processed to yield a properly sialylated glycoprotein **(E)**. Blue squares, *N*-acetylglucosamine; red circles, mannose; green squares, galactose; orange squares, sialic acid. Data from Hamilton and Gerngross, *Curr. Opin. Biotechnol*. **18:**387–392, 2007.



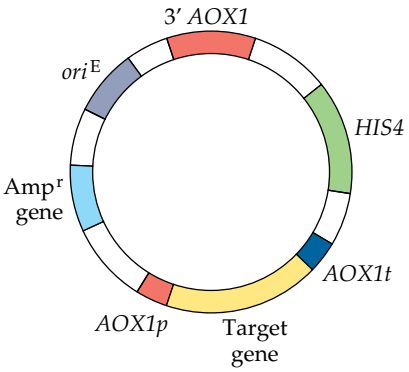


Figure 3.34 *P. pastoris* integrating expression vector. The target gene is cloned between the promoter (*AOX1p*) and termination-polyadenylation sequence (*AOX1t*) of the *P. pastoris* alcohol oxidase 1 gene. The *HIS4* gene encodes a functional histidinol dehydrogenase of the histidine biosynthesis pathway. The ampicillin resistance (Amp^r) gene and an origin of replication (ori^E) function in *E. coli*. The segment marked 3′ *AOX1* is a piece of DNA from the 3′ end of the alcohol oxidase 1 gene of *P. pastoris*. A double recombination event between the *AOX1p* and 3′ *AOX1* regions of the vector and the homologous segments of chromosome DNA results in the integration of the DNA carrying the target gene and the *HIS4* gene into a *P. pastoris* chromosome (Fig. 3.35B).

genes (mouse and human) for proteins that catalyze the synthesis, transport to the Golgi apparatus, and addition of sialic acid to the terminal galactose on the protein precursor were inserted into the *P. pastoris* genome. Properly sialylated recombinant proteins, including monoclonal antibodies such as trastuzumab (Herceptin), which is used for the treatment of breast cancer (see chapter 5), have been produced by the "humanized" *P. pastoris*. Accurate glycan structure is critically important for therapeutic antibodies, because it is required for interaction with immune cells that kill cancerous cells.

***P. pastoris* expression vectors.** During growth on methanol, enzymes required for catabolism of this substrate are expressed at very high levels with alcohol oxidase, the first enzyme in the methanol utilization pathway, encoded by the gene *AOX1*, representing as much as 30% of the cellular protein. Transcription of *AOX1* is tightly regulated; in the absence of methanol, the *AOX1* gene is completely turned off but responds rapidly to the addition of methanol to the medium. Therefore, the *AOX1* promoter is an excellent candidate for producing large amounts of recombinant protein under controlled conditions. Moreover, the induction of the cloned gene can be timed to maximize recombinant protein production during large-scale fermentations. In contrast to *S. cerevisiae*, *P. pastoris* does not synthesize ethanol; therefore, very high cell densities of *P. pastoris* are attained, with the concomitant secretion of large quantities of protein. In addition, *P. pastoris* normally secretes very few proteins, thus simplifying the purification of secreted recombinant proteins.

Many *P. pastoris* expression vectors have been devised, each one having more or less the same format. The basic features include a target gene under the control of promoter and transcription termination sequences from the *P. pastoris AOX1* gene, an *E. coli* origin of replication and selectable marker gene, and a yeast selectable marker gene (Fig. 3.34). The addition of a signal sequence from either the *P. pastoris* phosphatase *PHO1* gene or another yeast gene facilitates the secretion of a recombinant protein. To avoid the problems of plasmid instability during long-term growth, most heterologous genes are designed to be integrated into the *P. pastoris* genome, often within the *AOX1* gene or the *HIS4* gene for histidine biosynthesis. Both the target gene and a yeast selectable marker gene are inserted together into a specific chromosome site by either a single (Fig. 3.35A) or a double (Fig. 3.35B) recombination event. The *P. pastoris* expression system has been used to produce more than 100 different biologically active proteins from bacteria, fungi, invertebrates, plants, and mammals, including humans. Many of these proteins, such as the hepatitis B virus surface antigen, human serum albumin, and bovine lysozyme, are identical to the native proteins and thus authentic.

Other yeast expression systems. Authentic heterologous proteins for industrial and pharmaceutical uses have also been generated in other yeasts. For example, the cDNAs for the α- and β-globin chains of human hemoglobin A were produced in the methylotrophic yeast *Hansenula polymorpha*. The thermotolerant dimorphic yeasts *Arxula adeninivorans* and *Yarrowia lipolytica* have demonstrated promising potential as hosts for high levels of heterologous-protein expression. These yeasts can grow at temperatures of up to 48°C and can survive at higher temperatures (55°C) for several hours. At higher temperatures, the fungi grow in a mycelial

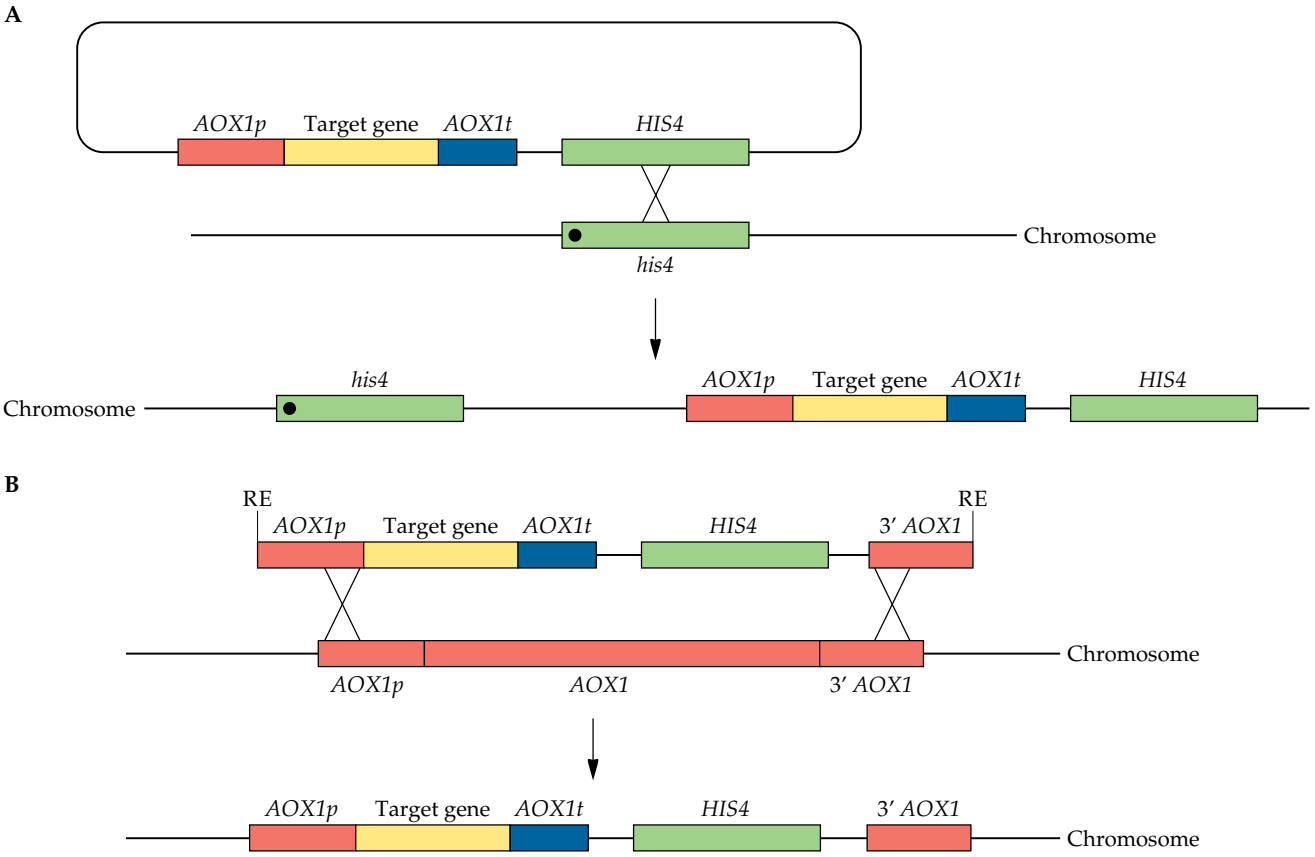


Figure 3.35 Integration of DNA into a specific *P. pastoris* chromosome site by single **(A)** or double **(B)** recombination. **(A)** A single recombination (×) between the *HIS4* gene of an intact circular plasmid and a chromosome *his4* mutant gene results in the integration of the entire vector, including the target gene with the promoter (*AOX1p*) and transcription–polyadenylation sequence (*AOX1t*) from the *AOX1* gene, into the chromosome. The integrated DNA is flanked by recombined mutant *his4* and functional *HIS4* genes. The dot in the *his4* gene represents the mutation. **(B)** A double recombination (×) between the cloned *AOX1p* and 3′ *AOX1* DNA segments of a restriction endonuclease (RE) linearized DNA fragment from the vector and the corresponding homologous chromosome regions results in the integration of the target gene and a functional *HIS4* gene. The chromosome *AOX1* coding region is lost as a result of the recombination event.

form and revert to budding cells below 42°C. Some secreted proteins, such as glucoamylase and invertase, are produced at higher levels in mycelia. Cell morphology also influences posttranslational modification, with O-linked glycosylation predominating in budding cells, while N-glycosylation occurs in both mycelial and budding cells. An additional advantage of *A. adeninivorans* is the ability to grow on a wide range of inexpensive carbon and nitrogen sources.

It is often necessary to try several host types in order to find the one that produces the highest levels of a biologically active recombinant protein. Differences in the processing and productivity of a particular protein can occur among different yeast strains. For example, both *S. cerevisiae* and *H. polymorpha* produced a truncated version of the human protein interleukin-6 (IL-6), whereas *A. adeninivorans* produced a full-length

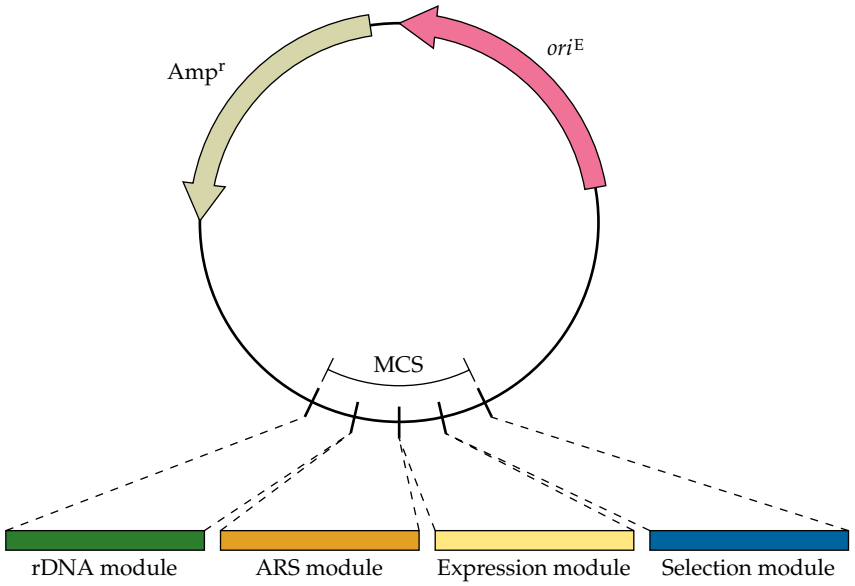


Figure 3.36 A wide-range yeast vector system for expression of heterologous genes in several different yeast hosts. The basic vector contains a multiple cloning site (MCS) for insertion of selected modules containing appropriate sequences for chromosomal integration (rDNA module), replication (ARS module), selection (selection module), and expression (expression module) of a target gene in a variety of yeast host cells (Table 3.13 shows examples of interchangeable modules). Sequences for maintenance (*ori*^{E}) and selection (Amp^{r}) of the vector in *E. coli* are also included.

Table 3.13 Examples of modules available for wide-range yeast vector systems[a]

Module and gene	Donor
Integration	
25S rDNA	*A. adeninivorans*
18S rDNA	*A. adeninivorans*
18S rDNA	*H. polymorpha*
Replication	
2 µm ARS	*S. cerevisiae*
ARS1	*S. cerevisiae*
HARS[b]	*H. polymorpha*
Selection	
URA3	*S. cerevisiae*
LEU2	*S. cerevisiae*
ALEU2m	*A. adeninivorans*
ATRP1m	*A. adeninivorans*
HIS4	*P. pastoris*
Expression	
MOX promoter	*H. polymorpha*
AOX1 promoter	*P. pastoris*
TEF1 promoter	*A. adeninivorans*
GAA promoter	*A. adeninivorans*
RPS7 promoter	*Y. lipolytica*

[a]Data from Gellissen et al., *FEMS Yeast Res.* **5:**1079–1096, 2005.

[b]HARS, *Hansenula* ARS.

version of the protein. The construction of a wide-range yeast vector for expression in several fungal species has facilitated this trial-and-error process (Fig. 3.36). The basic vector contains features for propagation and selection in *E. coli* and a multiple cloning site for insertion of interchangeable modules that are chosen for a particular yeast host, including a promoter to drive expression of the heterologous gene, a sequence for vector integration into a specific site in the fungal genome, a suitable origin of replication, and selectable markers to complement a range of nutritional auxotrophies or to confer resistance to antifungal compounds, such as hygromycin B (Table 3.13). In other words, by selecting from a range of available modules, customized vectors can be rapidly and easily constructed for expression of the same gene in several different yeast cells to determine which host is optimal for heterologous-protein production.

Expression in filamentous fungi. Expression vectors with appropriate transcription and translation control elements for the expression of recombinant proteins in filamentous fungi are also commercially available. Distinct from unicellular yeasts, filamentous fungi are multicellular, microscopic fungi that produce long, branching strands of cells called hyphae. This group of fungi includes the common mold genera *Penicillium*, *Rhizopus*, *Trichoderma*, and *Aspergillus*. Many species of these genera of filamentous fungi are a rich natural resource for commercially important metabolites and enzymes and have also been used as cell factories for the production of recombinant proteins for the food, beverage, pulp and paper, and pharmaceutical industries (Table 3.14). Similar to yeast, filamentous fungi can grow rapidly on inexpensive media, secrete large amounts of

Table 3.14 Some recombinant proteins produced by filamentous fungal expression systems

Recombinant protein	Host cell	Main application
α-Amylase	*Aspergillus niger*, *Aspergillus oryzae*	Starch processing, food industry
Aspartyl protease	*Aspergillus nidulans*, *A. oryzae*	Food industry
Cellulase	*Trichoderma reesei*	Textile, pulp and paper industries
Chymosin	*A. niger*	Food industry
Immunoglobulin G	*A. niger*	Pharmaceutical industry
Insulin	*A. niger*	Pharmaceutical industry
Interleukin-6	*A. niger*	Pharmaceutical industry
Laccase	*A. niger*, *T. reesei*	Textile, pulp and paper industries
Manganese peroxidase	*A. niger*	Chemical industry
Lactoferrin	*A. oryzae*	Pharmaceutical industry
Lipase, thermophilic	*A. oryzae*	Detergent
Lysozyme	*A. niger*	Pharmaceutical industry
Phytase	*T. reesei*	Food industry
Xylanase	*A. niger*, *T. reesei*	Textile, pulp and paper, food industries

proteins, process eukaryotic mRNA, and carry out many posttranslational modifications. However, an additional advantage of using filamentous fungi as hosts for the production of mammalian proteins is their ability to add mammalian-like sugars to proteins. Prior to transformation, single protoplast cells are prepared from the multicellular hyphae by enzymatically removing the cell wall. Alternatively, intact conidia (external spores) may be transformed. Vectors carrying the target genes are commonly introduced into protoplasts or conidia by electroporation (chapter 2), microprojectile bombardment (chapter 10), agitation with glass beads, or vacuum infiltration of *Agrobacterium tumefaciens* (chapter 10). Following selection of transformed cells, hyphae are regenerated.

In sum, fungal expression systems play an important role in the production of heterologous proteins for research, industrial, and medical applications. However, experience has shown that no one system is able to produce an authentic version of every heterologous protein. For this and other reasons, gene expression systems that use insect or mammalian cells have been developed.

Baculovirus–Insect Cell Expression Systems

Baculoviruses are a large, diverse group of double-stranded DNA viruses that specifically infect arthropods, including many insect species, and are not infectious to other animals. During the infection cycle, two forms of baculovirus (occluded form and budded form) are produced (Fig. 3.37). Infection is initiated when the occluded form of the virus is ingested by the insect larvae. In this form, the viral nucleocapsids (virions) are clustered in a matrix that is made up of the protein polyhedrin, which protects the virions from degradation in the environment. The occluded virions packaged in this protein matrix are referred to as a polyhedron. Following ingestion, the virus is taken up into the midgut of the insect, the polyhedrin matrix dissolves due to the alkaline gut environment, and the virions enter midgut cells. The virions migrate to the nucleus, where they are uncoated, releasing

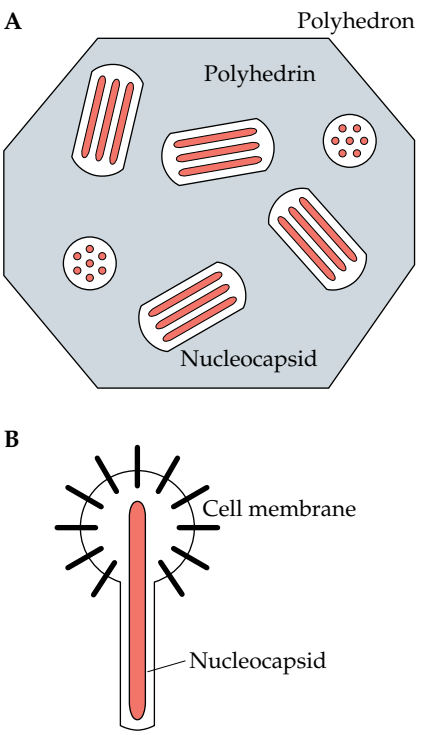


Figure 3.37 Forms of the baculovirus AcMNPV. **(A)** In the occluded form, clusters of nucleocapsids (occluded virions) are embedded in a polyhedrin matrix. The structure is referred to as a polyhedron. **(B)** During budding, a nucleocapsid becomes enveloped by the membrane of an infected cell.

the DNA for genome replication, synthesis of viral proteins, and production of new virions. Within the insect midgut, the infection can spread from cell to cell as viral particles (single nucleocapsids) bud off from an infected cell and infect other midgut cells. This form of the virus, known as the budded form, is not embedded in a polyhedrin matrix and is not infectious to other individual insect hosts, although it can infect cultured insect cells. Plaques produced in insect cell cultures by the budded form of baculovirus have a different morphology from the occluded form. During the late stages of the infection cycle in the insect host, about 36 to 48 hours after infection, the polyhedrin protein is produced in massive quantities and continues for 4 to 5 days, until the infected cells rupture and the host organism dies. Occluded virions are released and can infect new hosts.

The promoter for the polyhedrin (*polyh*) gene is exceptionally strong, and transcription from this promoter can account for as much as 25% of the mRNA produced in cells infected with the virus. Moreover, the polyhedrin protein is not required for virus production. Consequently, it was reasoned that replacement of the polyhedrin gene in the viral genome with a coding sequence for a heterologous protein, followed by infection of cultured insect cells, would result in the production of large amounts of the heterologous protein. Furthermore, because of the similarity of posttranslational modification systems between insects and mammals, it was thought that the recombinant protein would mimic closely, if not precisely, the authentic form of the original protein. Baculoviruses have been highly successful as delivery systems for introducing target genes for production of high levels of heterologous proteins in insect cells. More than a thousand different proteins have been produced using this system, including several vaccines that have been approved for veterinary or human use (Table 3.15). Notably, an influenza vaccine containing recombinant hemagglutinin, a glycoprotein found on the surface of influenza virus, is produced in insect cells.

The specific baculovirus that has been used extensively as an expression vector is *Autographa californica* multiple nuclear polyhedrosis virus (AcMNPV). *A. californica* (the alfalfa looper) and over 30 other insect species are infected by AcMNPV. This virus also grows well on many insect cell lines. The most commonly used cell line for genetically engineered AcMNPV is derived from the fall armyworm, *Spodoptera frugiperda*. In these cells, the polyhedrin promoter is exceptionally active, and during infections with wild-type baculovirus, high levels of polyhedrin are synthesized.

Table 3.15 Some recombinant products produced by baculovirus–insect cell expression systems[a]

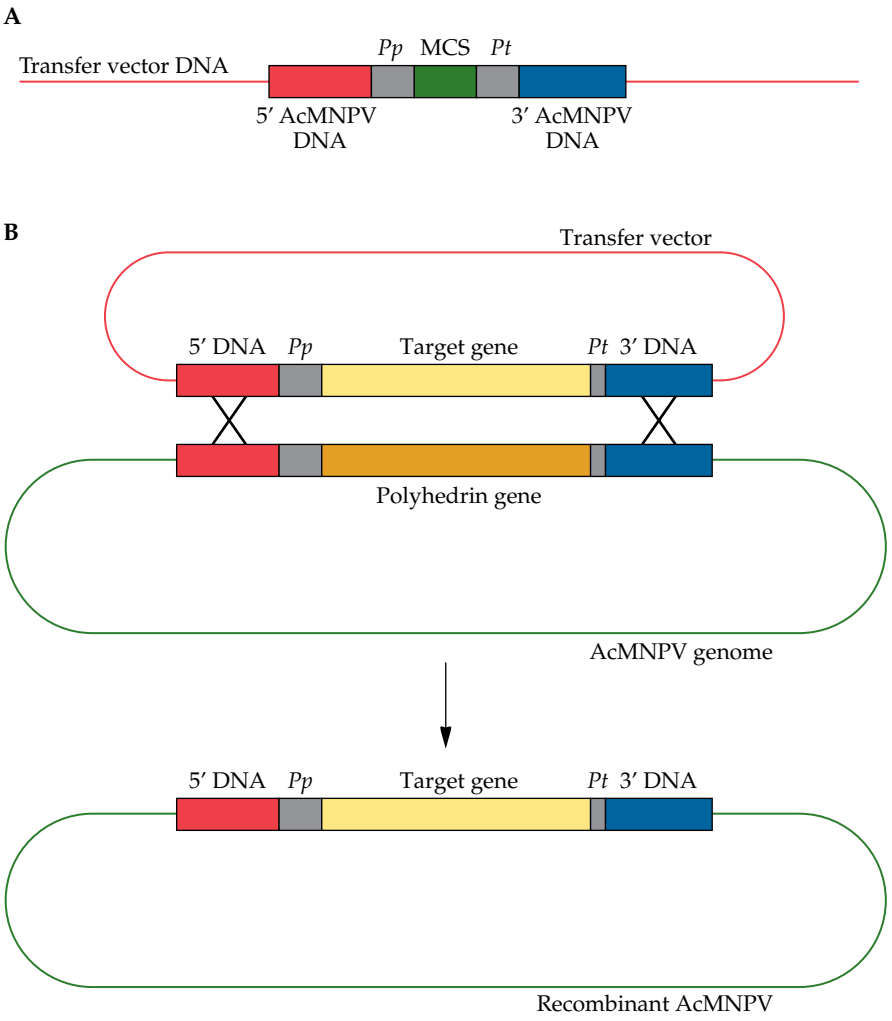
Recombinant product	Use	Year available
E2 glycoprotein	Vaccine against swine fever virus	2001
Porcine circovirus ORF2	Vaccine against porcine circovirus type 2	2005
Human papillomavirus L1 proteins (serotypes 16 and 18)	Vaccine against cervical cancer	2007
Prostatic acid phosphatase-granulocyte macrophage colony-stimulating factor	Immunotherapy against prostate cancer	2010
Adeno-associated virus vector with lipoprotein lipase transgene	Gene therapy against familial lipoprotein lipase deficiency	2012
Influenza hemagglutinin	Influenza vaccine	2013

[a]Data from van Oers et al., *J. Gen. Virol*. **96:**6–23, 2015.

Baculovirus Expression Vectors

The first step in the production of a recombinant AcMNPV that is used to deliver the target gene into the insect host cell is to create a transfer vector. The transfer vector is an *E. coli*-based plasmid that carries a segment of DNA from AcMNPV (Fig. 3.38A) consisting of the polyhedrin promoter region with flanking upstream AcMNPV DNA, a multiple cloning site, the polyhedrin termination and polyadenylation signal regions with flanking

Figure 3.38 Insertion of a heterologous gene into a baculovirus genome. **(A)** Organization of the expression unit of a baculovirus (AcMNPV) transfer vector. The target gene is inserted into the multiple cloning site (MCS) that lies between the polyhedrin gene promoter (*Pp*) and polyhedrin gene transcription termination (*Pt*) sequences. The AcMNPV DNA upstream from the polyhedrin promoter (5′ AcMNPV DNA) and downstream from the polyhedrin transcription termination sequence (3′ AcMNPV DNA) provides sequences for integration of the expression unit by homologous recombination into an AcMNPV genome. **(B)** Replacement of the polyhedrin gene in the AcMNPV genome with an expression unit from a transfer vector. A double crossover event (×) between homologous DNA segments of the transfer vector and the AcMNPV genome results in the integration of the expression unit into the AcMNPV genome.



downstream AcMNPV DNA. The coding region for the polyhedrin gene has been deleted from this segment of DNA. The flanking AcMNPV DNA sequences provide regions for homologous recombination with AcMNPV. A heterologous gene is inserted into the multiple cloning site between the polyhedrin promoter and termination sequences, and the transfer vector is propagated in *E. coli*.

Next, insect cells in culture are cotransfected with AcMNPV DNA and the transfer vector carrying the cloned gene. Within some of the doubly transfected cells, a double-crossover recombination event occurs at homologous polyhedrin gene sequences on the transfer vector and in the AcMNPV genome, and the cloned gene with the polyhedrin promoter and termination regions becomes integrated into the AcMNPV DNA (Fig. 3.38B) with the concomitant loss of the polyhedrin gene. Virions lacking the polyhedrin gene produce distinctive zones of cell lysis (occlusion-negative plaques), from which recombinant baculovirus is isolated.

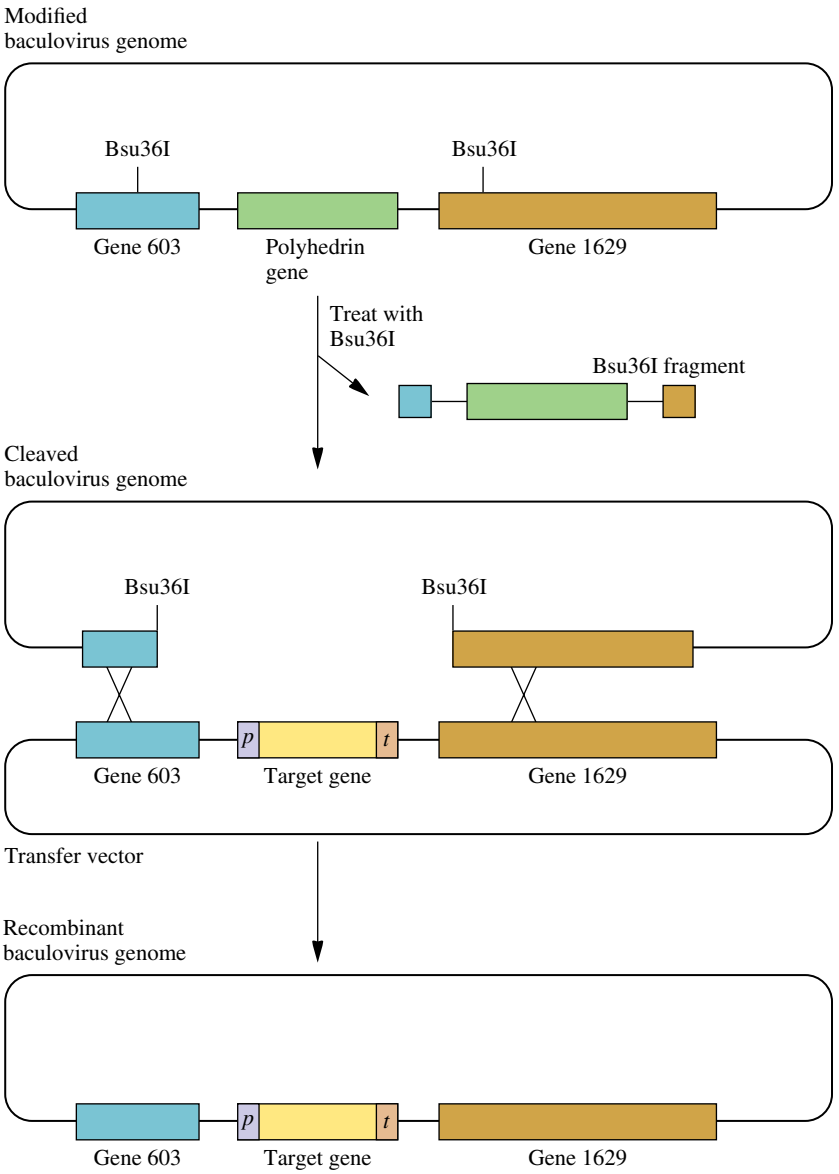
Linearization of the normally circular 134-kbp AcMNPV genome before transfection into insect cells substantially increases the frequency of recombinant plaques. The AcMNPV genome was engineered with two Bsu36I sites that were placed on either side of the polyhedrin gene (Fig. 3.39). One is in gene 603 and the other is in gene 1629, which is essential for viral replication. When DNA from this modified baculovirus is treated with Bsu36I and transfected into insect cells, no viral replication occurs because a segment of the essential gene (1629) is missing. As part of this system, a transfer vector is constructed with the target gene between intact versions of gene 603 and gene 1629. This transfer vector is introduced into insect cells that were previously transfected with linearized, replication-defective AcMNPV DNA that is missing the segment between the two Bsu36I sites. A double-crossover event both reestablishes a functional version of gene 1629 and incorporates the cloned gene into the AcMNPV genome (Fig. 3.39). With this system, over 90% of the baculovirus plaques are recombinant.

To eliminate the need to use plaque assays to identify and purify recombinant viruses, several methods have been developed that introduce the target gene into the baculovirus genome at a specific nucleotide sequence by recombination, either in an intermediate bacterial host, such as *E. coli*, or *in vitro* outside a host cell. Transfection of insect cells is required only for the production of the heterologous protein. AcMNPV DNA can be maintained in *E. coli* as a plasmid known as a bacmid, which is a baculovirus–plasmid hybrid molecule. In addition to AcMNPV genes, the bacmid contains an origin of replication for propagation in *E. coli*, a kanamycin resistance gene for selection of the bacmid, and an integration site (attachment site) that is inserted into the *lacZ′* gene without impairing its function (Fig. 3.40A and B). Another component of this system is the transfer vector that carries the target gene cloned between the polyhedrin promoter and a terminator sequence. In the transfer vector, the target gene expression unit (expression cassette) and a gentamicin resistance gene are flanked by DNA attachment sequences that can bind to the attachment site in the bacmid (Fig. 3.40B). An ampicillin resistance gene lies outside the expression cassette for selection of the transfer vector.

Bacterial cells carrying a bacmid are cotransformed with the transfer vector and a helper plasmid that encodes the specific proteins (transposition proteins) that mediate recombination between the attachment sites on the

transfer vector and on the bacmid and that carries a tetracycline resistance gene (Fig. 3.40B). After recombination, the DNA segment that is bounded by the two attachment sites on the transfer vector (the expression cassette carrying the target gene) is transposed into the attachment site on the bacmid, destroying the reading frame of the *lacZ′* gene (Fig. 3.40C).

Figure 3.39 Production of recombinant baculovirus. Single Bsu36I sites are engineered into gene 603 and a gene (1629) that is essential for AcMNPV replication. These genes flank the polyhedrin gene in the AcMNPV genome. After baculovirus genomic DNA with two engineered Bsu36I sites is treated with Bsu36I, the segment between the Bsu36I sites is deleted. Insect cells are cotransfected with a Bsu36I-treated baculovirus DNA and a transfer vector with a target gene under the control of the promoter (*p*) and terminator (*t*) elements of the polyhedrin gene and the complete sequences of both genes 603 and 1629. A double crossover event (×) generates a recombinant baculovirus with a functional gene 1629. With this system, almost all of the progeny baculoviruses are recombinant.



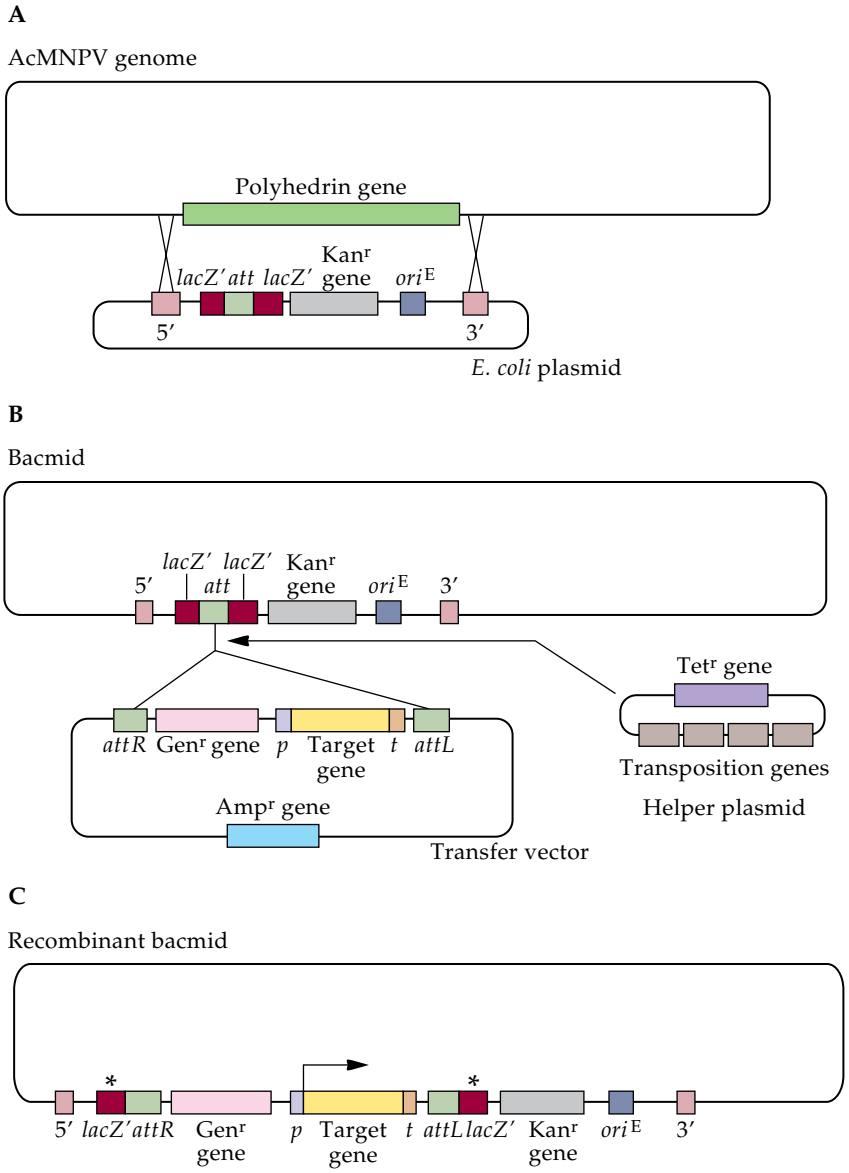


Figure 3.40 Construction of a recombinant bacmid. **(A)** An *E. coli* plasmid is incorporated into the AcMNPV genome by a double crossover event (×) between DNA segments (5′ and 3′) that flank the polyhedrin gene to create a shuttle vector (bacmid) that replicates in both *E. coli* and insect cells. The gene for resistance to kanamycin (Kan^r), an attachment site (*att*) that is inserted in-frame in the *lacZ′* sequence, and an *E. coli* origin of replication (*ori*^{E}) are introduced as part of the plasmid DNA. **(B)** The transposition proteins encoded by genes of the helper plasmid facilitate the integration (transposition) of the DNA segment of the transfer vector that is bounded by two attachment sequences (*attR* and *attL*). The gene for resistance to gentamicin (Gen^r) and a target gene that is under the control of the promoter (*p*) and transcription terminator (*t*) elements of the polyhedrin gene are inserted into the attachment site (*att*) of the bacmid. The helper plasmid and transfer vector carry the genes for resistance to tetracycline (Tet^r) and ampicillin (Amp^r), respectively. **(C)** The recombinant bacmid has a disrupted *lacZ′* gene (*) and therefore cells that are transfected with a recombinant bacmid are not able to produce functional β-galactosidase. The right-angled arrow denotes the site of initiation of transcription of the target gene after transfection of the recombinant bacmid into an insect cell.

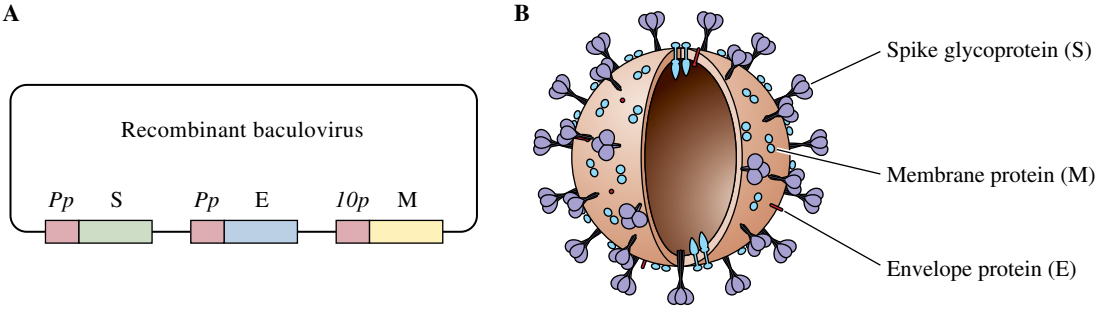
Consequently, bacteria with recombinant bacmids produce white colonies in the presence of IPTG and 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal). White colonies that are resistant to kanamycin and gentamicin and sensitive to both ampicillin and tetracycline carry only a recombinant bacmid and no transfer and helper plasmids. After all of these manipulations, the integrity of the cloned gene can be confirmed by PCR. Finally, recombinant bacmid DNA can be transfected into insect cells, where the cloned gene is transcribed and the heterologous protein is produced.

The simultaneous expression of two or more cloned genes can lead to the formation of functional multimeric protein complexes. This can be accomplished by transfecting insect cells with a single recombinant baculovirus expressing multiple proteins. AcMNPV is particularly amenable to carrying large insertions, up to 38 kb, with several foreign genes due to its flexible envelope. In one study, the genes for three different envelope structural proteins from the human severe acute respiratory syndrome coronavirus were expressed simultaneously at a high level from a single baculovirus vector (Fig. 3.41A). The proteins were found to assemble spontaneously and stably into virus-like particles (Fig. 3.41B). Virus-like particles, composed of the assembled envelope or protein coat of a virus but without the nucleic acid genome, are the basis for some subunit vaccines (see chapter 7). One such commercially available vaccine (Cervarix), produced using the baculovirus-insect system, protects against cervical cancer caused by human papillomavirus types 16 and 18.

Mammalian Glycosylation and Processing of Proteins in Insect Cells

Although insect cells can process proteins in a manner similar to that of other eukaryotes, some mammalian proteins produced in *S. frugiperda* cell lines are not authentically glycosylated. For example, insect cells do not normally add galactose and terminal sialic acid residues to N-linked glycoproteins. Where these glycans are normally added to mannose residues during the processing of some proteins in mammalian cells, insect

Figure 3.41 Production of virus-like particles using a baculovirus–insect cell expression system. **(A)** Viral spike (S), membrane (M), and envelope (E) proteins, which compose the envelope of the human SARS-CoV, are expressed in insect cells from a single recombinant baculovirus vector carrying all three viral genes. **(B)** Following expression, the S, M, and E proteins assemble to form a SARS-CoV virus-like particle that resembles the envelope of the original virus but does not contain the viral genetic material. The virus-like particle is a candidate vaccine for protection against SARS. *Pp*, polyhedrin promoter; *10p*, baculovirus p10 promoter.



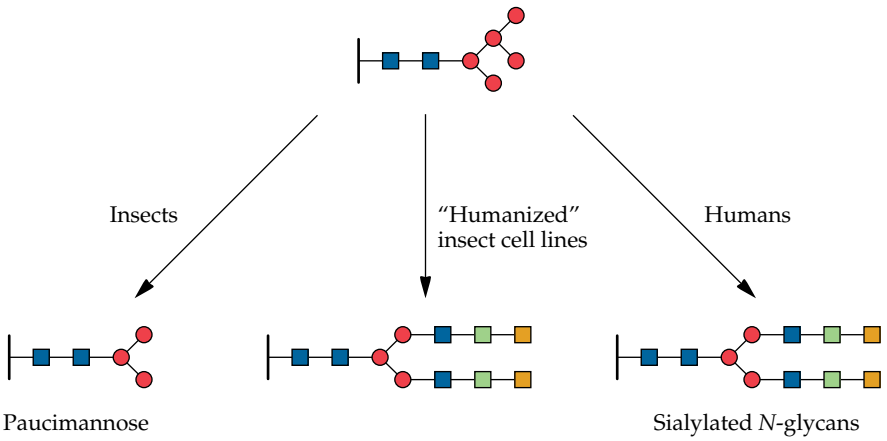


Figure 3.42 *N*-glycosylation of proteins in the Golgi apparatus of insect, human, and "humanized" insect cells. While the sugar residues added to *N*-glycoproteins in the endoplasmic reticulum are similar in insect and human cells, further processing in the Golgi apparatus yields a trimmed oligosaccharide (paucimannose) in insect cells and an oligosaccharide that terminates in sialic acid in human cells. To produce recombinant proteins for use as human therapeutic agents, "humanized" insect cells have been engineered to express several enzymes that process human glycoproteins accurately. Blue squares, *N*-acetylglucosamine; red circles, mannose; green squares, galactose; orange squares, sialic acid.

cells will trim the oligosaccharide to produce paucimannose (Fig. 3.42). Consequently, the baculovirus system cannot be used for the production of several important mammalian glycoproteins. To ensure the production of "humanized" glycoproteins with accurate glycosylation patterns, insect cell lines have been constructed that express a combination of mammalian glycosyltransferases (Fig. 3.42).

Further improvements to prevent undesirable processing of heterologous proteins in insect cells are the removal of the genes encoding chitinase and the protease cathepsin from the AcMNPV genome. Cathepsin is normally produced late in the infection cycle to facilitate the release of new virions from the insect host. It also reduces the yield of heterologous proteins through proteolytic cleavage. Chitinase is produced in conjunction with cathepsin and is thought to function in the proper folding of cathepsin and in the degradation of the host insect exoskeleton. It is secreted at very high levels from baculovirus-infected insect host cells and can compete with secreted target proteins for the secretory apparatus, thereby reducing yields of the target protein. Coexpression of chaperones to ensure proper folding of the target protein has also resulted in increased yields of functional heterologous proteins.

Baculovirus-infected *S. frugiperda* cells generally die within about three to four days following infection, which limits the period of time for recombinant protein production. To extend the viability of baculovirus-infected cells, the gene encoding the viral antiapoptotic protein vankyrin was integrated into the chromosome of the host insect cells. Constitutive expression of vankyrin delayed cell death and increased yields of a recombinant protein three-fold at three days postinfection. Transcription of host genes ceases soon after viral infection, and therefore, to further increase recombinant protein production, the vankyrin gene, under the

control of the strong baculovirus p10 promoter, was inserted into a baculovirus vector. At 3 days postinfection, about 90% of host cells infected with the vankyrin-encoding baculovirus vector were viable, compared with about 55% of cells that did not express vankyrin. Production of substantially higher levels of recombinant human erythropoietin in vankyrin-expressing cells compared to cells that did not produce vankyrin was maintained for 10 days. A synergistic effect on yields of a recombinant protein was observed in insect cells carrying both a chromosomal and a baculovirus-borne copy of the vankyrin gene.

Mammalian Cell Expression Systems

Between 2014 and 2018, 84% of recombinant proteins approved for human therapeutic use were produced in cultured mammalian cells (the remainder were produced in *E. coli* and *S. cerevisiae*, and to a lesser extent in insect cell cultures). Most of these were antibodies that require accurate post-translational modification. Chinese hamster ovary (CHO) cells are most commonly used because they produce proteins with human-like glycans and have been adapted for growth in high-density suspension cultures in serum-free medium, which not only reduces costs, but also facilitates purification of the target protein and reduces the risk of contamination with animal-derived material. They are receptive to transfection and can achieve long-term (stable) gene expression and high yields of heterologous proteins. Other host cell lines are derived from mouse myelomas, baby hamster kidney (BHK), and human embryo kidney (HEK 293). Human host cells have the advantage of authentically glycosylating human proteins; however, they are susceptible to infection by human viruses, which is a biosafety risk if the recombinant proteins will be used in humans.

Mammalian cells have been used for some time to produce therapeutic proteins, and vectors carrying suitable expression signals have been developed. Current efforts are aimed at improving productivity through the development of high-producing cell lines, increasing the stability of production over time, and increasing expression by manipulating the chromosomal environment in which the recombinant genes are integrated.

Vector Design

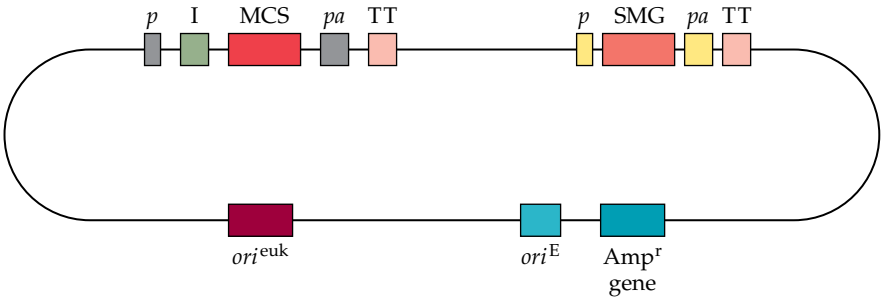
Most cloning vectors constructed for the expression of heterologous genes in mammalian cells are based on the genomes of viruses that infect mammalian cells. Many vectors are derived from the simian vacuolating virus 40 (SV40) genome, which can be replicated in several mammalian species. The genome of this virus is a small (5.2 kb) double-stranded DNA molecule that encodes four proteins. One protein (T-antigen) expressed early in the infection cycle controls replication of viral DNA. The other structural proteins expressed later in the infection cycle from overlapping reading frames compose the capsid. For use as a cloning vector, much of the viral sequence is removed and replaced with a target gene under the control of appropriate mammalian expression signals; the SV40 origin of replication is retained. Although many cloning vectors are based on SV40 DNA, its use is restricted to small inserts (~5 kb) because only a limited amount of DNA can be packaged into the viral capsid. Other vectors that can accommodate larger amounts of cloned DNA are derived from adenovirus, bovine papillomavirus, which can be maintained as a multicopy

plasmid in some mammalian cells, and adeno-associated virus, which can integrate into specific sites in the host chromosome.

Baculovirus delivery systems have also been developed to express recombinant proteins in mammalian cells. Although baculoviruses cannot replicate in mammalian cells, they can be transduced into these cells with a transduction efficiency reaching 100% in some cases, where they enter the nucleus and transiently express target genes that were inserted in the viral genome. In addition, stable artificial chromosome expression systems have been developed for some mammalian cell lines. These carry specific sequences for integration of one or more copies of a target gene by recombination.

Transcription signals. All mammalian expression vectors tend to have similar features and are not very different in design from other eukaryotic expression vectors. A typical mammalian expression vector (Fig. 3.43) contains a eukaryotic origin of replication, usually from an animal virus, such as SV40. The promoter sequences that drive expression of the cloned gene(s) and the selectable marker gene(s), and the transcription termination sequences (polyadenylation signals), must be eukaryotic and are frequently derived from either human viruses or mammalian genes. Strong constitutive promoters, such as the SV40 early promoter or the human cytomegalovirus (CMV) intermediate-early promoter and enhancer, both of which function in numerous cell types, and efficient polyadenylation signals are preferred. Expression of a target gene may be increased by placing the sequence for an intron between the promoter and the multiple cloning site within the transcribed region. Inducible promoters are often used when continuous synthesis of the heterologous protein is toxic to the host cell. These may be constructed by inserting a repressor-binding site, such as the *lac* operator from the *E. coli lac* operon, between the promoter and the target gene (Fig. 3.44). Included on the vector is the gene encoding the LacI repressor that binds to the *lac* operator, thereby blocking the promoter and preventing expression of the target gene. Addition of IPTG and/or lactose, which readily penetrate mammalian cells, prevents the LacI repressor from binding to the *lac* operator and induces expression of the target gene.

Figure 3.43 Generalized mammalian expression vector. Target genes inserted into the multiple cloning site (MCS) and selectable marker gene (SMG) are under the control of eukaryotic promoter (*p*), polyadenylation (*pa*), and transcription termination (TT) sequences. An intron (I) enhances the production of recombinant protein. Propagation of the vector in *E. coli* and mammalian cells depends on the origins of replication ori^{E} and ori^{euk}, respectively. The ampicillin resistance (Amp^{r}) gene is used for selecting transformed *E. coli*.



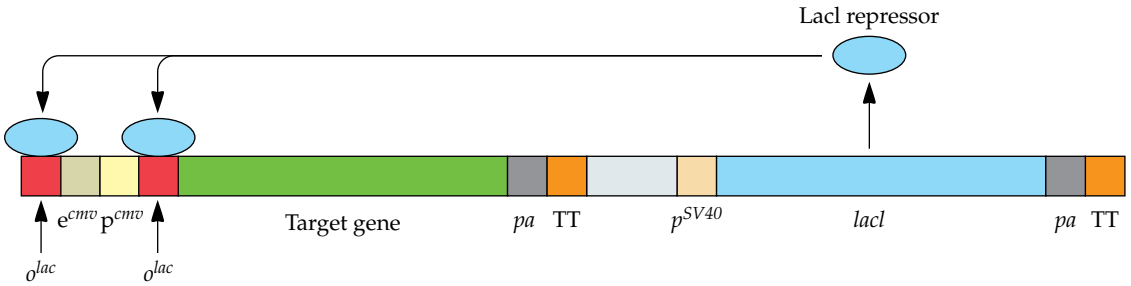
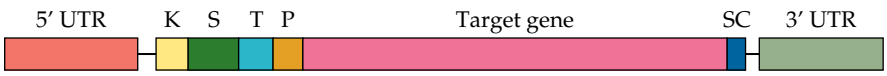


Figure 3.44 Regulation of heterologous gene expression in mammalian cells. The target gene is cloned into a vector under the control of the strong CMV promoter (p^{CMV}) and enhancer (e^{CMV}) with polyadenylation (*pa*) and transcription termination (TT) sequences. A repressor gene, such as *lacI* encoding the LacI repressor from the *E. coli lac* operon, is also present on the vector and is expressed from the SV40 promoter (p^{SV40}). The LacI repressor binds to *lac* operator (o^{lac}) sequences that were inserted up- and downstream of p^{CMV} and prevents transcription of the target gene. In the presence of IPTG and/or lactose, LacI does not bind to o^{lac}, and the recombinant protein is expressed.

Translation signals. For the best results, a target gene must be equipped with translation control sequences (Fig. 3.45). Initiation of translation in higher eukaryotic organisms depends on a specific sequence of nucleotides surrounding the start (AUG) codon in the mRNA called the Kozak sequence, e.g., GCCGCC(A/G)CC**AUG**G in vertebrates. The corresponding DNA sequence for the Kozak sequence is placed at the 5′ end of the target gene, often followed by a signal sequence to facilitate secretion, a protein sequence (tag) to enhance the purification of the recombinant protein, and a proteolytic cleavage sequence that enables the tag to be removed from the recombinant protein. A stop codon is required for translation to cease at the correct location. Finally, the sequence content of the 5′ and 3′ untranslated regions (UTRs) is important for efficient translation and mRNA stability. Either synthetic 5′ and 3′ UTRs or those from the human β-globin gene are used in mammalian expression vectors. The codons encoded by the target gene may also require modification to suit the translational preferences of the host cell.

The majority of mammalian cell expression vectors carry a single target gene that encodes a functional polypeptide. However, the active form of some commercially important proteins consists of two different protein subunits. For example, human thyroid-stimulating hormone is a two-subunit protein (heterodimer), and both hemoglobin and antibodies are tetramers with two copies of each subunit, $\alpha_2\beta_2$ and H_2L_2, respectively. It is possible to clone the gene or cDNA for each subunit of a multimeric protein, synthesize and purify

Figure 3.45 Translation control elements. A target gene can be fitted with various sequences that enhance translation and facilitate both secretion and purification, such as a Kozak sequence (K), signal sequence (S), protein affinity tag (T), proteolytic cleavage site (P), and stop codon (SC). The 5′ and 3′ UTRs increase the efficiency of translation and contribute to mRNA stability.



each subunit separately, and then mix the subunits together in a test tube. Unfortunately, relatively few multisubunit proteins are properly assembled *in vitro*. In contrast, *in vivo* assembly of dimeric and tetrameric proteins is generally quite efficient. Consequently, various strategies have been devised for the production of two different recombinant proteins within the same cell.

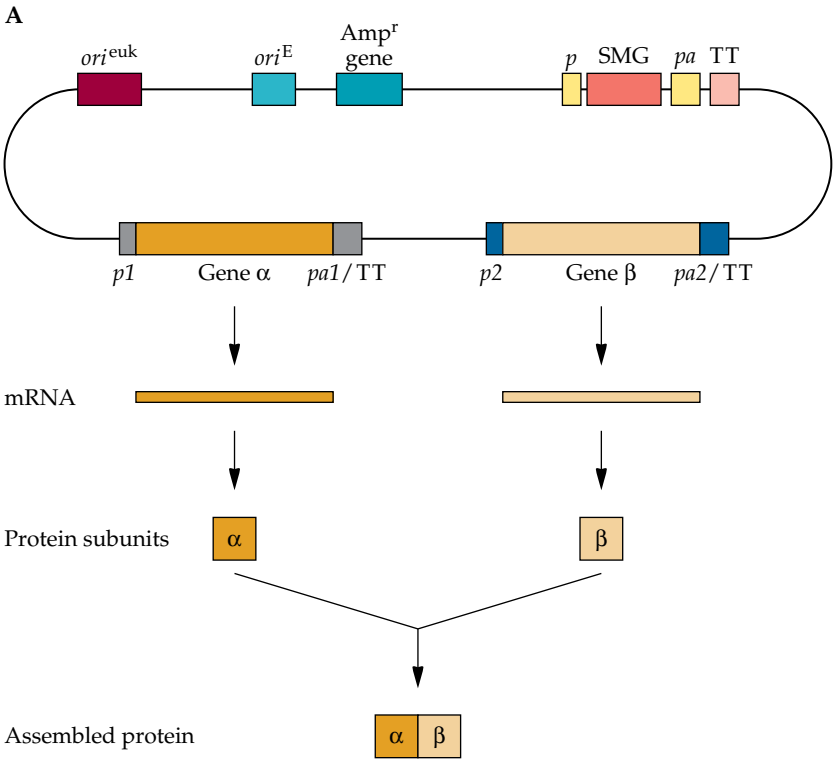
To produce hetero-dimeric or -tetrameric proteins, single vectors that carry two cloned genes have been developed. The two genes are placed under the control of independent promoters and polyadenylation signals (double-cassette vectors) (Fig. 3.46A). Alternatively, to ensure that equal amounts of the recombinant proteins are synthesized, "bicistronic" vectors have been constructed with the two cloned genes separated from each other by a DNA sequence that contains an internal ribosomal entry site (IRES) (Fig. 3.46B). IRESs are found in mammalian virus genomes, and after transcription, they allow simultaneous translation of different proteins from a polycistronic mRNA molecule. Transcription of a "gene α–IRES–gene β" construct is controlled by one promoter and polyadenylation signal. Under these conditions, a single "two-gene" (bicistronic) transcript is synthesized, and translation proceeds from the 5′ end of the mRNA to produce one of the subunits (subunit α) and internally from the IRES element to produce the second subunit (subunit β) (Fig. 3.46B).

Selectable Markers for Mammalian Expression Vectors

For the most part, the systems that are used to select transfected mammalian cells are the same as those for other eukaryotic host cells. A number of bacterial marker genes have been adapted for eukaryotic cells (Table 3.16). For example, the bacterial *neo* gene, which encodes neomycin phosphotransferase, is often used to select transfected mammalian cells. In eukaryotic cells, G-418 (Geneticin), which is phosphorylated and thereby inactivated by neomycin phosphotransferase, replaces neomycin as the selective agent because neomycin is not an effective inhibitor of eukaryotic protein synthesis.

Some selection schemes are designed not only to identify transfected cells, but also to increase heterologous-protein production by amplifying the copy number of the expression vector. The dihydrofolate reductase–methotrexate system falls into this category. Dihydrofolate reductase catalyzes the reduction of dihydrofolate to tetrahydrofolate, which is required for the production of purines. Dihydrofolate reductase-deficient cells transfected with vectors carrying the dihydrofolate reductase gene can grow on media lacking the nucleotide precursors hypoxantine and

Figure 3.46 Expression vectors for production of hetero-dimeric or -tetrameric proteins in mammalian cells. **(A)** Two-gene expression vector. The target genes (gene α and gene β) encode subunits of a protein dimer (αβ). The target genes are inserted into a vector and are under the control of different eukaryotic promoter (*p*), polyadenylation (*pa*), and transcription termination (TT) sequences. Each subunit is translated from a separate mRNA, and a functional protein dimer (αβ) is assembled. **(B)** Bicistronic expression vector. Each target gene (gene α and gene β) is inserted into a vector on either side of a sequence for an IRES. The two genes and the IRES sequence form a transcription unit under the control of a single eukaryotic promoter. Translation of the mRNA occurs from the 5′ end and internally (right-angled arrows). Both subunits (α and β) are synthesized and assembled into a functional protein dimer (αβ). Both expression vectors carry origins of replication for *E. coli* (ori^{E}) and mammalian cells (ori^{euk}), a selectable marker (Amp^{r}) for selecting transformed *E. coli*, and a selectable marker gene (SMG) that is under the control of eukaryotic promoter (*p*), polyadenylation (*pa*), and transcription termination (TT) sequences.


A
ori^euk
ori^E
Amp^r gene
p
SMG
pa
TT
p1
Gene α
pa1/TT
p2
Gene β
pa2/TT
mRNA
Protein subunits
α
β
Assembled protein
α
β

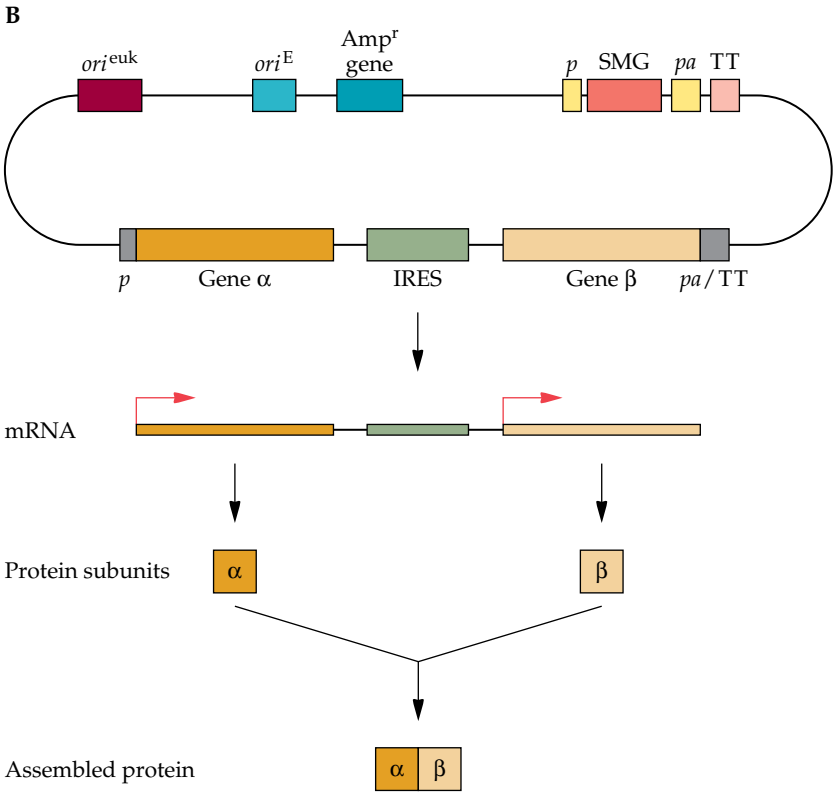

B
ori^euk
ori^E
Amp^r gene
p
SMG
pa
TT
p
Gene α
IRES
Gene β
pa/TT
mRNA
Protein subunits
α
β
Assembled protein
α
β

Table 3.16 Selective marker gene systems for mammalian cells[a]

Selective agent	Action of selective agent	Marker gene	Action of marker gene protein
Xyl-A	Damages DNA	Adenine deaminase (*ada*)	Deaminates Xyl-A
Blasticidin S	Inhibits protein synthesis	Blasticidin S deaminases (*Bsr*, *BSD*)	Deaminates blasticidin S
Bleomycin	Breaks DNA strands	Bleomycin-binding protein (*Ble*)	Binds to bleomycin
G-418 (Geneticin)	Inhibits protein synthesis	Neomycin phosphotransferase (*neo*)	Phosphorylates G-418
Histidinol	Produces cytotoxic effects	Histidinol dehydrogenase (*hisD*)	Oxidizes histidinol to histidine
Hygromycin B	Inhibits protein synthesis	Hygromycin B phosphotransferase (*Hph*)	Phosphorylates hygromycin B
MSX	Inhibits glutamine synthesis	Glutamine synthetase (*GS*)	Cells that produce excess glutamine synthetase survive
MTX	Inhibits DNA synthesis	Dihydrofolate reductase (*dhfr*)	Cells that produce excess dihydrofolate reductase survive
PALA	Inhibits purine synthesis	Cytosine deaminase (*codA*)	Lowers cytosine levels in the medium by converting cytosine to uracil
Puromycin	Inhibits protein synthesis	Puromycin *N*-acetyltransferase (*Pac*)	Acetylates puromycin

[a]MSX, methionine sulfoximine; MTX, methotrexate; PALA, *N*-(phosphoacetyl)-L-aspartate; Xyl-A, 9-β-D-xylofuranosyl adenine.

thymidine, whereas untransfected cells cannot. The enzyme is inhibited by methotrexate. Sensitivity to methotrexate can be overcome if the cell produces excess dihydrofolate reductase, and as the methotrexate concentration is increased over time, the dihydrofolate reductase gene in cultured cells is amplified. It is not unusual for methotrexate-resistant cells to contain hundreds of dihydrofolate reductase genes. The standard dihydrofolate reductase–methotrexate protocol entails transfecting dihydrofolate reductase-deficient cells with an expression vector carrying a dihydrofolate reductase gene as the selectable marker gene and treating the cells with methotrexate. After the initial selection of transfected cells, the concentration of methotrexate is gradually increased, and eventually cells with very high copy numbers of the expression vector are selected.

Engineering Mammalian Cell Hosts for Enhanced Productivity

Increasing host cell viability. In large-scale bioreactors, depleted nutrients and accumulation of toxic cell waste can limit the viability and density of cells as they respond to stress by inducing cell death, also known as apoptosis. One method to improve cell growth and viability under culture conditions in bioreactors is to prevent the tumor suppressor protein p53, which is a transcription factor, from activating the cell death response pathway. The mouse double mutant 2 protein (MDM2) binds to protein p53 and prevents it from acting as a transcription factor (Fig. 3.47). MDM2 also marks p53 for degradation. CHO cells were transfected with

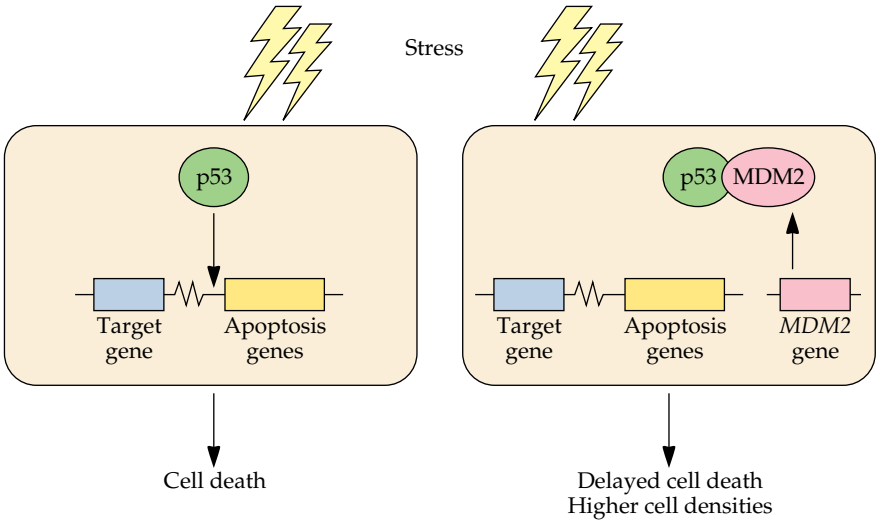


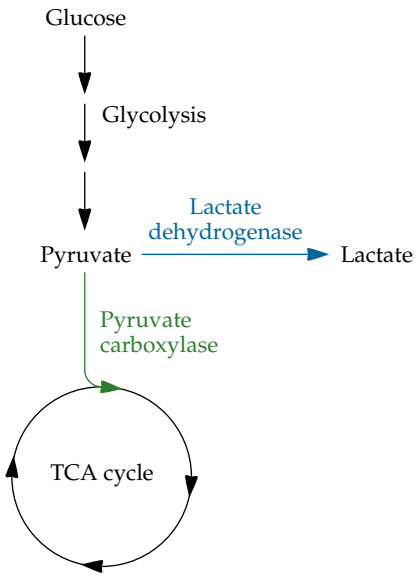
Figure 3.47 Strategy to increase yields of recombinant mammalian cells. Cell death (apoptosis), stimulated by the transcription factor p53, can lead to decreased yields of recombinant mammalian cells grown under stressful conditions in large bioreactors. To prevent cell death, the gene encoding MDM2 is introduced into mammalian cells. The MDM2 protein binds to p53 and prevents it from inducing expression of proteins required for apoptosis. Engineered cells not only showed delayed cell death, but also achieved higher cell densities in bioreactors.

plasmids containing a regulatable *MDM2* gene and cultured under conditions that mimicked the late stages of cell culture and in nutrient-limited medium. Cultures expressing *MDM2* had higher cell densities and delayed cell death compared to nontransfected cells, especially in nutrient-deprived medium.

Many cultured mammalian cells are unable to achieve high cell densities in cultures because toxic metabolic products accumulate in the culture medium and inhibit cell growth. Although efforts are made to optimize the culture conditions, inevitably, nutrients essential for optimal cell growth, including oxygen, are reduced. Under low-oxygen conditions, many cells, including CHO cells, secrete the acidic waste product lactate as they struggle to obtain energy from glucose. Under these conditions, pyruvate, an intermediate compound produced during the metabolism of glucose, is converted to lactate by lactate dehydrogenase rather than entering into the tricarboxylic acid cycle, where it is further oxidized through the activity of pyruvate carboxylase (Fig. 3.48). To counteract the acidification of the medium from lactate secretion, the human pyruvate carboxylase gene was cloned into an expression vector under the control of the CMV promoter and the SV40 polyadenylation signals and was transfected into CHO cells. When the pyruvate carboxylase gene was stably integrated into the CHO genome and expressed, the enzyme was detected in the mitochondria, where glucose is degraded. During the exponential growth phase, the rate of lactate production was reduced by 27% in the engineered cells, and the overall yield of lactate decreased by up to 15%.

Many of the eukaryotic DNA viruses from which the vectors used in mammalian cells are derived maintain their genomes as multicopy

Figure 3.48 Production of lactate under low-oxygen conditions. When oxygen is present, pyruvate, which is formed from glucose during glycolysis, is converted by the enzyme pyruvate carboxylase to an intermediate compound in the tricarboxylic acid (TCA) cycle. This metabolic pathway is important for the generation of cellular energy and for the synthesis of biomolecules required for cell proliferation. However, under low-oxygen conditions, such as those found in large bioreactors, pyruvate carboxylase has a low level of activity. Under these conditions, lactate dehydrogenase converts pyruvate into lactate, which yields a lower level of energy. Cultured cells secrete lactate, thereby acidifying the medium.



episomes in the host cell nucleus. These viruses produce proteins, such as the large-T antigen in SV40 and the nuclear antigen 1 protein in Epstein-Barr virus, that help to maintain the episomes in the host nucleus and to ensure that each host cell produced after cell division receives a copy of the episome. To increase the copy number of the target gene by increasing the episome copy number, cell lines have been engineered to express the SV40 large-T antigen or Epstein-Barr nuclear antigen 1.

Increasing protein secretion. Many proteins of therapeutic value, such as antibodies and interferon, are secreted. However, the high levels of these proteins that are desirable from a commercial standpoint can quickly overwhelm the capacity of the cell secretory system. Thus, protein processing is a major limiting step in the achievement of high target protein yields. Although high levels of recombinant protein production have been found to increase the levels of proteins associated with proper protein folding and secretion in the endoplasmic reticulum, the levels are usually not sufficient for optimal protein processing. Researchers have therefore devised methods to increase the capacity for protein secretion by engineering cell lines with enhanced production of components of the secretion apparatus. In this regard, an effective strategy may be to simultaneously overexpress several, if not all, of the proteins that make up the secretory mechanism. This can be achieved through the enhanced production of the transcription factor X box protein 1 (Xbp-1), a key regulator of the secretory pathway. Normally, full-length, unspliced *xbp-1* mRNA is found in nonstressed cells and is not translated into a stable, functional protein (Fig. 3.49A). However, when unfolded or misfolded proteins accumulate in the endoplasmic reticulum, a ribonuclease is activated that specifically cleaves *xbp-1* mRNA (Fig. 3.49B). This results in the production of a functional transcription factor that activates the expression of a number of proteins of the secretion apparatus. A truncated *xbp-1* gene that encodes an actively translated form of *xbp-1* mRNA (Fig. 3.49C) was overexpressed under the control of the CMV promoter in CHO cell lines that were previously constructed to express the recombinant proteins human erythropoietin, human γ-interferon, and human monoclonal antibodies either stably or transiently. Expression of the genes encoding proteins of the secretion apparatus that are controlled by Xbp-1 was found to increase in response to elevated levels of Xbp-1. Although overexpression of Xbp-1 did not increase the production of recombinant proteins in stable cell lines in which the heterologous gene is inserted in a chromosome, a significant increase was observed in cell lines engineered to express the recombinant proteins transiently from a plasmid-encoded gene.

Chromosomal Integration and Environment

A major consideration for high levels and long-term stability of heterologous-protein production is the site of integration of the target gene into the mammalian cell genome. Expression of high levels of protein from plasmid vectors is transient and inevitably results in loss of the vector, which cannot be propagated in mammalian cells, or death of the host cell. To overcome this problem, stable cell lines have been generated in which the heterologous gene is integrated into the chromosome, often facilitated by the CRISPR-Cas system. However, the site of integration can have a significant impact on the levels of target protein produced. Genomic

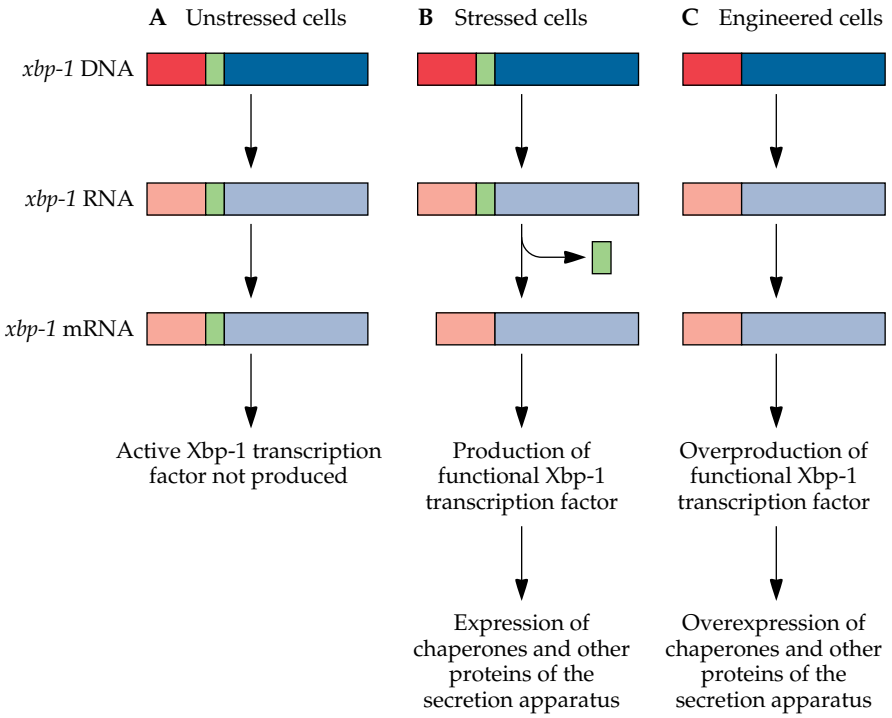


Figure 3.49 Strategy to increase yields of secreted recombinant proteins from mammalian cells by simultaneously upregulating the expression of several proteins in the secretion apparatus. The expression of chaperones and other proteins of the secretion apparatus is controlled by the transcription factor Xbp-1. **(A)** In unstressed cells, the intron (green box) is not cleaved from the *xbp-1* transcript, and therefore, functional Xbp-1 transcription factor is not produced. **(B)** However, in stressed cells that have accumulated misfolded proteins, an endoribonuclease cleaves the transcript to yield mature *xbp-1* mRNA (the red and blue boxes represent exons) that is translated into a stable, functional transcription factor. **(C)** Recombinant CHO cells were transfected with a truncated gene including only the *xbp-1* exons and overproduced a functional Xbp-1 transcription factor that directed the production of high levels of proteins required for protein secretion.

DNA is associated with a great number of proteins, including the major histone proteins, around which the DNA is coiled, that compact (condense) the DNA in the nucleus. The DNA and associated packaging proteins are known as chromatin. While much of the genome is highly condensed (heterochromatin) and contains silent genes or genes with low levels of expression, other regions are less condensed (euchromatin) and contain actively transcribed genes; the degree of condensation is influenced by epigenetic changes such as DNA methylation and covalent addition of chemical groups to histones. For enhanced expression and stability, the target gene should be integrated into euchromatin, rather than heterochromatin. Because a larger portion of the genome is in the heterochromatin form, there is a greater chance that the target gene will be inserted into one of these regions.

Techniques to relax the chromatin structure and thereby increase the expression of introduced genes include modifying host strains to express proteins that alter the chromatin structure at the site of vector integration or inserting DNA elements that prevent chromosome condensation together

with the target gene. One approach to alter the epigenetic environment surrounding the inserted gene is to increase histone acetylation. The extent of histone acetylation is determined by the relative activities of two host cell enzymes, histone acetyltransferase, which adds acetyl groups to lysines on histone proteins, and histone deacetylase, which removes acetyl groups from the histone. The relative influences of these two enzymes at a given promoter are determined by specific transcription factors that recruit one or the other of the enzymes to the promoter. Increased histone acetylation, which leads to increased gene transcription, can be accomplished either by increasing the expression of histone acetyltransferase or by decreasing the activity of histone deacetylase. One effective strategy to do this is to target histone acetyltransferase specifically to the site of target gene insertion to ensure that the target gene is actively and continuously transcribed. One group of researchers created a stable CHO cell line in which histone acetyltransferase was produced as a fusion protein with the bacterial LexA protein that binds to specific DNA sequences (Fig. 3.50). To test this fusion protein, the green fluorescent protein (GFP) reporter gene was employed as a target gene and was integrated into a CHO chromosome under the control of the CMV promoter with the LexA-binding sequence inserted upstream. A gene encoding resistance to the antibiotic Zeocin was coupled to the reporter gene by an IRES element and therefore was also under the control of the CMV promoter. Stable cells with an active CMV promoter were established by addition of Zeocin (which kills nonresistant cells by intercalating into DNA) to the culture medium. Production of GFP, determined by measuring the emission of green fluorescence, was severalfold higher in cells that expressed the LexA–histone acetyltransferase fusion protein than in those that expressed the LexA protein alone (Fig. 3.50A). The LexA protein specifically binds to the LexA recognition site upstream of the gene encoding GFP and brings with it the fused histone acetyltransferase protein that acetylates histones associated with the promoter region and promotes a higher level of GFP transcription. Moreover, expression remained stable, although at a lower level, for at least 4 months in some of the clones.

To improve expression levels over a longer period, the construct was further modified to include a DNA segment known as a stabilizing and antirepressor (STAR) element on both sides of the expression cassette to block repression (Fig. 3.50B). Repression can occur when heterochromatin forms due to the association of the heterochromatin protein HP1 with methylated histones. This stimulates further histone deacetylation and methylation and, consequently, greater HP1 activity. Insertion of the relatively small (<2-kb) STAR elements was found to counteract the activity of HP1 and other heterochromatin-associated repressor proteins. Flanking the expression cassette with the antirepressor elements resulted in higher levels of GFP expression that were maintained over a longer period of time.

Other DNA elements that improve heterologous-protein expression by modifying heterochromatin structure are the ubiquitous chromatin-opening elements and scaffold/matrix attachment regions. Ubiquitous chromatin-opening elements are sequences of DNA normally found near the promoters of housekeeping genes that are constitutively expressed at high levels due to enhanced histone acetylation. Inclusion of the ubiquitous chromatin-opening element from the promoter of the highly expressed CHO elongation factor 1 alpha gene in an expression vector increased recombinant protein expression in CHO cells 6- to 35-fold. Scaffold/matrix attachment regions were also found to enhance the

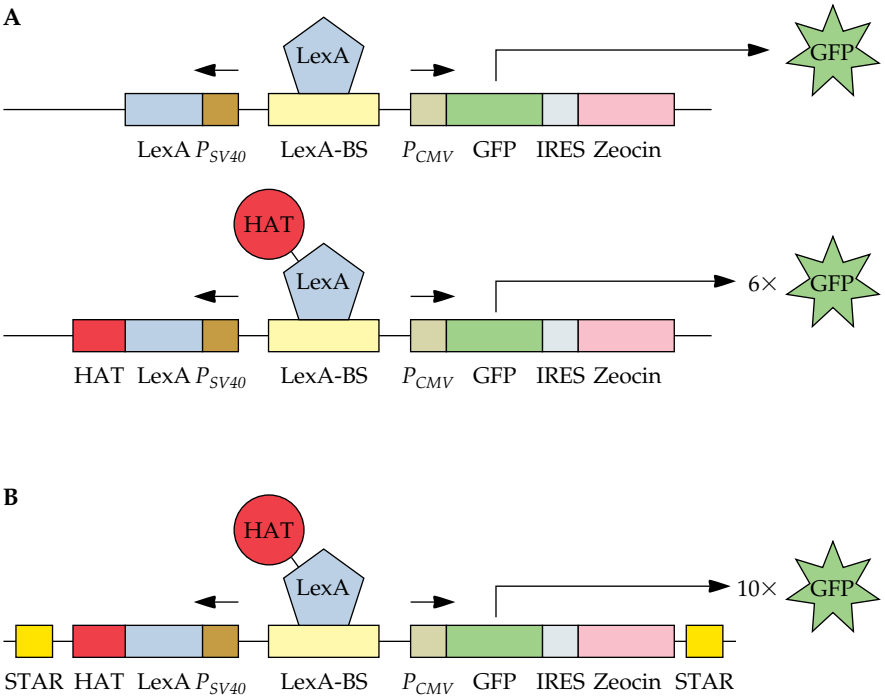


Figure 3.50 Strategies to increase expression of recombinant proteins in mammalian cells by altering chromatin structure. Local "relaxation" of chromosome condensation, which leads to increased transcription of genes in the region, can be achieved by the addition of an acetyl group to DNA-packaging proteins known as histones. Histone acetylation is catalyzed by the enzyme histone acetyltransferase (HAT). **(A)** To increase the expression of a recombinant protein, HAT was directed to the site of target gene (GFP gene) insertion in a mammalian chromosome. HAT was expressed as a fusion protein with the LexA protein that binds to a specific DNA sequence (LexA-BS) inserted upstream of the CMV promoter (P_{CMV}) that directs expression of GFP. Production of the HAT-LexA fusion protein under the control of the SV40 promoter (P_{SV40}) increased expression of GFP 6-fold compared to production of the LexA protein alone. **(B)** Insertion of STAR elements on both sides of the expression cassette further increased GFP expression. The gene encoding resistance to the antibiotic Zeocin was included as a selectable marker and was expressed from an IRES. The arrows above the promoter boxes indicate the direction of transcription.

production of heterologous protein in CHO cells. These elements, found in the chromosomes of many eukaryotes, bind to protein complexes in the nucleus that arrange regions of the chromosome into loops. It is thought that these DNA loops contain transcriptionally active genes that are regulated in a coordinated fashion. Although scaffold/matrix attachment regions from the human β-globin gene and the chicken lysozyme gene were found to increase expression of a target gene, not all scaffold/matrix attachment regions have a positive effect on gene expression.

In sum, mammalian cell expression systems are as versatile and effective as other eukaryotic expression systems. However, industrial production of a recombinant protein with engineered mammalian cells is costly, mainly due to high culture costs. Consequently, less expensive expression systems are favored unless authenticity of an important recombinant protein can be obtained only with mammalian cells.

Protein Engineering

The physical and chemical properties of natural proteins are sometimes not well suited to a medical, industrial, or other application. In some cases, a protein that is better suited to a particular task may be obtained by using a gene from an organism that grows in an unusual, often extreme, environment. In addition to isolating natural genes that encode proteins with useful properties, directed or random mutagenesis and selection schemes can be used to create a mutant form of a gene that encodes a protein with the desired properties. In directed mutagenesis, a specific amino acid is targeted for change by substituting nucleotides in the corresponding coding sequence in the gene. In random mutagenesis, the amino acid changes that will result in the desired properties are unknown. A large number of mutant proteins, each with a different amino acid change, are generated by randomly altering individual nucleotides within a gene and then tested for the desired properties.

By using directed or random mutagenesis, proteins with enhanced characteristics can be created for therapeutic or industrial applications. For example, the catalytic efficiency, allosteric regulation, cofactor requirement, or substrate specificity of an enzyme may be improved. The latter may decrease undesirable side effects of a therapeutic enzyme. The thermal tolerance and/or pH stability of a protein may be increased, enabling the mutant protein to be used under conditions that would inactivate the native version. The sensitivity of a protein to cellular proteases may be decreased, which would increase the recoverable yield of the protein.

Directed Mutagenesis

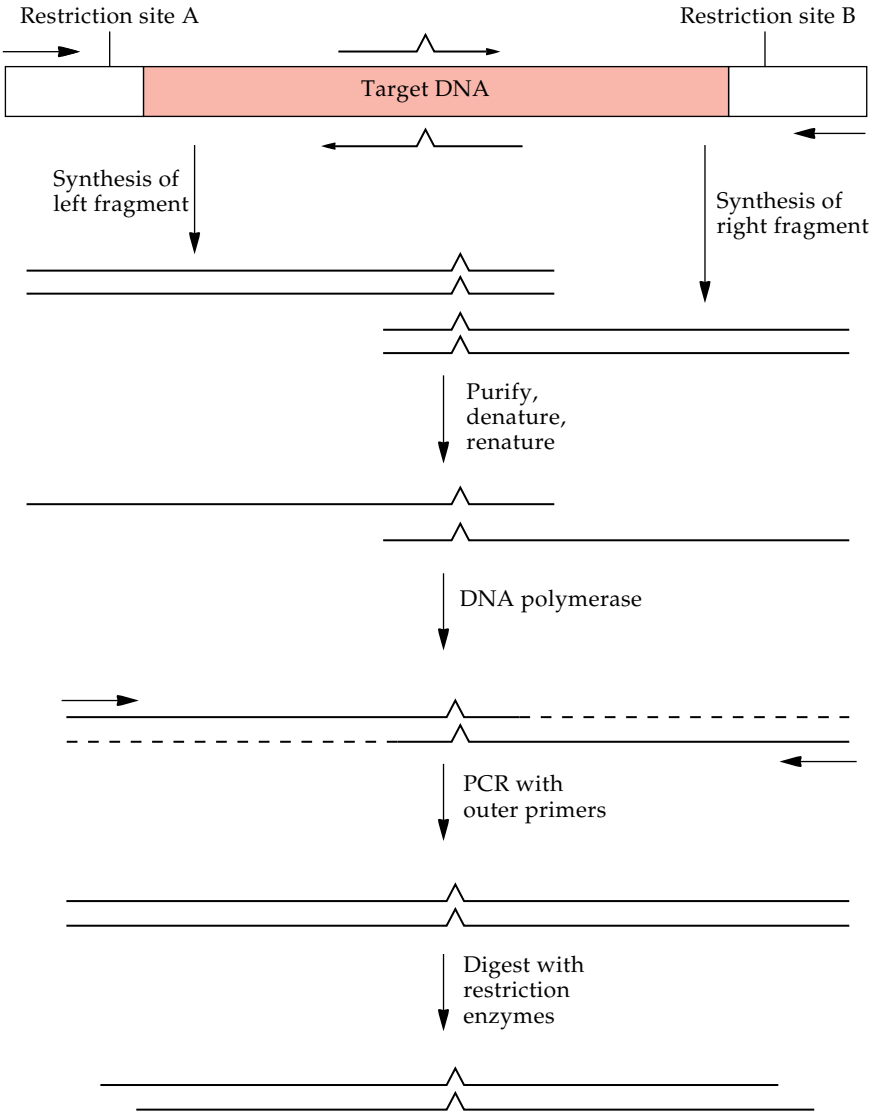
In some cases, it is possible to predict in advance which individual amino acids or short sequences of amino acids contribute to a particular physical, kinetic, or chemical property of a protein. The process for altering specific amino acids by changing the coding sequence at a targeted site in a gene is called directed mutagenesis. It is important to keep in mind that a particular property of a protein may be the consequence of two or more amino acids that are far apart from each other in the linear (primary) sequence but are juxtaposed as a result of the folding of the protein. In this case, two or more amino acids may have to be changed to produce a protein with the desired properties. Predicting which amino acids of a protein should be changed to attain a specific property generally requires that the three-dimensional structure of the protein, or a similar protein, has been well characterized by X-ray crystallographic analysis. However, for many proteins, such detailed information is often lacking, so directed mutagenesis becomes a trial-and-error strategy in which changes are made to those nucleotides that are most likely to yield a particular change in a protein property. Then, of course, the protein encoded by each mutated gene has to be tested to ascertain whether the mutagenesis process has indeed generated the desired change.

Site-Directed Mutagenesis by Overlap Extension PCR

A straightforward directed mutagenesis protocol introduces defined nucleotide substitutions into a gene by overlap extension PCR. Four oligonucleotide primers are used in the PCR; two flanking primers anneal to the ends of the target gene (often a cloned gene), and the other two overlapping,

internal primers carry the mutation. While the 3′ end of a primer must be perfectly complementary to the annealing site on the template DNA to prime DNA synthesis, mismatched nucleotides at the 5′ end do not affect the reaction. The target gene is initially amplified in two separate reactions to generate overlapping ("left" and "right") fragments (Fig. 3.51). In each of the

Figure 3.51 Site-directed mutagenesis by overlap extension PCR. The left and right portions of the target DNA are amplified separately by PCR. The primer pairs are shown by horizontal arrows. Oligonucleotides that carry the mutations are depicted as a line with a spike; a spike denotes a position that contains a nucleotide that is not found in the native gene. The amplified fragments are purified, denatured to make them single stranded, and then reannealed. Regions of overlap are formed between complementary mutation-producing oligonucleotides. The single-stranded regions are made double stranded with DNA polymerase, and then the entire fragment is amplified by PCR. The resultant product is digested with restriction endonucleases to facilitate cloning into a vector that has been digested with the same enzymes.

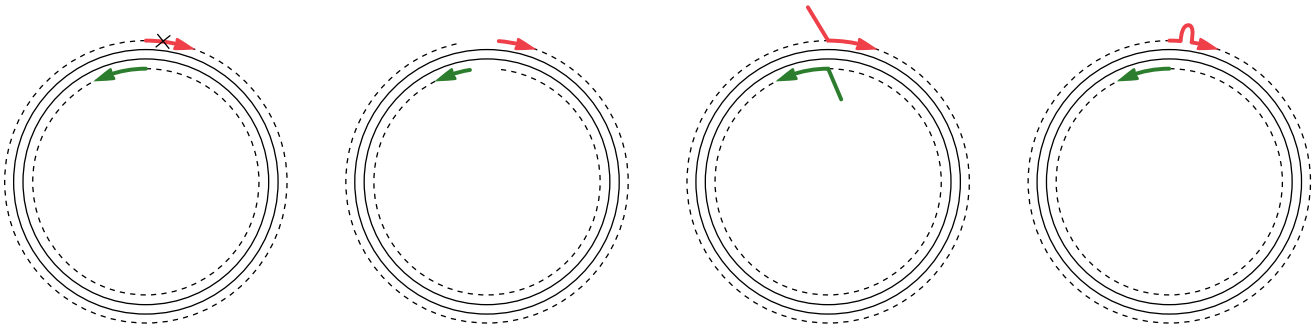


two reactions, one of the PCR primers (an internal primer) carries the mutation and the other is a flanking primer. After PCR amplification, the products are purified and the left and right fragments are combined. Denaturation and reannealing of the mixed fragments produce some DNA molecules that hybridize in the overlapping, complementary, mutated sequences. DNA polymerase is added to extend the strands to form double-stranded DNA molecules. These molecules are amplified by PCR with the flanking primers to enrich for full-length DNA molecules. The amplified DNA is then cloned into a suitable plasmid vector; this is facilitated by inclusion of suitable restriction enzyme sites in the 5′ ends of the flanking primers. This procedure results in the production of an altered gene that has mutated sites in the region of the overlap of the internal oligonucleotides.

Site-Directed Mutagenesis by Inverse PCR

Nucleotide substitutions, deletions, or insertions can be introduced into a target gene that has been cloned into a plasmid in a procedure known as inverse PCR. In this case, the entire plasmid is amplified, which restricts the size of the plasmid to less than about 10 kb. The oligonucleotide primers used in the inverse PCR anneal to adjacent sequences in the target gene but are divergently oriented, that is, their 3′ ends are directed away from each other. For point mutations, nucleotide changes are introduced in the middle of one of the primers (Fig. 3.52). To create deletion mutations, primers must flank the region of target DNA to be deleted and be perfectly matched to their annealing sites (Fig. 3.52). To create mutations with long insertions, a stretch of mismatched nucleotides is added to the 5′ end of one or both primers, while for mutations with short insertions, a stretch of nucleotides is designed in the middle of one of the primers (Fig. 3.52). PCR amplification yields linear double-stranded DNA products that are circularized by ligation with T4 DNA ligase. Ligation requires that the 5′ ends of the linear DNA molecules are phosphorylated, and therefore either the primers must be phosphorylated or the PCR products must be phosphorylated using the enzyme polynucleotide kinase. Finally, the recircularized plasmid DNA is transformed into *E. coli* by any standard procedure. Since this protocol yields a very high

Figure 3.52 Overview of the basic methodology to introduce point mutations, insertions, or deletions into DNA cloned into a plasmid. The forward and reverse primers are shown in red and green, respectively. The solid circles represent template DNA. The dotted lines represent newly synthesized DNA. The X indicates an altered nucleotide(s).



frequency of plasmids with the desired mutation, screening three or four clones by sequencing the target DNA is usually sufficient to find one with the desired mutation. Given its simplicity and effectiveness, this procedure has come to be widely used to introduce a specified point mutation, insertion, or deletion into a cloned gene.

Mutant Proteins with Unusual Amino Acids

Essentially any protein can be altered by substituting one amino acid for another using directed mutagenesis. However, this approach is limited to the 20 amino acids that are normally used in protein synthesis. One way to increase the diversity of the proteins formed after mutagenesis is to introduce synthetic, nonstandard amino acids with unique side chains at specific sites. To do this, *E. coli* was engineered to produce a novel transfer RNA (tRNA) that can be covalently bound to (charged with) a nonstandard amino acid and participates in ribosomal protein synthesis to incorporate the nonstandard amino acid at a specified position in a recombinant protein. The same host cell was also engineered to produce a new aminoacyl-tRNA synthetase that aminoacylates (charges) only the novel tRNA with the nonstandard amino acid, which is provided in the culture medium; the novel tRNA is not recognized by any of the existing *E. coli* aminoacyl-tRNA synthetases. A tyrosine-tRNA and tyrosine-tRNA synthetase pair from the archaebacterium *Methanococcus jannaschii* was used as a starting point for this system. The amino acid specificity of the tyrosine-tRNA synthetase from *M. jannaschii* was altered by random mutagenesis of its gene so that, instead of tyrosine, it catalyzes the addition of *O*-methyl-L-tyrosine onto a mutant form of the tyrosine-tRNA. The mutant tRNA contains an anticodon that is complementary to the stop codon UAG. It can therefore insert the nonstandard amino acid into a protein in places where the mRNA contains the codon UAG, which normally directs the cessation of protein synthesis. That is, UAG, the least-used stop codon in *E. coli*, was reassigned to encode a modified amino acid. A target gene was modified by site-directed mutagenesis to contain a 5′-TAG-3′ in that portion of the DNA that encodes the amino acid that was targeted for change to *O*-methyl-L-tyrosine (Fig. 3.53). The modified gene was expressed in the *E. coli* strain that produces the *O*-methyl-L-tyrosine-tRNA, resulting in a full-length target protein containing the modified amino acid. Had the mutant gene been expressed in wild-type *E. coli*, a truncated version of the protein would have been produced. To prevent the read-through of mRNAs of other cellular proteins that normally contain a UAG stop codon, insertion of amino acids at these sites is always less than 100% and is often dependent upon the nucleotides surrounding the stop codon.

Several novel tRNA/aminoacyl-tRNA synthetase pairs have been developed to insert more than 200 nonstandard amino acids with diverse physical, chemical, and biological properties into specified sites in proteins to produce functional proteins with altered activities compared with the native form. In addition to reassigning stop codons, redundant codons that are rarely used by a host cell may also be exploited to incorporate several different nonstandard amino acids into a protein (Fig. 3.54). It is also possible to engineer tRNAs with anticodons that recognize extended four- or five-base codons that have been introduced into genes encoding recombinant proteins.

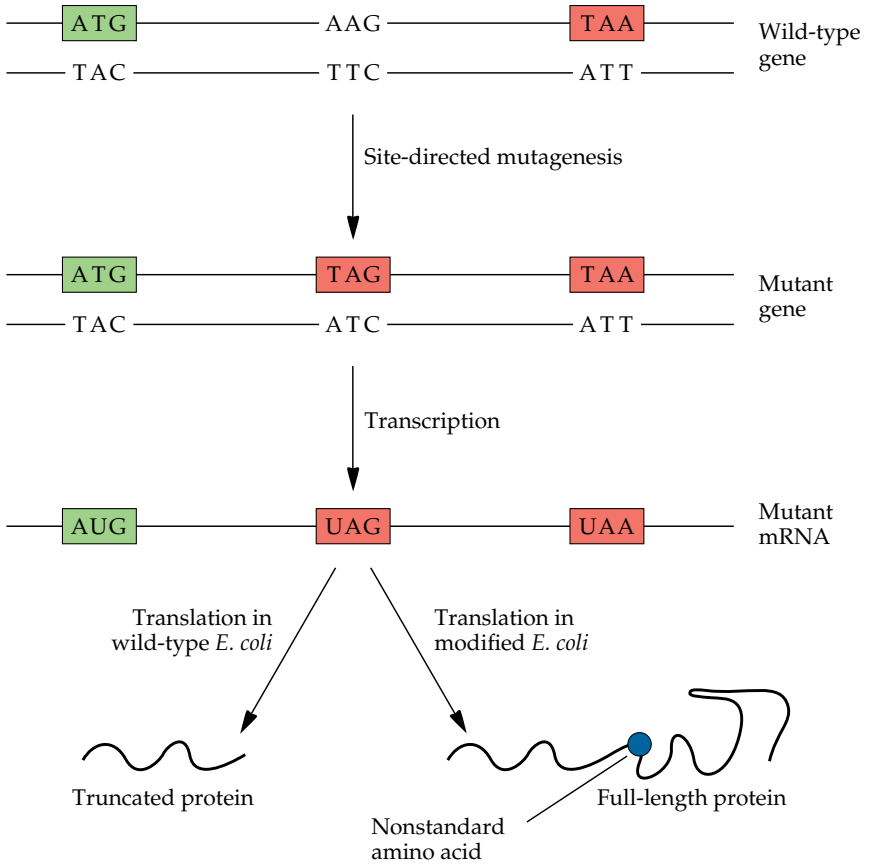
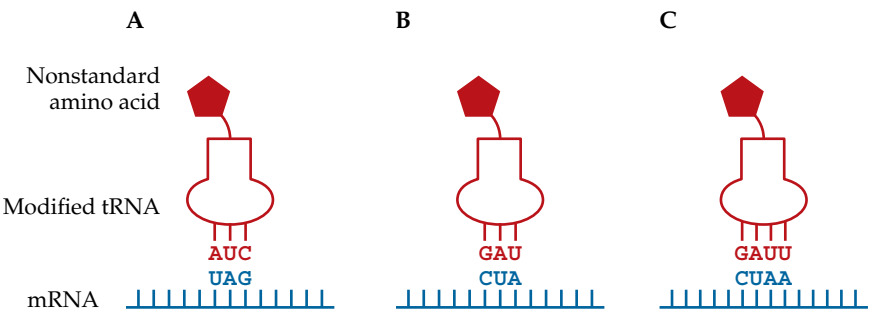


Figure 3.53 Production of a protein containing a nonstandard amino acid. The start codon is highlighted in green, and the stop codons are in red. The inserted nonstandard amino acid is shown in blue.

Figure 3.54 Codons in mRNA may be altered by site-directed mutagenesis of target genes to encode a nonstandard amino acid. Rarely used stop codons **(A)** or redundant codons **(B)** may be reassigned to encode a nonstandard amino acid that is covalently bound to a modified tRNA. Extended codons of four or more bases **(C)** may be recognized by modified tRNAs. Note that in the example shown here, recognition of the first three bases (CUA) of the four-base codon by a natural tRNA (leucine-tRNA) would result in a reading frame shift; however, this codon is rarely used by *E. coli*, and therefore proteins with the incorrect amino acids would be rare.



Random Mutagenesis

In those cases where the amino acid changes that will result in the desired properties are unknown, a library of mutated sequences is generated by randomly altering individual nucleotides within a structural gene. Most of the mutations will decrease the functioning of the encoded protein, and therefore an efficient screening process is required to identify proteins with the rare mutations that result in beneficial changes. This approach, often requiring multiple rounds of random mutagenesis and selection to acquire a protein with the desired activity, is sometimes referred to as directed evolution.

Error-Prone PCR

Some of the temperature-stable DNA polymerases that are used to amplify target DNA by PCR occasionally insert incorrect nucleotides during DNA replication. If one is attempting to amplify a DNA with high fidelity, this is obviously a problem. On the other hand, if the construction of a library of mutants of the target gene is the objective, then this approach is a useful method for random mutagenesis. Error-prone PCR is performed using DNA polymerases that lack proofreading activity, such as *Taq* DNA polymerase. The error rate may be increased by increasing the concentration of Mg^{2+} to stabilize noncomplementary base pairs. Addition of Mn^{2+} and/or unequal amounts of the four deoxynucleoside triphosphates to the reaction buffer may also increase the error rate. The primer annealing sites on the template DNA define the region to be altered, and the number of nucleotide substitutions per template increases with the number of PCR cycles and the length of the template. Following error-prone PCR, the randomly mutagenized DNA is cloned into an expression vector, introduced into a host cell, and screened for altered or improved protein activity. The DNA from those clones that encode the desired activity is isolated and sequenced to determine the relevant changes to the target DNA.

Random Mutagenesis with Degenerate Oligonucleotide Primers

In addition to introducing a specific nucleotide substitution into a gene, overlap extension PCR can be used to incorporate any of the four nucleotides at defined positions to generate all the possible amino acid changes in a particular region of a protein. This pattern of sequence degeneracy is achieved by programming an automated DNA synthesis reaction to add a low level (usually a few percent) of each of the three alternative nucleotides each time a particular nucleotide is added during the synthesis of an oligonucleotide primer (Fig. 3.55). In this way, the oligonucleotide primer preparation contains a heterogeneous set of DNA sequences that will generate a series of mutations that are clustered in a defined portion of the target gene. The degenerate oligonucleotides are employed as "internal" PCR primers to amplify the left and right portions of the target gene in separate reactions (Fig. 3.51). Mixing, denaturing, and annealing the left and right fragments produces some DNA molecules that overlap by complementarity and can be extended by DNA polymerase to produce a library of altered genes that have mutated sites in the region of the overlap of the degenerate oligonucleotides.

Mutagenesis using degenerate oligonucleotides confers two advantages over targeted mutagenesis. (i) Detailed information regarding the roles of

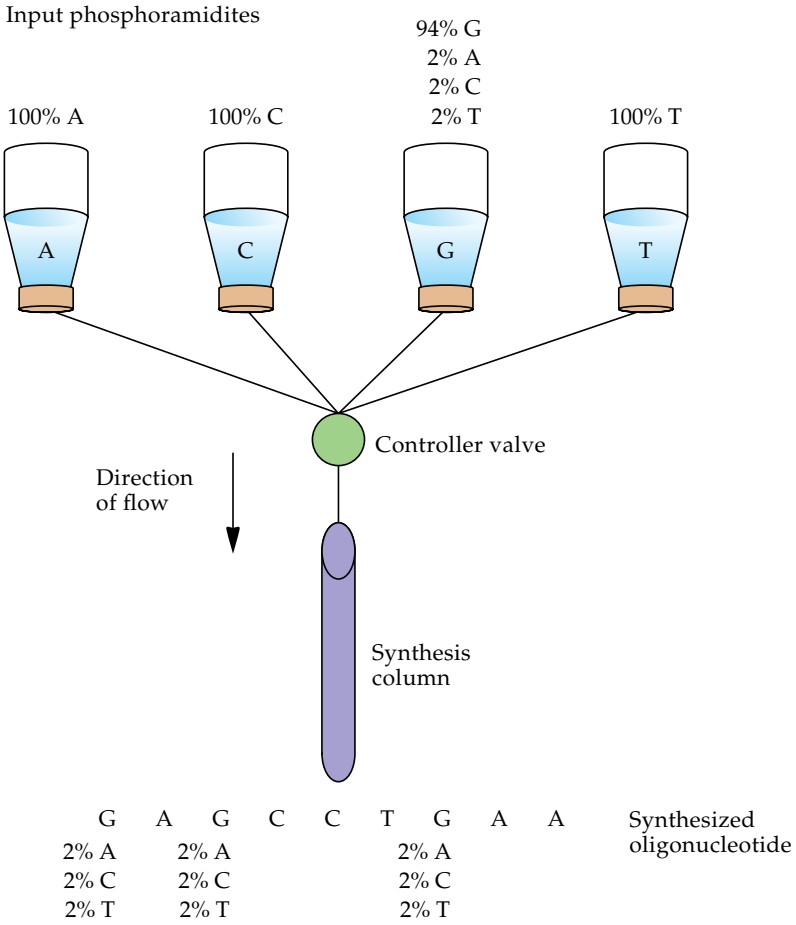


Figure 3.55 Chemical synthesis of oligonucleotide primers with any of the four nucleotides at defined positions. In this case, the flask with G phosphoramidite consists of a mixture of nucleotides, such as 94% G, 2% A, 2% C, and 2% T, leading to a mixture of oligonucleotides that may have A, C, or T at the sites where G is the specified nucleotide.

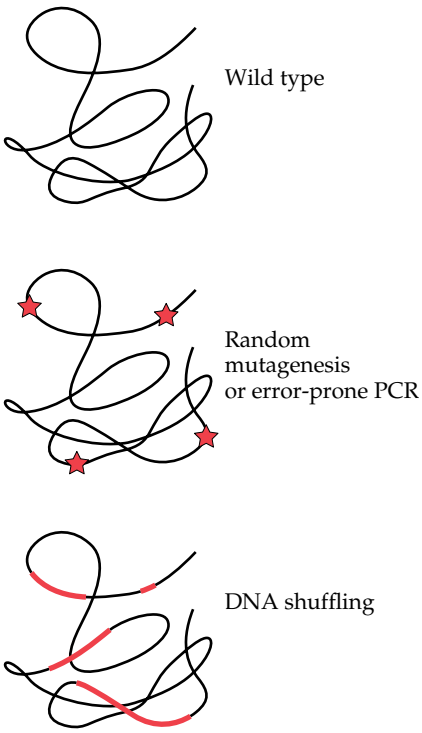


Figure 3.56 Amino acid changes may be introduced at random positions in a protein by either error-prone PCR or PCR using degenerate oligonucleotide primers, both of which cause single amino acid substitutions (red stars), and by DNA shuffling, in which genes are formed with large regions from different sources (red lines).

particular amino acids in the functioning of the protein is not required. (ii) Unexpected mutants encoding proteins with a range of interesting and useful properties may be generated because the introduced changes are not limited to one amino acid. Of course, should none of the mutants yield a protein with the properties that are being sought, then it may be necessary to repeat the entire procedure with a set of degenerate primers that is complementary to a different region of the gene.

DNA Shuffling

Some biologically important proteins, such as α-interferon (IFN-α), are encoded by a family of several related genes, with each protein having slightly different biological activity. If all, or at least several, of the genes or cDNAs for a particular protein have been isolated, it is possible to recombine portions of these genes or cDNAs to produce hybrid or chimeric forms (Fig. 3.56).

This "DNA shuffling" is done with the expectation that some of the hybrid proteins will have unique properties or activities that were not encoded in any of the original sequences. Also, some of the hybrid proteins may combine important attributes of two or more of the original proteins, e.g., high activity and thermostability.

The simplest way to shuffle portions of similar genes is through the use of common restriction enzyme sites (Fig. 3.57). Digestion of two or more of the DNAs that encode the native forms of similar proteins with one or more restriction enzymes that cut the DNAs in the same place, followed by ligation of the mixture of DNA fragments, can potentially generate a large number of hybrids. For example, two DNAs, each of which has three unique restriction enzyme sites, can be recombined (shuffled) to produce 14 different hybrids in addition to the original DNA (Fig. 3.57).

Another way to shuffle DNA involves combining several members of a gene family, fragmenting the mixed DNA at random sites with deoxyribonuclease I (DNase I), and then performing a PCR. During PCR, gene fragments from different members of a gene family cross-prime each other after DNA fragments bind to one another by complementary base pairing in regions of high homology (Fig. 3.58). The final products are amplified by PCR using terminal primers. After 20 to 30 PCR cycles, a large number of full-length, hybrid DNAs will be generated. The hybrid DNAs are then cloned to create a library that can be screened for the desired activity. Although DNA shuffling works well with gene families or with genes from different families that have a high degree of homology, the technique is not especially useful when proteins have little or no homology. Thus, the DNAs must be very similar to one another or the PCR will not proceed. To remedy this situation and combine the genes of dissimilar proteins, several variations of the DNA-shuffling protocol have been described.

One procedure that was developed to combine the genes of dissimilar proteins and that does not rely on PCR amplification of DNA fragments is called nonhomologous random recombination. In this procedure, DNAs from different sources (either defined or random DNA sequences, or a mixture of both) are combined and then partially digested with DNase I (Fig. 3.59). These DNA fragments, which include a wide variety of sizes, are made blunt ended by digestion with the enzyme T4 DNA polymerase. This enzyme both fills in 5′ overhanging nucleotides and degrades 3′ overhanging nucleotides. The DNA fragments are then mixed with a synthetic DNA molecule that forms a hairpin loop and contains a specific restriction enzyme site. The entire mixture is ligated by the addition of the enzyme T4 DNA ligase, which results in the formation of extended mosaic DNA molecules of variable lengths with a hairpin at each end. Ligation of the hairpins prevents further addition of fragments to the molecules. The average length of the hairpin-flanked molecules is dictated by the ratio between the blunt-ended DNA fragments and the hairpins added to the ligation reaction. Finally, restriction enzyme digestion removes the hairpin loops so that the resulting sticky-ended DNA fragments can be inserted into plasmid vectors and tested for various activities. Because this process randomly recombines DNA fragments, only a very small fraction of the recombined DNAs are likely to encode the desired activity.

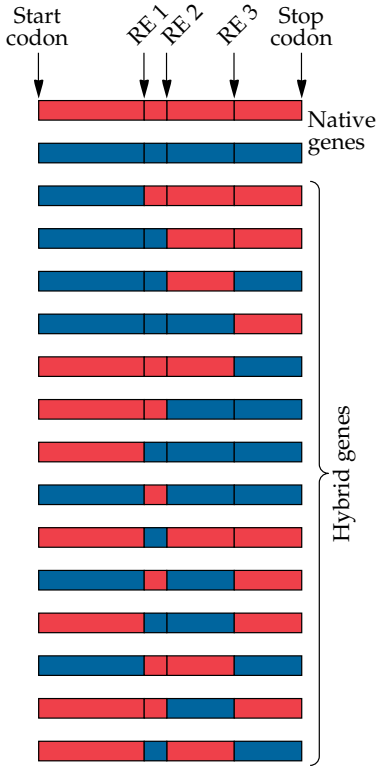
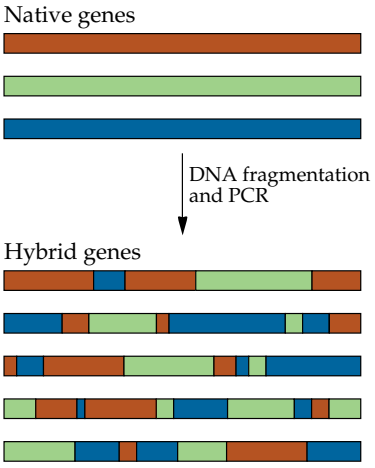


Figure 3.57 The 14 different hybrid genes that can be generated by combining restriction enzyme fragments from two genes from the same gene family that have three different restriction sites in common. RE, restriction enzyme.

Figure 3.58 Some of the hybrid DNAs that can be generated during PCR amplification of three members of a gene family.



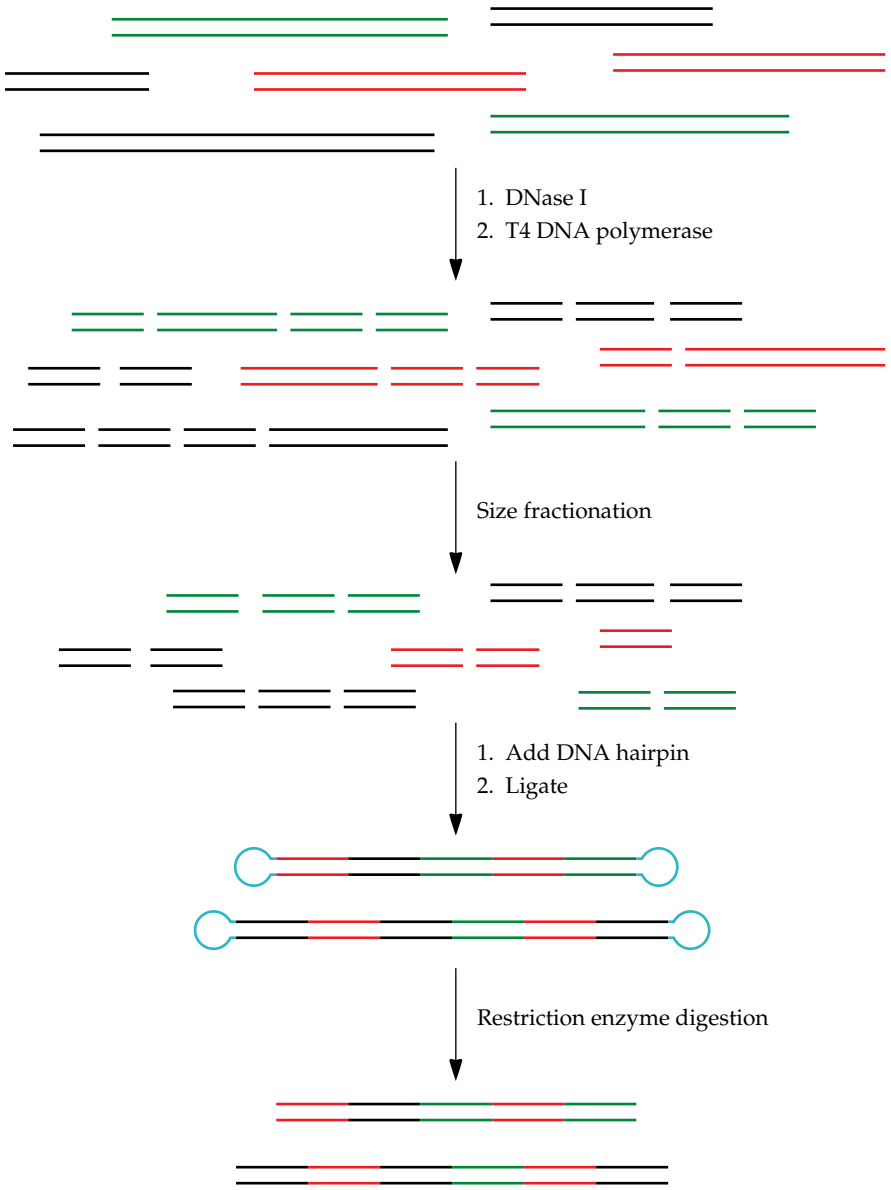


Figure 3.59 Nonhomologous random recombination. Different DNAs (shown in different colors) are mixed together, partially digested with DNase I, and blunted at the ends by digestion with T4 DNA polymerase. The fragments are size fractionated and then ligated with synthetic hairpin DNAs to form extended DNA molecules with a hairpin on both ends. After restriction enzyme digestion to remove the hairpin ends and generate sticky ends, the DNA molecules are ligated into plasmid vectors.

Examples of Protein Engineering

Increasing Protein Stability

Proteins have evolved to perform a particular function for a microorganism, animal, or plant under natural conditions and are often not well suited for highly specialized biotechnology applications. For example, most

enzymes are easily denatured by the high temperature and the presence of organic solvents that are used in some industrial processes. Although thermotolerant enzymes can be isolated from thermophilic microorganisms, these organisms often lack the particular enzyme that is required for an industrial process. However, directed mutagenesis can be used to create a protein that will not readily unfold under the conditions in which it will be employed. The addition of disulfide bonds, through the introduction of specifically placed cysteines, can usually significantly increase the stability of a protein (Fig. 3.9). Extra disulfide bonds may perturb the normal functioning of a protein, and therefore the activity, as well as the stability, of a modified protein must be tested.

In one example, a receptor antagonist protein was engineered for increased stability to enhance its effectiveness as a therapeutic agent. Low-density lipoprotein receptor-related protein 1 (LRP1) is a cell surface signaling protein that binds lipoproteins and other ligands and removes them from the bloodstream. The receptors also remove blood coagulation proteins, which leads to bleeding episodes in individuals with hemophilia. Following synthesis in the endoplasmic reticulum, LRP1 is escorted to the Golgi by a chaperone protein, receptor-associated protein (RAP). RAP is denatured in the acidic environment of the Golgi, thereby releasing the receptor proteins for subsequent processing and transport to the cell membrane. Exploiting its high affinity for LRP1, RAP can also be administered exogenously to inhibit binding of blood coagulation proteins to the receptors and thereby prevent bleeding episodes in hemophiliacs.

Following administration of RAP, the LRP1–RAP complex that forms at the cell membrane is taken up by the cell in an endosome, which results in low pH-induced denaturation of RAP and recycling of LRP1 back to the cell surface (Fig. 3.60A). Thus, the acid sensitivity of RAP limits its potential as a therapeutic agent. Researchers reasoned that introduction of a disulfide bond would increase the acid stability of RAP and thereby prevent its dissociation from LRP1. The structure of RAP is known, and computer modeling was used to predict optimal sites for introduction of cysteines. Using oligonucleotide-directed mutagenesis, the coding sequences for two amino acids in domain D3 of RAP, a tyrosine at position 260 and a threonine at position 297, were altered to encode cysteines (Fig. 3.60B). In addition, four histidines in domain D3 were changed to phenylalanine to prevent histidine protonation at low pH that causes RAP to unfold and dissociate from LRP1 (Fig. 3.60C). The two sets of mutations, that is, the introduction of the cysteines and the elimination of the histidines, were introduced into RAP separately and in combination. At pH 7.4, the wild-type and mutant RAP proteins were properly folded (measured as the portion of α-helical content of the protein). At pH 5.5, wild-type RAP unfolded, and its affinity for LRP1 was reduced 181-fold (Table 3.17). In contrast, at low pH, mutant RAP with the disulfide bond and mutant RAP with histidines replaced with phenylalanine were substantially folded. Moreover, RAP with the combined mutations remained properly folded. Importantly, the LRP1 binding affinity of the mutant RAP proteins at pH 5.5 was reduced only 22- to 53-fold compared to that at pH 7.4 (Table 3.17). Treatment of LRP1-expressing human fibroblast cells with combined mutant RAP significantly inhibited ligand (α_2-macroglobulin) uptake compared to treatment with wild-type

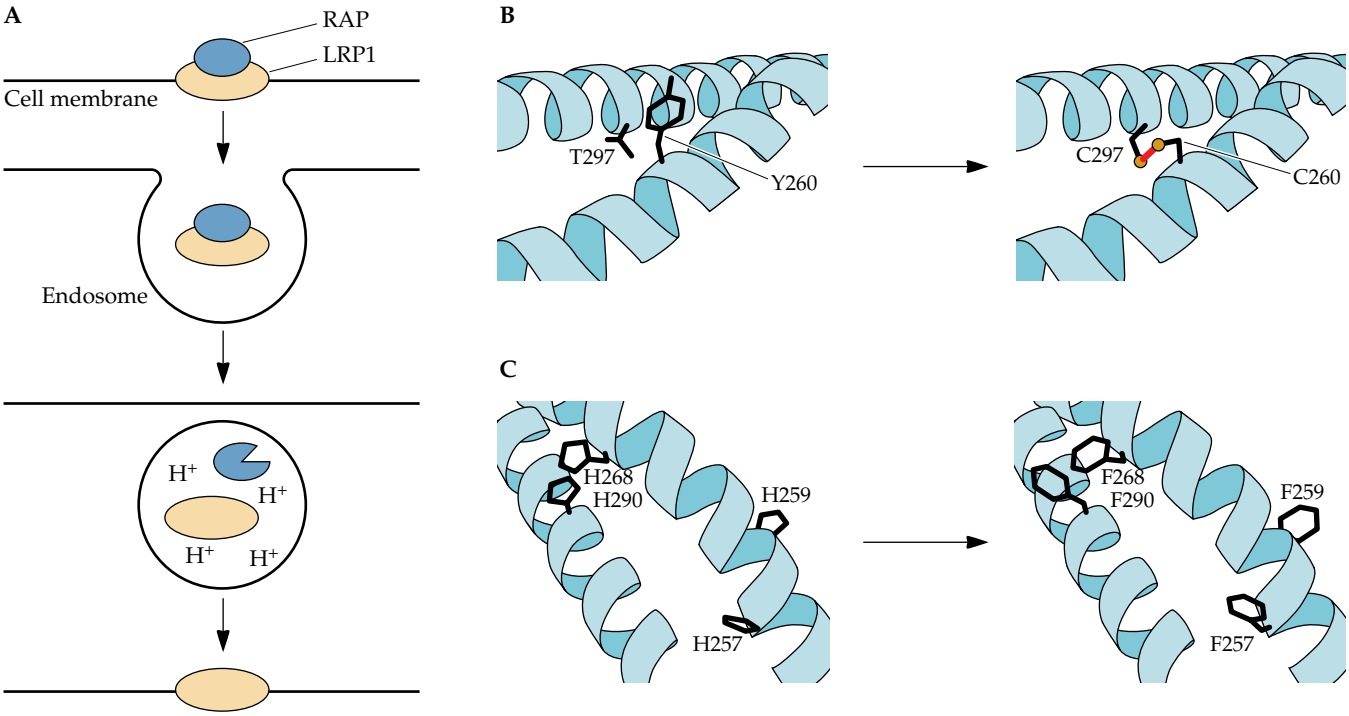


Figure 3.60 Directed mutagenesis of receptor-associated protein (RAP) to increase acid stability. **(A)** Following binding of exogenous RAP to low-density lipoprotein receptor-related protein 1 (LRP1) in the cell membrane, the protein complex is taken into the cell by endocytosis. Acid-sensitive wild-type RAP is denatured in the acidic endosome and releases LRP1, which is recycled back to the cell membrane. **(B)** To increase the stability of RAP, a disulfide bond was introduced into domain D3. Tyrosine at position 260 (Y260) and threonine at position 297 (T297) were changed to cysteines (C260 and C297) by site-directed mutagenesis. **(C)** To further increase acid stability, four histidines (H257, H259, H268, H290) that are protonated at low pH were changed to phenylalanine (F257, F259, F268, F290). Data from Prasad et al., *J. Biol. Chem*. **290:**17262, 2015.

RAP, indicating that recycling of the receptor was prevented to a greater extent by the acid-stable, mutant RAP. A similar result was found in mice. This study shows that engineering RAP for increased stability generates a more potent inhibitor of LRP1 than wild-type RAP and is a suitable treatment for diseases mediated by LRP1 clearance of blood proteins.

Table 3.17 Binding affinity of LRP1 for mutant RAP[a,b]

RAP	Mutations	KD at pH 7.4	KD at pH 5.5	Fold change in KD
Wild-type	None	0.68	123	181
RAP disulfide	Y260C, T297C	1.67	58.2	35
RAP quad	H:F H257F, H259F, H268F, H290F	1.38	72.7	53
RAP combined	Y260C, T297C, H257F, H259F, H268F, H290F	0.96	21.5	22

[a]Data from Prasad et al., *J. Biol. Chem*. **290:**17262–17268, 2015.
[b]KD, dissociation constant (nM) fold change = KD at pH 5.5/KD at pH 7.4.

Modifying Protein Specificity

The enzyme tissue plasminogen activator (tPA) is a serine protease that is medically useful for the dissolution of blood clots. However, tPA is rapidly cleared from the circulation, so it must be administered by infusion. Therefore, to be effective with this form of delivery, high initial concentrations of tPA must be used. Unfortunately, under these conditions, tPA can cause nonspecific internal bleeding. Thus, a long-lived tPA that has an increased specificity for fibrin in blood clots and is not prone to induce nonspecific bleeding would be desirable. These three properties were introduced by directed mutagenesis into the gene for the native form of tPA. First, changing threonine at position 103 (Thr-103) to asparagine (Asn) causes the enzyme to persist in rabbit plasma approximately 10 times longer than the native form. Second, changing the amino acids lysine-histidine-arginine-arginine (Lys-His-Arg-Arg) at positions 296 to 299 to alanines (Ala-Ala-Ala-Ala) produces an enzyme that is much more specific for fibrin than is the native form. Third, changing Asn-117 to glutamine (Gln) causes the enzyme to retain the level of fibrinolytic activity found in the native form. Moreover, combining these three mutations in a single construct allows all three activities to be expressed simultaneously (Table 3.18). Since its approval in 2000 by the U.S. Food and Drug Administration and in 2001 by the European Medicines Agency, the triple-mutant tPA, known as tenecteplase and produced in CHO cells, has saved the lives of patients suffering from acute myocardial infarctions; it is currently being studied in phase III clinical trials (see chapter 13) for the treatment of ischemic strokes.

Increasing Ligand Binding Affinity

Using random mutagenesis, it is possible to generate antibodies *in vitro* that are directed against a wide range of antigens or that have enhanced antigen affinity. Antibodies have enormous potential as diagnostic or therapeutic agents (see chapter 5); however, natural antibodies often do not

Table 3.18 Stabilities and activities of various modified versions of tPA[a,b]

tPA variant	Modification(s)	Stability in plasma	Fibrin binding	Activity in plasma	Activity vs clots
1	Thr(103) → Asn	10	0.34	0.68	0.56
2	LysHisArgArg(296–299) → AlaAlaAlaAla	0.85	0.93	0.13	1.01
3	Thr(103) → Asn, LysHisArgArg(296–299) → AlaAlaAlaAla	5.3	0.33	0.13	0.65
4	Thr(103) → Asn, Asn(117) → Gln	3.4	1.0	1.13	1.17
5	LysHisArgArg(296–299) → AlaAlaAlaAla, Asn(117) → Gln	1.2	1.33	0.16	1.38
6	Thr(103) → Asn, LysHisArgArg(296–299) → AlaAlaAlaAla, Asn(117) → Gln	8.3	0.87	0.06	0.85

[a]Data from Keyt et al., *Proc. Natl. Acad. Sci. USA* **91:**3670–3674, 1994.

[b]All of the values shown are normalized to the wild type. Plasma stability is the reciprocal of the time it takes for plasma clearance; larger numbers indicate a more stable derivative. Fibrin specificity is reflected by a high activity versus clots and a low activity in plasma.

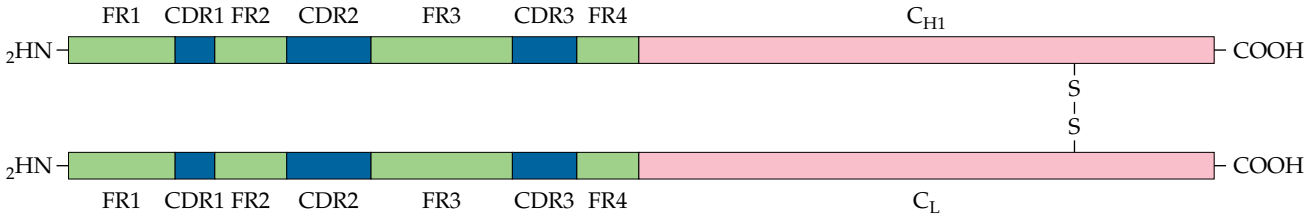
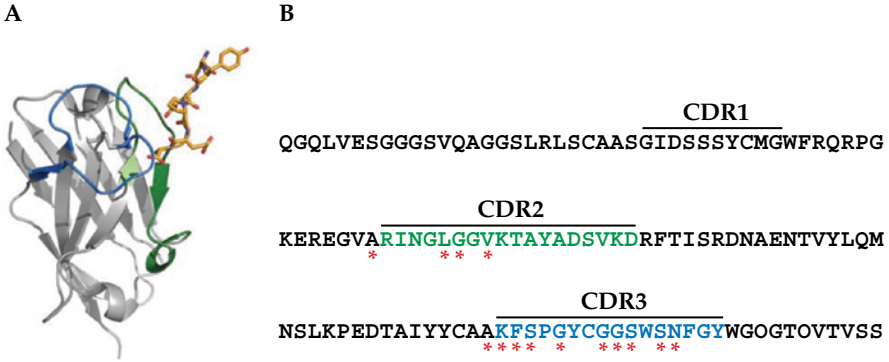


Figure 3.61 Structure of a Fab molecule. FR, framework region; CDR, complementarity-determining region. C_{H1} and C_L are constant domains from the heavy and light chains of the antibody molecule, respectively. The N-terminal (NH_2) and C-terminal (COOH) ends of each polypeptide, as well as a disulfide bridge (-S-S-), are indicated.

have high enough affinity for some therapeutic applications. The portion of an antibody molecule that contains the ability to bind to an antigen is sometimes called a Fab fragment (see chapter 4), and within this fragment are hypervariable complementarity-determining regions (CDRs) separated by relatively invariant framework regions (Fig. 3.61). Together, the six CDRs, three from the variable part of the antibody light chain and three from the variable part of the heavy chain (see chapter 4), determine the specificity of an antibody molecule. Altering one or more of the amino acids in the CDRs can change the antigen specificity and/or binding affinity of the antibody.

To generate antibodies with improved binding to α-synuclein, a protein implicated in Parkinson's disease, mutations were introduced into the variable region of the antibody heavy-chain gene. Starting with the experimentally determined three-dimensional structure of the antibody variable region bound to the C-terminal domain of α-synuclein, two key CDRs (CDR2 and CDR3) involved in antigen binding were targeted for mutagenesis (Fig. 3.62). Amino acids within the two CDRs that

Figure 3.62 Random mutagenesis to increase the affinity of the antibody for α-synuclein. **(A)** The structure of the heavy-chain variable region of the antibody (gray) bound to the C-terminal region of α-synuclein (yellow). Antibody regions CDR2 (green) and CDR3 (blue) are important for interaction with α-synuclein. **(B)** The primary sequence of the heavy-chain variable region of the antibody. The asterisks indicate the amino acid positions where random mutations were introduced in the corresponding DNA. Modified from Tiller et al., *Front. Immunol.* **8:**986, 2017, © 2017 Tiller, Chowdhury, Li, Ludwig, Sen, Maranas and Tessier, CC BY 4.0.



were predicted to be important for antibody structure and stability were excluded. Because multiple mutations are often required to increase antigen affinity, degenerate oligonucleotide primers were employed to introduce mutations at 14 sites in the CDR regions by PCR. The DNA sequences encoding the N-terminal CDR2-containing and CDR3-containing regions were amplified separately and then assembled by PCR for cloning and expression in *S. cerevisiae* (see chapter 2). Using this approach, an antibody library with 9.4×10^6 variants was screened to identify several antibodies with greater than 5-fold increased affinity for α-synuclein. These antibodies each carry between four and six altered codons. Two mutants showed improved α-synuclein specificity compared to the wild-type antibody; however, one showed decreased stability likely due to structural changes in the CDRs as a consequence of the introduced mutations.

Modifying Cofactor Requirements

Subtilisins are a group of nonspecific serine proteases that are secreted into growth medium by Gram-positive bacteria and are widely used as biodegradable cleaning agents in laundry detergents. All subtilisins bind tightly (affinity constant [K_a] = ~10^7 M) to one or more molecules of calcium per molecule of enzyme, where calcium binding stabilizes the enzyme. Unfortunately, since subtilisins are used in industrial settings where there are a large number of chelating agents that can bind to and effectively remove calcium, these enzymes are rapidly inactivated under these conditions. To circumvent this problem, it is necessary first to abolish completely the ability of a subtilisin to bind calcium and then to attempt to increase the stability of this modified enzyme in the absence of bound calcium.

The starting point for the development of a modified subtilisin was an isolated subtilisin gene from *Bacillus amyloliquefaciens*. Prior to this work, the subtilisin protein had been well characterized, and its high-resolution X-ray crystallographic structure had been determined. Oligonucleotide-directed mutagenesis was used to construct a mutant form of the gene for this enzyme by deleting the nucleotides encoding the portion of the protein—amino acids 75 to 83—that is responsible for binding to calcium (Fig. 3.63). The protein without this stretch of amino acids does not bind calcium and, surprisingly, retains an overall conformation that is similar to that of the native form.

The next steps in the development of a stable subtilisin from one that lacked a calcium-binding domain entailed predicting which sites might contribute to stability and which amino acids should be placed at these sites. The researchers assumed that any of the amino acids that had previously interacted with the calcium-binding loop in the native form of the enzyme were potential candidates for change. In total, 10 amino acids were considered to be candidates for modification. Moreover, since it was not known *a priori* which particular amino acids might best contribute to stabilizing the enzyme molecule, random mutagenesis was used for each of these sites.

The amino acids selected for mutagenesis came from four separate regions of the protein: the N terminus (amino acids 2 to 5), the omega loop (amino acids 36 to 44), an α-helical region (amino acids 63 to 85), and a β-pleated region (amino acids 202 to 220). To identify the best amino acid at a

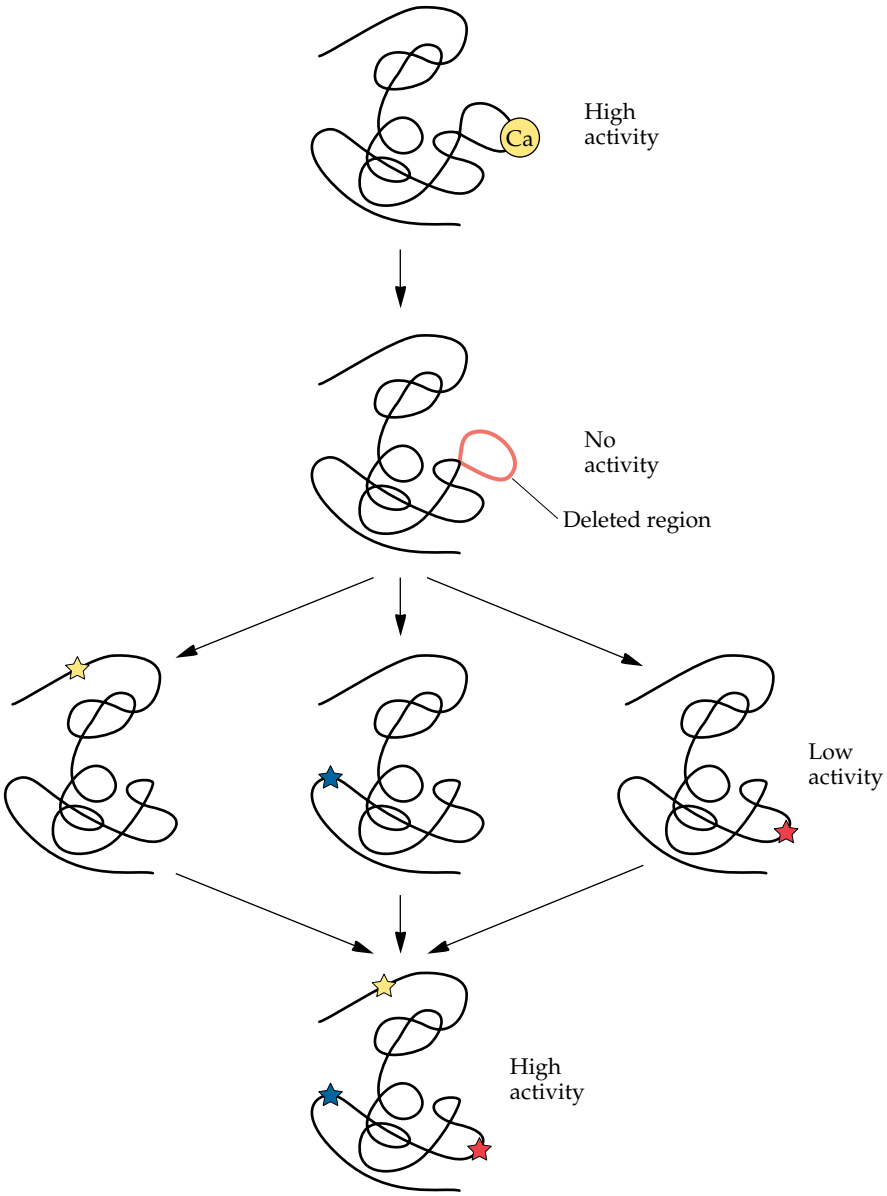


Figure 3.63 Genetic engineering of calcium-independent subtilisin. The native calcium-containing enzyme is highly active but lost almost all of its activity when the loop that binds calcium was deleted. The gene encoding the enzyme with the deleted calcium-binding region was subjected to random mutagenesis. After several rounds of mutagenesis, mutants with stabilizing mutations and a low level of activity were selected. Several of these mutations (indicated by colored stars) were combined into a single derivative to generate subtilisin that does not require calcium and has a high level of activity.

particular position, *Bacillus subtilis* clones expressing the mutated proteins were grown in the wells of microtiter plates, heated to 65°C for 1 hour, allowed to cool, and then assayed for subtilisin activity. It was necessary to express the active calcium-free subtilisin in the host *B. subtilis* because it was lethal when expressed in *E. coli*.

Table 3.19 Effects of random mutations of selected amino acid residues on the stability of a subtilisin lacking a calcium-binding domain[a]

Region of protein	Amino acid residue	Stabilizing mutation	Fold increase in half-life
N terminus	2	Gln → Lys	2.0
	3	Ser → Cys	17.0
	4	None found	None
	5	Pro → Ser	1.2
Omega loop	41	Asp → Ala	1.2
	44	Lys → Asn	1.2
α-helix	73	Ala → Leu	2.6
	74	None found	None
β-pleat structure	206	Gln → Cys	17.0
	214	None found	None

[a]Data from Strausberg et al., *Bio/Technology* **13:**669–673, 1995.

[b]The mutations at positions 3 and 206 to Cys occur in the same clone and provide such a high level of stability because of the formation of the disulfide bridge between these residues.

After the initial screening, stabilizing mutations were identified at 7 of the 10 positions that were examined (Table 3.19). When these stabilizing mutations were combined into a single gene, the enzyme that was produced had kinetic properties that were very similar to those of the native form of subtilisin. Moreover, the modified form of subtilisin was nearly 10 times more stable than the native form of the enzyme in the absence of calcium and, surprisingly, about 50% more stable than the native enzyme in the presence of calcium. This work demonstrates that complex properties of enzymes that involve a large number of different amino acids can be genetically engineered.

Decreasing Protease Sensitivity

The human serine protease mesotrypsin is overexpressed in many malignant tumors and is associated with invasiveness and metastasis of several cancers, including prostate, pancreatic, lung, colon, and breast cancers. It is therefore a potential target for inhibition in cancer therapy. One promising inhibitor is the protease inhibitor domain from the human amyloid precursor protein (APPI). However, similar to other potential protease inhibitors, APPI is susceptible to cleavage by mesotrypsin and does not bind to mesotrypsin with high affinity. Thus, APPI was subjected to random mutagenesis and screening to identify inhibitors with enhanced proteolytic stability and mesotrypsin affinity. First, the 300-bp DNA sequence encoding APPI was assembled by PCR from six overlapping oligonucleotides (see chapter 2) that contained codons optimized for expression in *S. cerevisiae* and *P. pastoris* and cloned into a vector. This sequence was then used as a template for error-prone PCR using low-fidelity Taq polymerase, and the resulting PCR products were cloned into a vector and transformed into the yeast strains to generate a library of 9×10^6 clones, each carrying a variant of APPI with 0 to 3 mutations at random positions throughout the molecule. The mutated APPI sequences were cloned into the vector in-frame with the coding sequence for the Aga2p agglutinin cell adhesion protein so that they were expressed as fusion proteins directed to the yeast cell surface (Fig. 3.64A). Display of APPI on the surface of the yeast cell facilitated screening of a large number of clones for enhanced

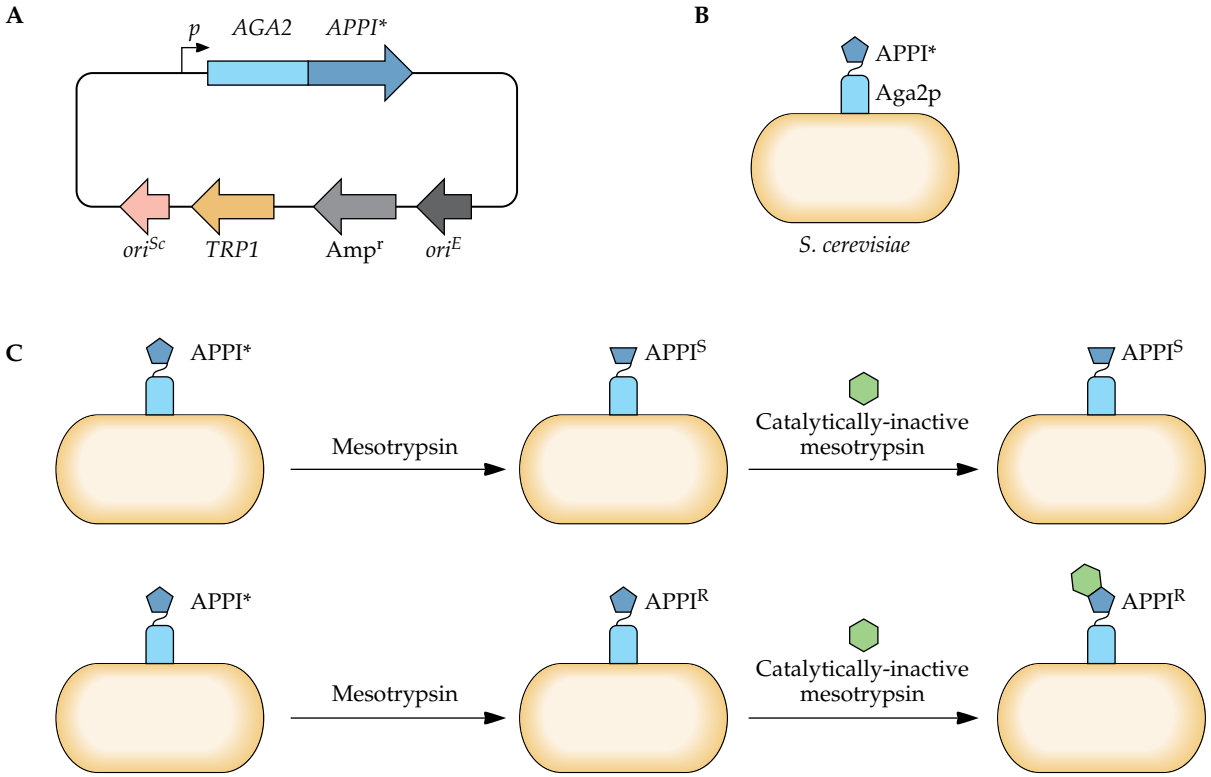


Figure 3.64 Random mutagenesis to decrease protease sensitivity of a protease inhibitor. **(A)** The gene encoding the protease inhibitor domain of the human amyloid precursor protein (*APPI*) was amplified by error-prone PCR and cloned into a vector in the same reading frame as the coding sequence for the yeast Aga2p agglutinin protein (*AGA2*) to generate a library of APPI variants (*APPI**) in *S. cerevisiae*. **(B)** The Aga2p fusion partner directed APPI* to the surface of *S. cerevisiae* cells to facilitate screening of protease-resistant variants. **(C)** The *S. cerevisiae* library was incubated with the protease mesotrypsin to cleave susceptible APPI variants ($APPI^S$), including wild-type APPI, and then with a labeled, mutant form of mesotrypsin that does not have proteolytic activity but binds to intact APPI to identify protease-resistant APPI ($APPI^R$). *p*, *GAL1* promoter; ori^{SC} and *TRP1*, origin of replication and selectable marker for *S. cerevisiae*, respectively; ori^E and Amp^r, origin of replication and selectable marker for *E. coli*, respectively. Data from Cohen et al., *Biochem. J.* **473:**1329–1341, 2016.

resistance to proteolysis by mesotrypsin (Fig. 3.64B). This was determined by incubating the yeast library first with mesotrypsin to cleave susceptible APPI variants and then with a mutant form of mesotrypsin that binds to APPI but does not have proteolytic activity (Fig. 3.64C). The mutant form of mesotrypsin was labeled with biotin for visualization of uncleaved, surface-displayed APPI by fluorescently labeled streptavidin. Wild-type APPI was rapidly cleaved by mesotrypsin, even at low concentrations of the protease. However, the screen identified several clones displaying mutant APPIs that were resistant to all concentrations of mesotrypsin tested. DNA sequencing revealed that three amino acid changes were common among the proteolytically resistant APPIs: M17G, I18F, and F34V, where the number refers to the position of the amino acid and the

letters before and after the number indicate the amino acids in the wild-type and mutants, respectively. APPI with mutations M17G or I18F not only showed proteolytic stability but also enhanced binding to mesotrypsin. Combining these mutations resulted in an APPI triple mutant with 83-fold greater proteolytic stability and 1,500-fold greater affinity for mesotrypsin compared to wild-type APPI. Importantly, the APPI variants with the individual and combined mutations inhibited mesotrypsin to a greater extent than wild-type APPI in solution and in cell assays. The triple mutant APPI strongly inhibited invasiveness of a highly aggressive, metastatic prostate cancer cell line; in these cells mesotrypsin production is essential for invasive behavior. Invasiveness was not suppressed by the same concentration of wild-type APPI. These results suggest that mutant APPI may be a promising new treatment for mesotrypsin-dependent metastatic cancers.

summary

The production of a protein requires that the encoding gene be properly transcribed and then that the mRNA be translated. To produce a heterologous protein in a prokaryote, a suitable prokaryotic promoter sequence is necessary for the initiation of transcription at the correct nucleotide site, and a terminator sequence at the end of the gene is essential for the cessation of transcription. The aim of many biotechnology applications is to produce large amounts of protein, so it is necessary to use a strong promoter that supports transcription at a high level. However, continuous transcription of a cloned gene drains the energy reserves of the host cell, and therefore it is also necessary to use a promoter system whose activity can be regulated. For translation, a ribosome-binding site is placed in the DNA region that precedes the translation initiation site (start codon), and a termination sequence (stop codon) is included at the end of the protein-coding sequence to ensure that translation stops at the correct amino acid. Codon optimization is often required for high-level production of foreign proteins in many host cells. If secretion of the protein is desired, the DNA sequence preceding the cloned gene should include a signal sequence in the same reading frame as the target gene. In addition, amino acid purification tags may be added to purify the recombinant protein by affinity chromatography.

High levels of expression of some foreign proteins in bacterial hosts can result in misfolded proteins that form insoluble inclusion bodies. This can be avoided by growing the bacterium at low temperatures, coexpressing chaperone proteins, overexpressing enzymes that catalyze the formation of disulfide bonds, and/or expressing the target protein as a fusion protein. To prevent rapid degradation of a recombinant protein by host proteolytic enzymes, the gene is altered to encode one or more additional amino acids at its N terminus or to remove protease recognition sequences. During large-scale production of recombinant proteins, plasmids may be unstable and lost from the population. To overcome this problem, researchers have developed protocols for integrating a cloned gene into a chromosomal site of the host organism. Under these conditions, the gene is maintained stably as part of the host DNA.

Although many heterologous proteins have been successfully synthesized in prokaryotic host cells, some proteins require eukaryote-specific posttranslational modifications, such as glycosylation, to be functional. Consequently, expression systems were devised for fungal, insect, and mammalian cells. With respect to the ease and likelihood of obtaining an authentic protein from a cloned gene, each of these systems has distinct merits and shortcomings. In other words, there is no single eukaryotic host cell that is capable of producing an authentic protein at a high level from every cloned gene.

All eukaryotic expression vectors have the same basic format. The target gene, which may be equipped with sequences that facilitate the secretion and purification of the heterologous protein, is under the control of eukaryotic promoter, polyadenylation, and transcription terminator sequences. To simplify both maintenance and recombinant DNA manipulations, eukaryotic expression vectors are routinely maintained in *E. coli*.

summary *(continued)*

Several different fungus-based expression systems have been developed for the production of heterologous proteins. The yeast *S. cerevisiae*, which is well characterized genetically and can be grown in large fermenters, has been used extensively for this purpose. Both episomal and integrating expression vectors, as well as artificial chromosomes, have been constructed. However, with *S. cerevisiae* as the host cell, a number of recombinant proteins are hyperglycosylated, and in some cases, protein yields are low because the capacity of the cell to properly fold and secrete proteins has been exceeded. Other yeast and filamentous fungal systems have been developed for the production of heterologous proteins. Of these, the methylotrophic yeast *P. pastoris* has been used successfully because of the low occurrence of hyperglycosylation, the ease of obtaining high cell densities, and the rapid and strong response of the *AOX1* promoter (usually used to drive the target gene) to methanol. A "humanized" strain of *P. pastoris* has been genetically altered to produce glycoproteins with glycosylation patterns that are identical to those found on the same proteins produced in human cells.

A large number of biologically active heterologous proteins have also been produced in insect cells grown in culture using baculoviruses to deliver the target gene into the insect host cell. This system is advantageous because posttranslational protein modification is similar in insects and mammals, and the baculoviruses used in these systems do not infect humans or other insect cells. The baculovirus most commonly used as a vector is AcMNPV. A target gene is inserted into the AcMNPV genome by homologous or site-specific recombination between sequences on a transfer vector carrying the target gene and the AcMNPV DNA. Recombinant AcMNPV DNA carrying the target gene is introduced into insect cells for heterologous-protein production. Improved insect host cells have been developed through genetic engineering to increase protein yields and to ensure that target proteins are properly glycosylated. In addition to production of a single protein of interest, the baculovirus–insect expression system is particularly amenable to producing functional multimeric protein complexes, such as virus-like particles, which are effective vaccines.

Many therapeutic proteins that require a full complement of posttranslational modifications are now produced in cultured mammalian cells, such as CHO cells. Most of the vectors that have been developed to introduce foreign genes into mammalian cells are based on mammalian viruses, especially SV40. The viral genome has been altered to remove some viral genes required for replication and viral-protein production and to include suitable mammalian transcription and translation signals to drive expression of the cloned gene. Expression of chromosomally integrated target genes can be increased by altering the epigenetic state of the insertion site through histone acetylation or insertion of chromatin-relaxing DNA elements. A major challenge for production of high levels of heterologous proteins in mammalian cell lines is preventing cell death, which is often induced by the stressful conditions of large-scale bioreactors. Strategies to improve cell growth and protein yields include genetically engineering host cells to block the transcription factor that induces apoptosis, to prevent accumulation of toxic metabolites in the culture medium, and to increase expression of proteins required for proper protein folding and secretion.

Natural proteins are often not well suited for biotechnology applications. For example, an enzyme may unfold, and therefore be inactivated, at high temperatures employed in an industrial process, or a therapeutic protein may be short lived due to protease sensitivity, necessitating administration of high, somewhat toxic, doses. Random or directed mutagenesis can be employed to alter the nucleotide sequence encoding a protein to improve its stability, activity, specificity, cofactor requirements, or protease resistance. Straightforward protocols have been developed to introduce nucleotide substitutions into a gene. When the specific amino acids that contribute to a property are known in advance, the defined nucleotide changes can be introduced on an oligonucleotide by overlap extension or inverse PCR. When the amino acid changes that will result in the desired property of a protein are unknown, libraries of randomly mutated sequences can be generated by performing a PCR under conditions that increase the error rate or by employing degenerate oligonucleotide primers. Most of the mutations will decrease the function of the encoded protein, and therefore the libraries must be screened to identify proteins with desired characteristics. Shuffling of DNA segments from two or more genes creates a large number of hybrid proteins that can also be screened for unique biological activity.

REFERENCES

Baghban R, Farajnia S, Rajabibazl M, Ghasemi Y, Mafi A, Hoseinpoor R, Rahbarnia L, Aria M. 2019. Yeast expression systems: overview and recent advances. *Mol Biotechnol* **61:**365–384.

Barnes LM, Dickson AJ. 2006. Mammalian cell factories for efficient and stable protein expression. *Curr Opin Biotechnol* **17:**381–386.

Baumgarten T, Ytterberg AJ, Zubarev RA, de Gier J-W. 2018. Optimizing recombinant protein production in the

Escherichia coli periplasm alleviates stress. *Appl Environ Microbiol* **84:**e00270-18.

Berger I, **Fitzgerald DJ**, **Richmond TJ.** 2004. Baculovirus expression system for heterologous multiprotein complexes. *Nat Biotechnol* **22:**1583–1587.

Browning DF, **Richards KL**, **Peswani AR**, **Roobol J**, **Busby SJW**, **Robinson C.** 2017. *Escherichia coli* "TatExpress" strains super-secrete human growth hormone into the bacterial periplasm by the Tat pathway. *Biotechnol Bioeng* **114:**2828–2836.

Burdette LA, **Leach SA**, **Wong HT**, **Tullman-Ercek D.** 2018. Developing Gram-negative bacteria for the secretion of heterologous proteins. *Microb Cell Fact* **17:**196–202.

Çelik E, **Çalik P.** 2012. Production of recombinant proteins by yeast cells. *Biotechnol Adv* **30:**1108–1118.

Chatterjee R, **Yuan L.** 2006. Directed evolution of metabolic pathways. *Trends Biotechnol* **24:**28–38.

Chen R. 2012. Bacterial expression systems for recombinant protein production: *E. coli* and beyond. *Biotechnol Adv* **30:**1102–1107.

Chong S, **Mersha FB**, **Comb DG**, **Scott ME**, **Landry D**, **Vence LM**, **Perler FB**, **Benner J**, **Kucera RB**, **Hirvonen CA**, **Pelletier JJ**, **Paulus H**, **Xu MQ.** 1997. Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element. *Gene* **192:**271–281.

Cohen I, **Kayode O**, **Hockla A**, **Sankaran B**, **Radisky DC**, **Radisky ES**, **Papo N.** 2016. Combinatorial protein engineering of proteolytically resistant mesotrypsin inhibitors as candidates for cancer therapy. *Biochem J* **473:**1329–1341.

Costa S, **Almeida A**, **Castro A**, **Domingues L.** 2014. Fusion tags for protein solubility, purification and immunogenicity in *Escherichia coli*: the novel Fh8 system. *Front Microbiol* **5:**63.

de Boer HA, **Comstock LJ**, **Vasser M.** 1983. The *tac* promoter: a functional hybrid derived from the *trp* and *lac* promoters. *Proc Natl Acad Sci USA* **80:**21–25.

Eijsink VGH, **Bjørk A**, **Gåseidnes S**, **Sirevåg R**, **Synstad B**, **van den Burg B**, **Vriend G.** 2004. Rational engineering of enzyme stability. *J Biotechnol* **113:**105–120.

Elleuche S, **Pöggeler S.** 2010. Inteins, valuable genetic elements in molecular biology and biotechnology. *Appl Microbiol Biotechnol* **87:**479–489.

Ernst JF. 1988. Codon usage and gene expression. *Trends Biotechnol* **6:**196–199.

Fong BA, **Wu W-Y**, **Wood DW.** 2010. The potential role of self-cleaving purification tags in commercial-scale processes. *Trends Biotechnol* **28:**272–279.

Gasser B, **Saloheimo M**, **Rinas U**, **Dragosits M**, **Rodríguez-Carmona E**, **Baumann K**, **Giuliani M**, **Parrilli E**, **Branduardi P**, **Lang C**, **Porro D**, **Ferrer P**, **Tutino ML**, **Mattanovich D**, **Villaverde A.** 2008. Protein folding and conformational stress in microbial cells producing recombinant proteins: a host comparative overview. *Microb Cell Fact* **7:**11–29.

Geisow MJ. 1991. Both bane and blessing: inclusion bodies. *Trends Biotechnol* **9:**368–369.

Gellissen G, **Kunze G**, **Gaillardin C**, **Cregg JM**, **Berardi E**, **Veenhuis M**, **van der Klei I.** 2005. New yeast expression platforms based on methylotrophic *Hansenula polymorpha* and *Pichia pastoris* and on dimorphic *Arxula adeninivorans* and *Yarrowia lipolytica*: a comparison. *FEMS Yeast Res* **5:**1079–1096.

Glick BR. 1995. Metabolic load and heterologous gene expression. *Biotechnol Adv* **13:**247–261.

Mahmoudi Gomari M, **Saraygord-Afshari N**, **Farsimadan M**, **Rostami N**, **Aghamiri S**, **Farajollahi MM.** 2020. Opportunities and challenges of the tag-assisted protein purification techniques: applications in the pharmaceutical industry. *Biotechnol Adv* **45:**107653.

Hamilton SR, **Gerngross TU.** 2007. Glycosylation engineering in yeast: the advent of fully humanized yeast. *Curr Opin Biotechnol* **18:**387–392.

Heckman KL, **Pease LR.** 2007. Gene splicing and mutagenesis by PCR-driven overlap extension. *Nat Protoc* **2:**924–932.

Jimenez-Rosales A, **Flores-Merino MV.** 2018. Tailoring proteins to re-evolve nature: a short review. *Mol Biotechnol* **60:**946–974.

Kaur J, **Sharma R.** 2006. Directed evolution: an approach to engineer enzymes. *Crit Rev Biotechnol* **26:**165–199.

Keyt BA, **Paoni NF**, **Refino CJ**, **Berleau L**, **Nguyen H**, **Chow A**, **Lai J**, **Pena L**, **Pater C**, **Ogez J.** 1994. A faster-acting and more potent form of tissue plasminogen activator. *Proc Natl Acad Sci USA* **91:**3670–3674.

Kjeldsen T, **Hach M**, **Balschmidt P**, **Havelund S**, **Pettersson AF**, **Markussen J.** 1998. Prepro-leaders lacking N-linked glycosylation for secretory expression in the yeast *Saccharomyces cerevisiae*. *Protein Expr Purif* **14:**309–316.

Kurokawa Y, **Yanagi H**, **Yura T.** 2000. Overexpression of protein disulfide isomerase DsbC stabilizes multiple-disulfide-bonded recombinant protein produced and transported to the periplasm in *Escherichia coli*. *Appl Environ Microbiol* **66:**3960–3965.

Kwaks THJ, **Otte AP.** 2006. Employing epigenetics to augment the expression of therapeutic proteins in mammalian cells. *Trends Biotechnol* **24:**137–142.

Kwaks THJ, **Sewalt RGAB**, **van Blokland R**, **Siersma TJ**, **Kasiem M**, **Kelder A**, **Otte AP.** 2005. Targeting of a histone acetyltransferase domain to a promoter enhances protein expression levels in mammalian cells. *J Biotechnol* **115:** 35–46.

Lipinszki Z, **Vernyik V**, **Farago N**, **Sari T**, **Puskas LG**, **Blattner FR**, **Posfai G**, **Gyorfy Z.** 2018. Enhancing the translational capacity of *E. coli* by resolving the codon bias. *ACS Synth Biol* **7:**2656–2664.

Lobstein J, **Emrich CA**, **Jeans C**, **Faulkner M**, **Riggs P**, **Berkmen M.** 2012. SHuffle, a novel *Escherichia coli* protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. *Microb Cell Fact* **11:**56–62.

Martinez-Morales F, **Borges AC**, **Martinez A**, **Shanmugam KT**, **Ingram LO.** 1999. Chromosomal integration of heterologous DNA in *Escherichia coli* with precise removal of markers and replicons used during construction. *J Bacteriol* **181:**7143–7148.

Myung S-H, **Park J**, **Han J-H**, **Kim T-H.** 2020. Development of the mammalian expression vector system that can be

induced by IPTG and/or lactose. *J Microbiol Biotechnol* **30:**1124–1131.

Ness JE, Welch M, Giver L, Bueno M, Cherry JR, Borchert TV, Stemmer WPC, Minshull J. 1999. DNA shuffling of subgenomic sequences of subtilisin. *Nat Biotechnol* **17:**893–896.

O'Flaherty R, Bergin A, Flampouri E, Mota LM, Obaidi I, Quigley A, Xie Y, Butler M. 2020. Mammalian cell culture for production of recombinant proteins: a review of the critical steps in their biomanufacturing. *Biotechnol Adv* **43:**107552.

Possee RD, Chambers AC, Graves LP, Aksular M, King LA. 2019. Recent developments in the use of baculovirus expression vectors, p 215–230. *In* Bonning BC (ed), *Insect Molecular Virology: Advances and Emerging Trends*. Caister Academic Press, Poole, United Kingdom.

Prasad JM, Migliorini M, Galisteo R, Strickland DK. 2015. Generation of a potent low density lipoprotein receptor-related protein 1 (LRP1) antagonist by engineering a stable form of the receptor-associated protein (RAP) D3 domain. *J Biol Chem* **290:**17262–17268.

Punt PJ, van Biezen N, Conesa A, Albers A, Mangnus J, van den Hondel C. 2002. Filamentous fungi as cell factories for heterologous protein production. *Trends Biotechnol* **20:**200–206.

Qiu J, Swartz JR, Georgiou G. 1998. Expression of active human tissue-type plasminogen activator in *Escherichia coli*. *Appl Environ Microbiol* **64:**4891–4896.

Rosano GL, Morales ES, Ceccarelli EA. 2019. New tools for recombinant protein production in *Escherichia coli*: a 5-year update. *Protein Sci* **28:**1412–1422.

Schröder M. 2008. Engineering eukaryotic protein factories. *Biotechnol Lett* **30:**187–196.

Seo JH, Bailey JE. 1985. Effects of recombinant plasmid content on growth properties and cloned gene product formation in *Escherichia coli*. *Biotechnol Bioeng* **27:**1668–1674.

Shi X, Jarvis DL. 2007. Protein N-glycosylation in the baculovirus-insect cell system. *Curr Drug Targets* **8:**1116–1125.

Simmons LC, Yansura DG. 1996. Translational level is a critical factor for the secretion of heterologous proteins in *Escherichia coli*. *Nat Biotechnol* **14:**629–634.

Smolskaya S, Andreev YA. 2019. Site-specific incorporation of unnatural amino acids into *Escherichia coli* recombinant protein: methodology development and recent achievement. *Biomolecules* **9:**255–272.

Song M, Sukovich DJ, Ciccarelli L, Mayr J, Fernandez-Rodriguez J, Mirsky EA, Tucker AC, Gordon DB, Marlovits TC, Voigt CA. 2017. Control of type III protein secretion using a minimal genetic system. *Nat Commun* **8:** 14737.

Steele KH, Stone BJ, Franklin KM, Fath-Goodin A, Zhang X, Jiang H, Webb BA, Geisler C. 2017. Improving the baculovirus expression vector system with vankyrin-enhanced technology. *Biotechnol Prog* **33:**1496–1507.

Strausberg SL, Alexander PA, Gallagher DT, Gilliland GL, Barnett BL, Bryan PN. 1995. Directed evolution of a subtilisin with calcium-independent stability. *Biotechnology (N Y)* **13:**669–673.

Thak EJ, Yoo SJ, Moon HY, Kang HA. 2020. Yeast synthetic biology for designed cell factories producing secretory recombinant proteins. *FEMS Yeast Res.* **20:**foaa009.

Tiller KE, Chowdhury R, Li T, Ludwig SD, Sen S, Maranas CD, Tessier PM. 2017. Facile affinity maturation of antibody variable domains using natural diversity mutagenesis. *Front Immunol* **8:**986.

Tobias JW, Shrader TE, Rocap G, Varshavsky A. 1991. The N-end rule in bacteria. *Science* **254:**1374–1377.

van Oers MM, Pijlman GP, Vlak JM. 2015. Thirty years of baculovirus-insect cell protein expression: from dark horse to mainstream technology. *J Gen Virol* **96:**6–23.

Van Zyl JHD, Den Haan R, Van Zyl WH. 2016. Overexpression of native *Saccharomyces cerevisia*e ER-to-Golgi SNARE genes increased heterologous cellulase secretion. *Appl Microbiol Biotechnol* **100:**505–518.

Wang L, Brock A, Herberich B, Schultz PG. 2001. Expanding the genetic code of *Escherichia coli*. *Science* **292:**498–500.

Widmaier DM, Tullman-Ercek D, Mirsky EA, Hill R, Govindarajan S, Minshull J, Voigt CA. 2009. Engineering the *Salmonella* type III secretion system to export spider silk monomers. *Mol Syst Biol* **5:**309.

review questions

1. What DNA sequence elements are required for expression of a cloned gene in a prokaryotic host?
2. What is a strong promoter? Why is a strong promoter not always desirable for expression of a cloned gene?
3. What is a regulatable promoter? How is the *E. coli lac* promoter used to regulate the expression of a cloned gene?
4. The promoter for gene 10 of the *E. coli* bacteriophage T7 is an example of a strong promoter. How is it used to express a cloned gene?
5. Why is codon optimization often required for production of high levels of a recombinant protein?
6. What are inclusion bodies, and how can their formation be avoided?
7. Describe a strategy to secrete a recombinant protein using *E. coli* as a host.
8. Discuss some strategies to purify a recombinant protein produced in a prokaryotic host. Consider that a protein may be used as a human therapeutic agent.
9. Why is it sometimes advantageous to integrate a target gene into the chromosomal DNA of a prokaryotic host? How might this be achieved?

10. During the course of integrating a target gene into the chromosomal DNA of the host bacterium, a marker gene may also be inserted into the chromosomal DNA. What strategy could be used to excise only the marker gene?

11. What are the major posttranslational modifications of eukaryotic proteins in the endoplasmic reticulum and Golgi apparatus?

12. Describe the features of a eukaryotic expression vector.

13. What criteria are used to decide if a particular recombinant protein should be produced in a yeast, insect, or mammalian cell system?

14. What are the advantages and disadvantages of the different classes of yeast vectors for producing a biotechnology product?

15. Describe some of the strategies that have been used to increase proper folding and secretion of recombinant proteins from yeast cells.

16. Discuss the salient features of a *P. pastoris* high-expression integrating vector system. How has *P. pastoris* been "humanized"?

17. What are baculoviruses?

18. Describe a strategy that can be used to insert a target gene into the baculovirus genome for expression in insect cells.

19. How have insect cell lines been improved for expression of recombinant proteins?

20. Describe the main features of an extrachromosomal mammalian-cell expression vector.

21. Why are yields of recombinant proteins produced by mammalian cells in large bioreactors generally low? How can yields be improved?

22. What is chromatin, and how does it affect gene expression? Describe some of the strategies that have been developed to increase expression levels of a target gene that is integrated into a chromosome of a eukaryotic host cell.

23. You have produced a recombinant enzyme for an industrial application but have found that the protein is unstable at moderately high temperatures. Describe how you would increase the stability of the enzyme. Assume that you have determined the DNA sequence of the gene encoding the enzyme and the three-dimensional structure of the enzyme.

24. How can nonstandard amino acids be incorporated into proteins, thereby producing an altered form of the target protein?

25. You have produced a recombinant enzyme for an industrial application but have found that the catalytic activity is low. Describe how you would increase the activity of the enzyme if the DNA sequence of the gene was known but the structure of the enzyme was not determined.

26. Outline two ways in which DNA shuffling may be used to generate hybrid genes.

27. How would you engineer a protein to be less sensitive to proteolytic digestion?

28. How can the gene(s) encoding the variable region of an antibody be modified to increase the affinity of the antibody for an antigen?

Immunological Approaches To Detect Protein Biomarkers
Antibodies
Agglutination
Enzyme-Linked Immunosorbent Assays
Protein Arrays To Detect Polygenic Diseases
Immunoassays for Protein Conformation-Specific Disorders
DNA-Based Diagnostic Approaches
Hybridization Probes
PCR-Based Detection Methods
CRISPR-Cas-Based Diagnostic Assays
DNA Microarrays
Whole-Genome Sequencing To Assess Genetic Disease Risk
Detecting RNA Signatures of Disease
Detection of Disease-Associated Changes in Gene Expression
Detection of RNA Signatures of Antibiotic Resistance in Bacteria
Detection of miRNA Signatures of Disease
Biofluorescent and Bioluminescent Systems
Fluorescent Proteins
Luciferase
Microbial Biosensors
SUMMARY
REFERENCES
REVIEW QUESTIONS

4 Molecular Diagnostics

Molecular diagnostics is concerned with the detection of specific proteins, nucleic acid sequences, or small molecules in humans, animals, plants, microorganisms, or environmental samples. The presence of a specific molecule may be symptomatic of an infectious or genetic disease, reveal the identity of an individual or its relatedness to other individuals or populations, establish the presence of an individual at a crime scene, or be indicative of microorganisms or transgenes in food or environmental samples. A diagnostic test must have high specificity for the target molecule (i.e., few false-positive results), have high sensitivity to detect low levels of the target (i.e., few false-negative results), and be sufficiently rapid and inexpensive for routine, and possibly high-throughput, analysis.

A major goal of diagnostic testing is to detect molecular biomarkers of disease as early as possible before the disease has significantly progressed, which limits treatment efficacy. Clinical diagnostic tests are also used to predict susceptibility to a disease or the response to a treatment, to determine disease prognosis (progress and outcome), and to monitor treatment efficiency. A disease biomarker is a specific molecule that has been determined to be present, or present at higher or lower levels, in diseased tissues compared to normal tissue. It can be a protein, specific sequence of DNA or RNA, or small metabolite that is analyzed qualitatively or quantitatively in body fluids such as blood, urine, sputum, or pus from a wound or in tissue acquired by biopsy. Samples that contain a small amount of the target molecule, for example, small-core needle biopsy samples or blood samples, may require some amplification of the target for accurate detection and quantification. The complex heterogeneity of clinical samples may

Molecular Biotechnology: Principles and Applications of Recombinant DNA,
Sixth Edition. Bernard R. Glick and Cheryl L. Patten.
© 2022 American Society for Microbiology.

also present a challenge for molecular diagnostics. Although many biomarkers have been identified to detect single-gene disorders such as cystic fibrosis or complex multigene disorders such as Alzheimer's disease, clinically relevant biomarkers are not well defined for many human diseases.

Immunological Approaches To Detect Protein Biomarkers

Proteins play a critical role in all cellular processes, including metabolism, communication, defense, reproduction, transport, and motility. Regulatory proteins maintain tight control over protein production to ensure that cells function normally. Dysregulation of these processes in diseased tissue is reflected in alterations in protein composition or levels of specific proteins. Characteristic changes in proteins have been used extensively to diagnose disease, either by detecting the presence or measuring levels of a specific protein biomarker or by determining protein profiles in polygenic diseases.

There are several advantages to using proteins as diagnostic biomarkers of disease. Abnormal levels of gene expression as a consequence of disease-associated mutations are more accurately quantified by measuring the protein directly rather than by measuring mRNA levels, which often do not precisely correlate with protein levels. Furthermore, proteins in diseased tissues may exhibit irregularities in posttranslational modification that cannot be detected using nucleic acids. Some diseases are a consequence of altered protein conformation (e.g., prion diseases and some neurodegenerative diseases) that may not be detected from nucleic acid sequences. On the other hand, the unique shape and often very low cellular levels of a protein are a challenge for development of a diagnostic assay. Antibodies can be produced that bind to target proteins with high affinity and specificity, and therefore immunological approaches, which employ antibodies, meet the criteria of sensitivity, specificity, and simplicity for diagnostic assays.

Antibodies

The molecules that are the basis for immunological approaches to diagnostics are antibodies. Antibodies, also known as immunoglobulins, are glycoproteins synthesized by B lymphocytes (plasma cells) that are activated by a specific antigen. They are secreted into the blood, lymphoid tissues, and mucosa of surface-exposed tissues such as the gastrointestinal, respiratory, and reproductive tracts, where they neutralize and eliminate extracellular microbes and toxins.

Structure of Antibodies

An antibody molecule has two regions with distinct functions: one region recognizes and binds to a specific molecule (antigen) of a pathogen that elicited its production, and the other effector region recruits immune cells such as phagocytes and molecules such as serum proteins (complement) that destroy the pathogen marked by the antibody. The antigen-binding region is highly variable in its amino acid sequence among antibody molecules and is referred to as the variable (V) region. The sequence variability endows an antibody with exquisite specificity of binding to a particular region of an antigen (epitope). On the other hand, the sequence in the region of the antibody that performs the effector functions does not vary among antibodies of a given class and is therefore termed the constant (C) region.

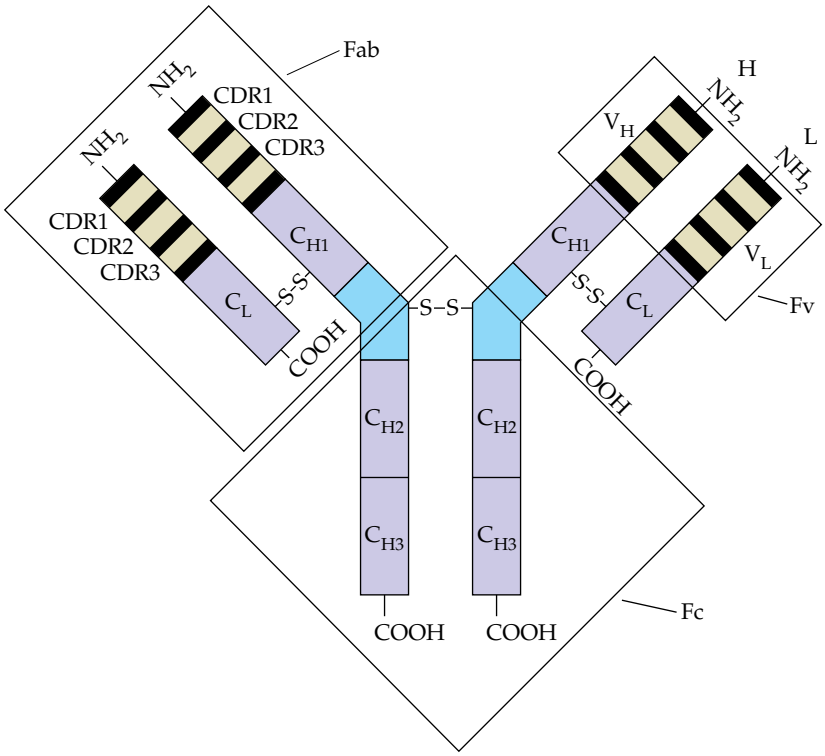


Figure 4.1 Structure of an antibody molecule. The H and L chains contain variable regions (V_L and V_H) with their CDRs (CDR1, CDR2, and CDR3) flanked by FRs (thick black lines) and constant domains (C_L, C_{H1}, C_{H2}, and C_{H3}). The Fv, Fab, and Fc portions of an antibody molecule are delineated. The N-terminal (NH_2) and C-terminal (COOH) ends of each polypeptide chain are indicated.

An antibody molecule consists of four polypeptides: two identical heavy (H) protein chains and two identical light (L) protein chains held together by covalent disulfide bonds and noncovalent interactions (hydrogen bonds, electrostatic forces, van der Waals forces, and hydrophobic forces) (Fig. 4.1). The H chains (50 kDa) are each of higher molecular mass than the L chains (25 kDa). The V regions of the H and L chains (V_H and V_L) together form the antigen recognition (binding) site of an antibody. The specific amino acids within this region that recognize and bind antigens are the three complementarity-determining regions (CDRs) (Fig. 4.1). The CDRs display the greatest variability in amino acid sequence of an antibody molecule. When the V_H and V_L domains are paired, the hypervariable regions (six in all) create a single hypervariable antigen-binding site (Fig. 4.1). In contrast, the amino acids in the framework regions (FRs) flanking the CDRs are relatively conserved among antibodies.

In addition to the V regions, each L chain contains one constant region (C_L), and each H chain contains three constant regions (C_{H1}, C_{H2}, and C_{H3}) (Fig. 4.1). The effector functions of an antibody molecule are mediated by the C regions of the H chain and are not associated at all with the L chain. When antibodies are digested with the proteolytic enzyme papain, three fragments are released: two identical (Fab) fragments, each of which contains an intact L chain linked by a disulfide bond to the C_{H1} region of the H chain, and one Fc fragment, which consists of two H chain fragments, each containing the C_{H2} and C_{H3} domains and joined by a disulfide bond (Fig. 4.1). Five different

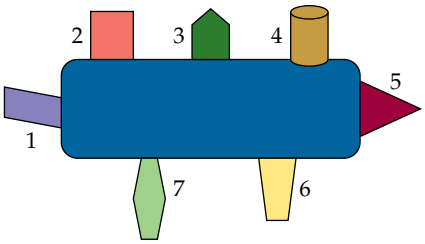


Figure 4.2 Schematic representation of a target antigen. The surface of the antigen depicted has seven (numbered 1 to 7) different antigenic determinants (epitopes). When an animal is immunized with this antigen, each antigenic determinant elicits the synthesis of a different antibody. Together, the different antibodies that interact with an antigen constitute a polyclonal antibody directed against that antigen.

classes of immunoglobulins (IgM, IgD, IgG, IgA, and IgE) are distinguishable by their Fc regions. After antigen binding by the Fab regions (specifically, the N-terminal portions known as the Fv region) of an intact antibody molecule, the Fc portion of the molecule stimulates an immune response. Together with other immune system proteins that bind to the Fc portion of antibody molecules, foreign pathogens are inactivated.

Polyclonal Antibodies

Upon stimulation by an antigen, each differentiated B cell synthesizes and secretes antibodies that recognize a single, discrete epitope (antigenic determinant) of the antigen and bind with high affinity. However, an antigen generally consists of several different epitopes (Fig. 4.2), and therefore, different B cells will produce different antibodies, each specific for only one of the epitopes. The set of heterogeneous antibodies that react with different epitopes of the same antigen is termed a polyclonal antibody; i.e., the antibodies in this set are produced by a group of several different B cells.

Polyclonal antibodies for experimental or diagnostic applications are produced by inoculating a laboratory animal (e.g., goat or rabbit) with an antigen. Often a subsequent, booster injection(s) with the same antigen is necessary. After several weeks, the antibodies are collected in the serum (antiserum). Production of high-concentration (titer) antisera usually requires the addition of an adjuvant, a substance such as alum or a bacterial component that effectively promotes the concentration of a coinoculated antigen and enhances the strength of an immune response.

Although polyclonal antisera are used extensively in biological assays, there are some disadvantages to their use. The effectiveness of a polyclonal antibody preparation varies from batch to batch because, at different times, the immune system of an animal responds more actively to different epitopes of the same antigen. Also, individual animals often respond differently to a particular antigen. This variation can affect the abilities of different preparations to neutralize antigens because different epitopes have different potencies (stimulating abilities). Hence, one batch of polyclonal antibody may have a low level of antibody molecules directed against a major epitope and not be as effective as a previous antibody preparation.

A fundamental objective for the application of antibodies as diagnostic agents or as components of therapeutic agents was to discover how to create a cell line that could be grown in culture and that would produce a single type of antibody molecule with a high affinity for a specific target antigen. Such a cell line would provide a consistent and continuous source of identical antibody molecules. Unfortunately, the B cells that synthesize antibodies do not reproduce in culture. However, it was envisioned that a hybrid cell type could be created to solve this problem. This hybrid would have the B-cell genetic components for producing antibodies and the cell division functions of a compatible cell type to enable the cells to grow in culture. It was known that normal B cells sometimes become cancer cells (myelomas) that acquire the ability to grow in culture while retaining many of the attributes of B cells. Thus, myeloma cells, especially those that did not produce antibody molecules, became candidates for fusion with antibody-producing B cells. In the mid-1970s, these ideas became reality.

Monoclonal Antibodies

To avoid problems with reproducibility of polyclonal antisera, it was desirable to make an unlimited supply of antibody molecules of homogeneous structure (derived from a single clone of B cells), high affinity, and

known specificity for a specific target antigen. This was achieved by the production of monoclonal antibodies from cultures of hybrid antibody-forming B cells (hybridomas) (Fig. 4.3).

The initial step leading to the production of a hybrid cell line that produces a single antibody entails the inoculation of a small animal (typically a mouse) with an antigen. After several inoculations over a period of a few weeks, the animals are tested to determine whether they have developed an immune response. If they have, they are sacrificed and their spleens are removed, washed, minced, and gently agitated to release individual cells, some of which are antibody-producing B cells. The splenic cell suspension is mixed with a suspension of myeloma cells that are genetically defective for the enzyme hypoxanthine-guanine phosphoribosyltransferase (HGPRT). The combined cell suspensions are mixed with 35% polyethylene glycol for a few minutes and then transferred to a growth medium containing hypoxanthine, aminopterin, and thymidine (HAT medium) (Fig. 4.3A).

The polyethylene glycol treatment facilitates fusion between cells, although the fusion events are rare and random. In the mixed population, there will be myeloma–spleen fusion cells, but also unfused myeloma cells, unfused spleen cells, myeloma–myeloma fusion cells, and spleen–spleen fusion cells. The HAT medium, however, allows only the myeloma–spleen fusion cells to grow, because none of the other cell types can proliferate in this medium. Unfused spleen cells and spleen–spleen fusion cells cannot grow in any culture medium. The HGPRT-deficient myeloma and myeloma–myeloma fusion cells cannot use hypoxanthine as a precursor for the biosynthesis of the purines guanine and adenine, which are essential for nucleic acid synthesis. However, they have a second, naturally occurring pathway for purine biosynthesis that utilizes the enzyme dihydrofolate reductase. Therefore, aminopterin is included in the medium because it inhibits dihydrofolate reductase activity. Hence, HGPRT-deficient myeloma and myeloma–myeloma fusion cells are unable to synthesize purines in HAT medium, so they die.

The spleen–myeloma fusion cells survive in HAT medium because the spleen cell contributes a functional HGPRT, which can utilize the hypoxanthine in the medium even though purine production by means of dihydrofolate reductase is blocked by aminopterin and because the cell division functions of the myeloma cell are active. Thymidine is provided to overcome the block in pyrimidine production that is caused by the inhibition of dihydrofolate reductase by aminopterin. About 10 to 14 days after the fusion treatment, only spleen–myeloma fusion cells have survived and grown in the HAT medium. These cells are then distributed into the wells of microtiter plates and grown on complete culture medium without HAT.

The next task is to identify those hybrid cells that produce antibody against the immunizing antigen. One common screening procedure uses the culture medium, which contains secreted antibodies. The medium is collected from the wells that have growing cells and is added to a well of another microtiter plate that has been precoated with the target antigen (Fig. 4.3B). If the culture medium contains an antibody (primary antibody) that recognizes an epitope of the antigen, it will bind to the antigen and not be washed away during a subsequent washing step. A second antibody (secondary antibody) that is specific for the primary (usually mouse) antibodies is added to the wells of the test plate. It will bind to any primary antibody that remains in the well.

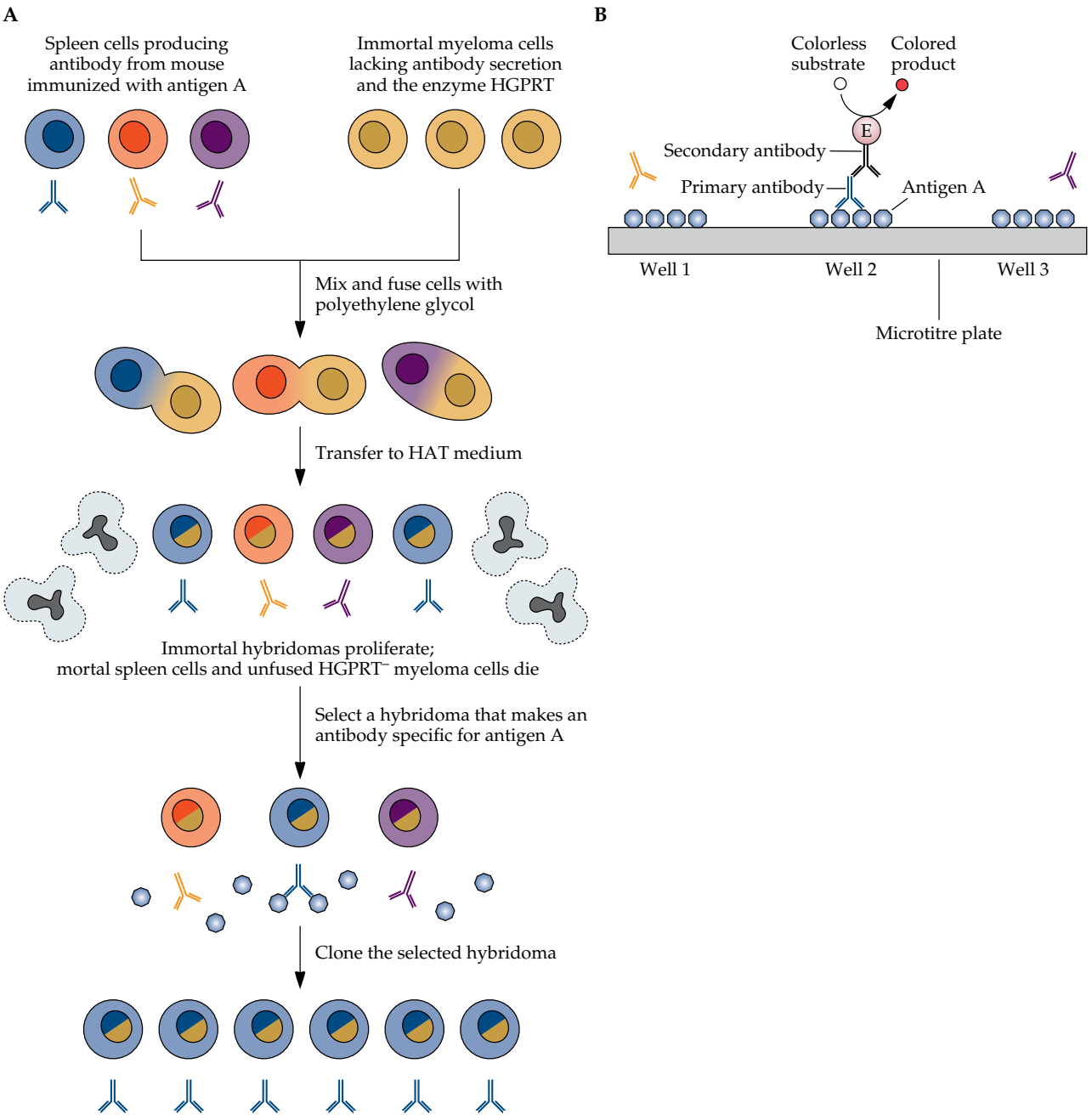


Figure 4.3 Creation of a hybridoma cell line that produces a monoclonal antibody. **(A)** Mice are immunized with an antigen and then given an intravenous booster 3 days before sacrifice to produce a large population of spleen cells secreting a specific antibody. Spleen cells die after a few days in culture, and therefore, to produce a continuous source of antibody, they are fused with immortalized myeloma cells (which do not produce antibodies) by using polyethylene glycol to produce a hybrid cell line called a hybridoma. Hybrid cells are selected on hypoxanthine-aminopterin-thymidine (HAT) medium. The *HGPRT* gene encoding the enzyme hypoxanthine-guanine phosphoribosyltransferase (HGPRT) contributed by the spleen cells allows fused spleen-myeloma cells to survive in HAT medium. These cells grow continuously in culture because of the malignant potential contributed by the myeloma cells. Unfused myeloma cells (which lack the *HGPRT* gene) and unfused spleen cells (depicted as cells with dark, irregular nuclei) die in HAT medium. Individual hybridoma cells are screened for antibody production. The cells that produce an antibody of the desired specificity are cloned by expansion from a single antibody-producing cell. The cloned hybridoma cells are grown in culture to produce large amounts of antibody. Since all cells of a hybridoma line make the same antibody, this antibody is termed monoclonal. **(B)** Assay to identify the hybridomas that produce antibodies that specifically bind to the target antigen. The medium from each cultured hybridoma is added to wells of a microtiter plate that are coated with the target antigen (antigen A). Antibodies (primary antibodies) secreted by the hybridomas into the culture medium added to well 2 bind to antigen A, while antibodies in the culture medium added to wells 1 and 3 do not and are subsequently washed away. Antibody-antigen binding is detected using a secondary antibody that binds to the primary antibody and is conjugated to an enzyme (E) that converts a colorless substrate into a colored product.

Before its use in the immunoassay, the secondary antibody is conjugated to an enzyme that can convert a colorless substrate to a colored compound. The presence of color in one of the test wells indicates that the original culture medium contained an antibody that was specific for the antigen. If the culture medium does not contain an antibody that binds to the antigen, then the first wash will remove the primary antibody. Therefore, when the secondary antibody is added, it has nothing to bind to and is removed by the second washing step. In a well where such a sequence of events occurs, the substrate solution remains colorless.

Those wells of the original microtiter plate whose media give a positive (color) response in the immunoassay may contain a mixture of cell fusions. These cells are therefore diluted with culture medium and seeded into fresh wells to establish cell lines from single cells (clones). After the clones have been cultured, their media are tested again to determine which cell lines (hybrid spleen–myeloma cells, or hybridomas) produce monoclonal antibodies that recognize the target antigen. If more than one specific hybridoma is isolated, further tests are conducted to determine whether the different clones produce antibody against the same or different epitopes. Each monoclonal antibody-producing clone can be maintained, more or less indefinitely, in culture. In addition, samples can be frozen in liquid nitrogen to provide a source of cells for future use.

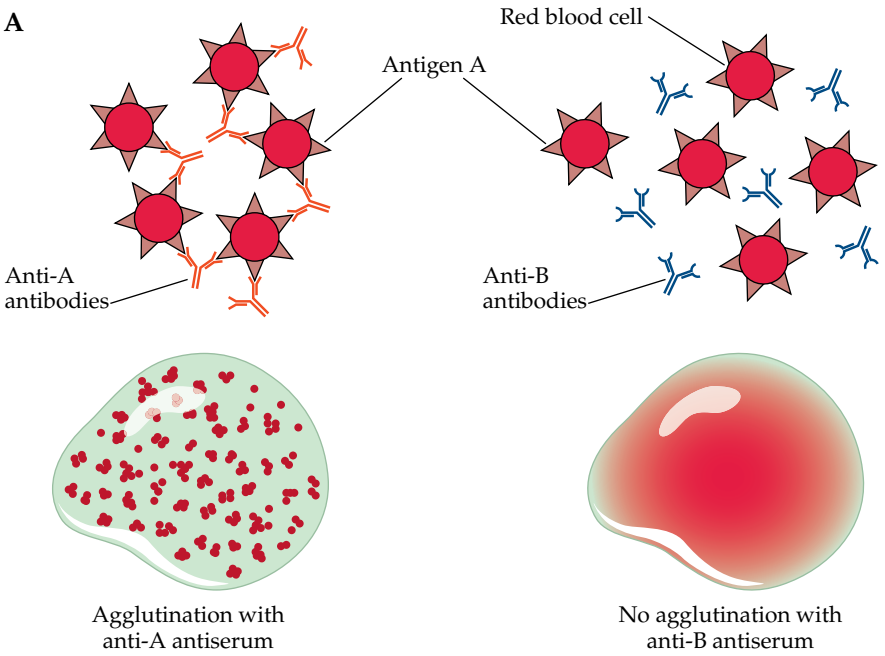
Many monoclonal antibodies have been developed for use as immunodiagnostic agents for a variety of compounds and pathogenic organisms (Table 4.1). As an alternative to the production of monoclonal antibodies in hybridoma cells in culture, monoclonal antibodies and parts of antibodies (Fab or Fv fragments) directed against a target antigen may be selected and produced in *Escherichia coli* (see chapter 5).

Table 4.1 Targets for diagnostic monoclonal antibodies

Polypeptide hormones
Chorionic gonadotropin
Growth hormone
Luteinizing hormone
Follicle-stimulating hormone
Thyroid-stimulating hormone
Prolactin
Tumor markers
Carcinoembryonic antigen
Prostate-specific antigen
Interleukin-2 receptor
Epidermal growth factor receptor
Cytokines
Interleukins 1–8
Colony-stimulating factor
Drug monitoring
Theophylline
Gentamicin
Cyclosporin
Miscellaneous targets
Thyroxine
Vitamin B_{12}
Ferritin
Fibrin degradation products
Tau protein
Infectious disease
Chlamydia
Herpes
Rubella
Hepatitis B
Legionella
Human immunodeficiency virus

Agglutination

Agglutination is a simple, inexpensive, rapid, and highly specific immunological test that is widely performed in diagnostic laboratories. For example, it is often used for human blood typing based on the presence of specific antigens on the surface of red blood cells, which vary among individuals. Antiserum containing antibodies against either A or B surface antigen is mixed with red blood cells. Clumping (agglutination) indicates the presence of the antigen (Fig. 4.4A). Some individuals produce only the A antigen (blood type A), which agglutinates with anti-A antiserum but not with anti-B antiserum (Fig. 4.4B). Blood type B individuals produce only the B antigen, which reacts with anti-B antiserum and not with anti-A antiserum. Others carry both antigens (blood type AB) or neither (blood type O). Blood samples from individuals with blood type O, the most common blood type, do not produce a positive agglutination result with either antiserum. This test, referred to as hemagglutination, is important for determining which blood to use in transfusions, as antibodies naturally produced against nonself red blood cell antigens would destroy the introduced blood. In some agglutination tests, the antigen or antibody employed to detect either a specific antibody or antigen, respectively, in patient samples is coated on the surface of small latex beads.



B

	Agglutination with serum	
Blood type	Anti-A	Anti-B
A	+	−
B	−	+
AB	+	+
O	−	−

Figure 4.4 Hemagglutination test to determine blood type. **(A)** Antiserum containing antibodies against A or B surface antigens is mixed with red blood cells carrying A surface antigens. Clumping (agglutination) in the presence of anti-A antiserum but not with anti-B antiserum indicates blood type A. **(B)** Agglutination responses for different blood types.

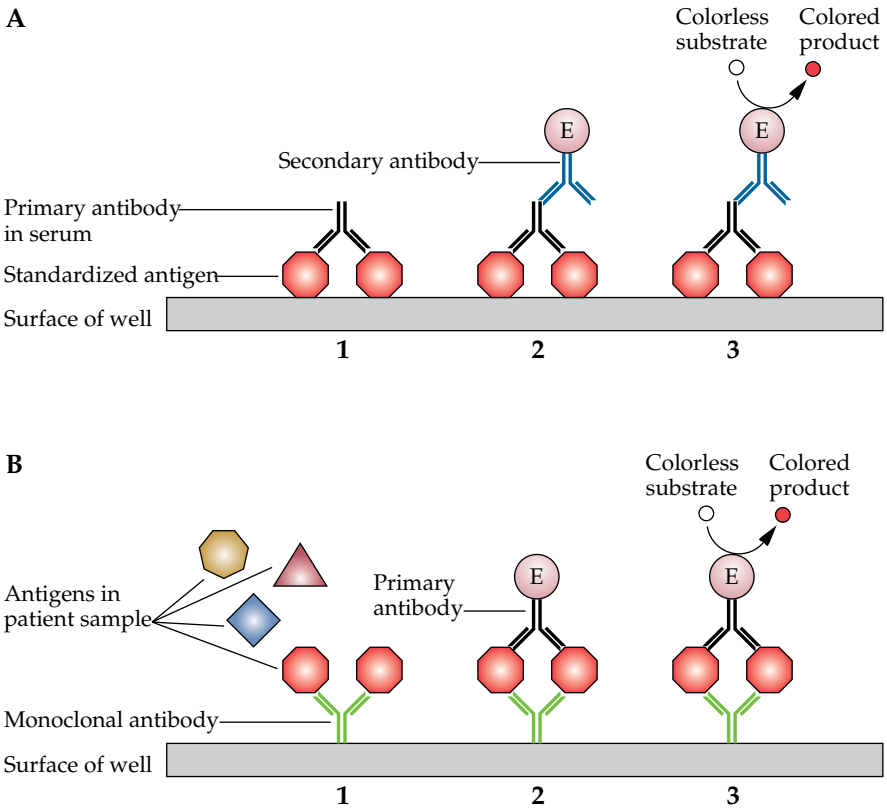
Enzyme-Linked Immunosorbent Assays

Enzyme-linked immunosorbent assays (ELISAs) are widely used for diagnostic detection. An ELISA measures antigens or antibodies produced against an antigen in a sample such as blood, urine, or tissues. It is based on the specific and high-affinity interaction between an antibody and an antigen and is a sensitive assay that can be used for rapid detection on a large scale (i.e., high-throughput assays). Detection relies on the activity of an enzyme that is covalently bound to an antibody employed in the assay. An indirect ELISA can detect the presence of specific antibodies in patient serum that indicates an immune response to the presence of a particular protein or pathogen. A sandwich ELISA detects the presence of a specific antigen in a sample and hence is sometimes referred to as an antigen capture assay.

In an indirect ELISA used for diagnostic purposes, a standardized antigen is bound to a solid support, usually the surface of a well in a

microtiter plate (Fig. 4.5A). A patient's serum sample is applied to the well, and specific (primary) antibodies in the serum bind to the immobilized antigens. A secondary antibody that binds specifically to the primary antibody is applied. When human serum is tested for the presence of a specific antibody, the secondary antibody is an anti-human immunoglobulin antibody that was raised in another animal, for example, a goat, by injecting the animal with human immunoglobulin. The secondary antibody is covalently bound (conjugated) to an enzyme such as alkaline phosphatase or horseradish peroxidase that catalyzes the conversion of a colorless substrate into a colored product (Fig. 4.6). A colorless substrate is applied, and the formation of a colored product by the enzyme indicates the presence of the secondary antibody and, hence, the primary antibody–antigen complex.

Figure 4.5 ELISAs. **(A)** Indirect ELISA. A specific antigen is immobilized on the surface of a well in a microtiter plate. (1) Serum is applied, and primary antibodies present in the serum bind to the antigen. Unbound antibodies are removed by a washing step. (2) A secondary antibody that binds specifically to the primary antibody is applied. (3) The secondary antibody is conjugated to an enzyme (E) that catalyzes the conversion of a colorless substrate into a colored product. The formation of a colored product indicates the presence of antibodies directed against the antigen in the serum sample. **(B)** Sandwich ELISA. Monoclonal antibodies specific for a target antigen are bound to the surface of a well in a microtiter plate. (1) A sample is added to the wells, and specific antigens present in the sample bind to the immobilized antibody. A wash step removes any unbound molecules. (2) A primary antibody is applied to detect the presence of bound antigen. (3) The primary antibody is conjugated to an enzyme that catalyzes the formation of a colored product that indicates the presence of the target antigen.



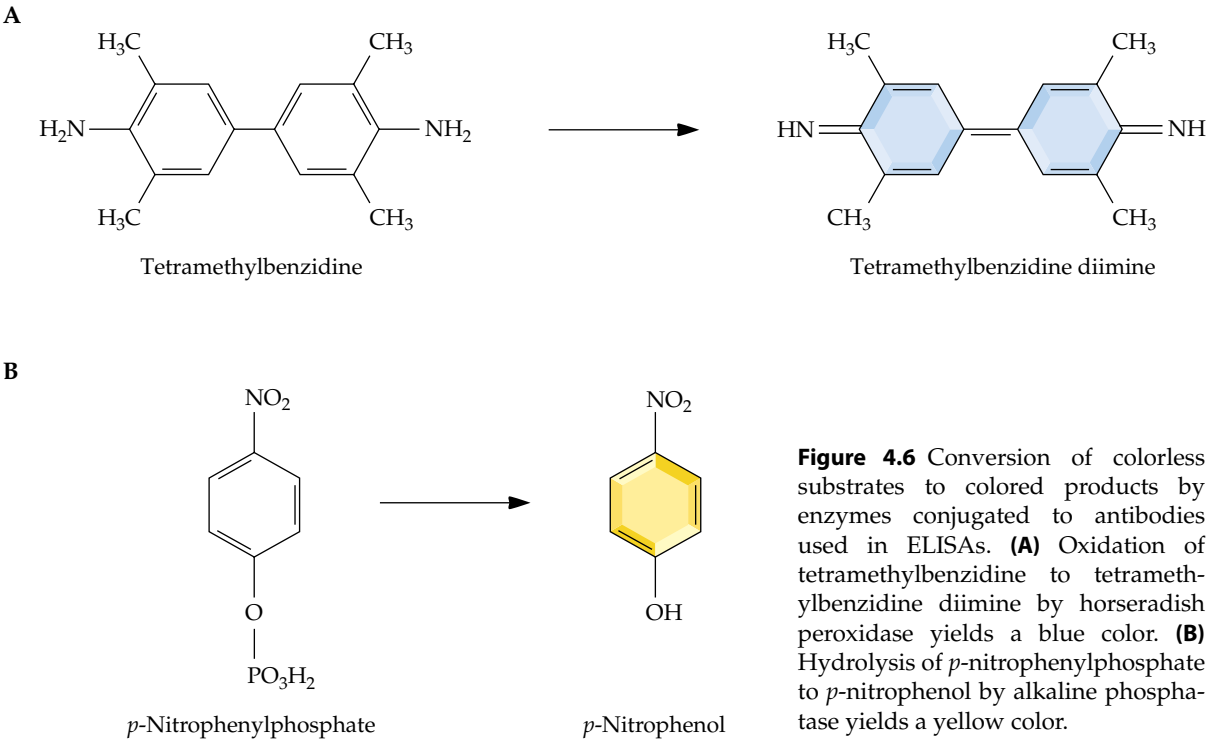


Figure 4.6 Conversion of colorless substrates to colored products by enzymes conjugated to antibodies used in ELISAs. **(A)** Oxidation of tetramethylbenzidine to tetramethylbenzidine diimine by horseradish peroxidase yields a blue color. **(B)** Hydrolysis of *p*-nitrophenylphosphate to *p*-nitrophenol by alkaline phosphatase yields a yellow color.

In contrast to an indirect ELISA, a sandwich ELISA directly detects a particular antigen in a complex sample. To capture the antigen, a monoclonal antibody that is specific for the target antigen is first bound to the surface of a microtiter plate (Fig. 4.5B). The sample is then added to a well; if the specific antigen is present in the sample, it binds to the immobilized antibody. A labeled primary antibody is added to detect the presence of bound antigen. The labeled antibody may detect a different epitope on the same antigen.

Detecting Pregnancy by a Sandwich ELISA

A sandwich ELISA is the basis for many diagnostic tests. In one home pregnancy test, the human chorionic gonadotrophin (hCG) protein produced during pregnancy by the developing placenta is detected in urine. Monoclonal and polyclonal antibodies specific for hCG are deposited on different regions of a membrane in a lateral flow device (Fig. 4.7). Urine is applied to one end of the membrane and is drawn through the membrane by capillary action to the region containing the anti-hCG monoclonal antibodies. If hCG molecules are present, they bind to the monoclonal antibodies. The antibodies with bound hCG are carried with the urine through the membrane to the site of the immobilized anti-hCG polyclonal antibodies. The polyclonal antibodies capture the hCG–monoclonal antibody complexes. Monoclonal antibodies that are not bound to hCG are carried away with the urine by capillary action. The monoclonal antibodies are conjugated to an enzyme that reacts with dye molecules contained in the region with the polyclonal antibodies. Formation of a colored product indicates a positive pregnancy test result. Farther downstream in the membrane, a control region contains anti-IgG antibodies that bind to the

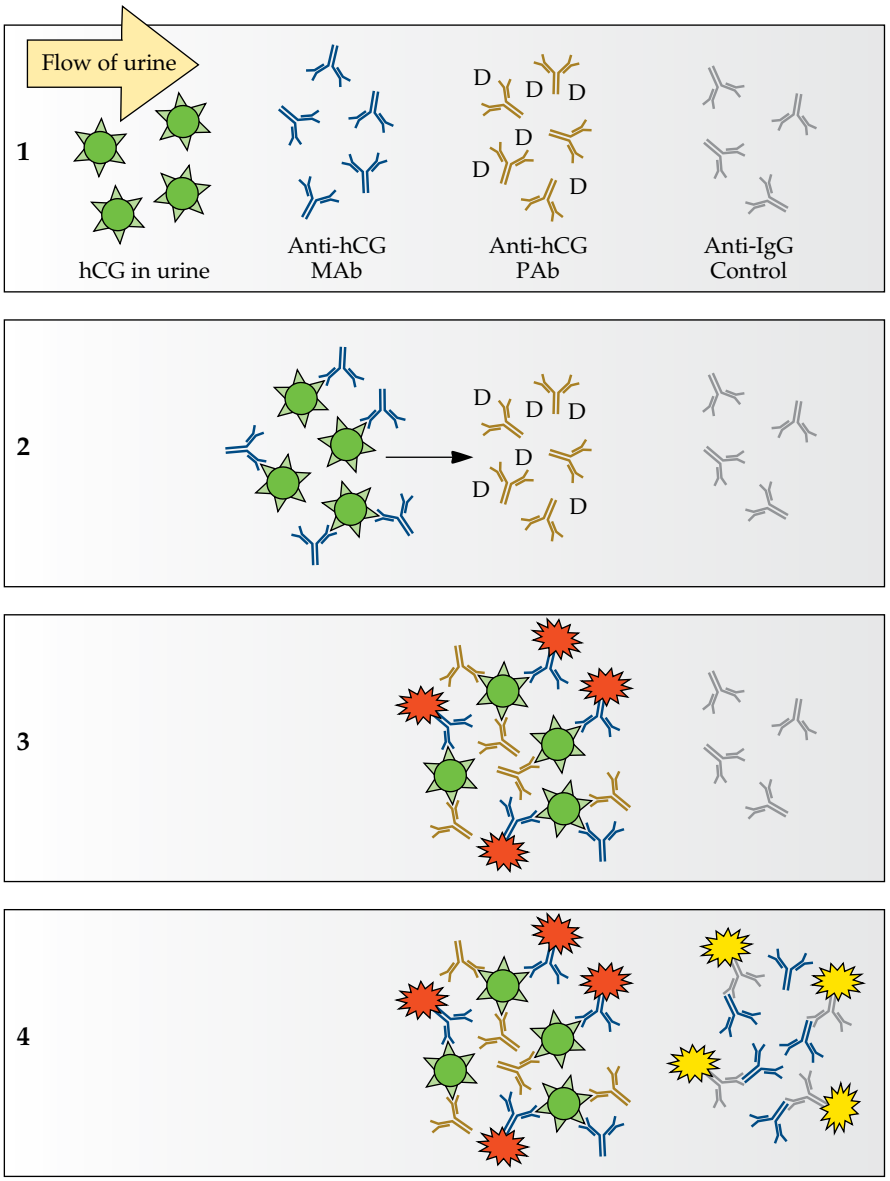


Figure 4.7 Pregnancy test using a sandwich ELISA. (1) Urine is applied to one end of the membrane. (2) Urine is drawn down the membrane by capillary action, and hCG molecules in the urine of a pregnant woman bind to the anti-hCG monoclonal antibodies (MAb). (3) The hCG molecules bound to the MAbs are subsequently carried with the flow of urine and are captured by the immobilized anti-hCG polyclonal antibodies (PAb). The MAbs are conjugated to an enzyme that reacts with dye molecules (D) to form a colored product (red star) that indicates a positive test result. (4) A control region contains anti-IgG antibodies that bind to the MAb whether or not hCG is present and produce a colored product (yellow star) to indicate that the test reagents are functional.

monoclonal antibodies whether or not hCG is present. A colored product in the control region indicates that the test components are functional. Timing of the test is important for accuracy because the hCG protein is produced only after implantation of the embryo in the uterine wall, which occurs 6 to

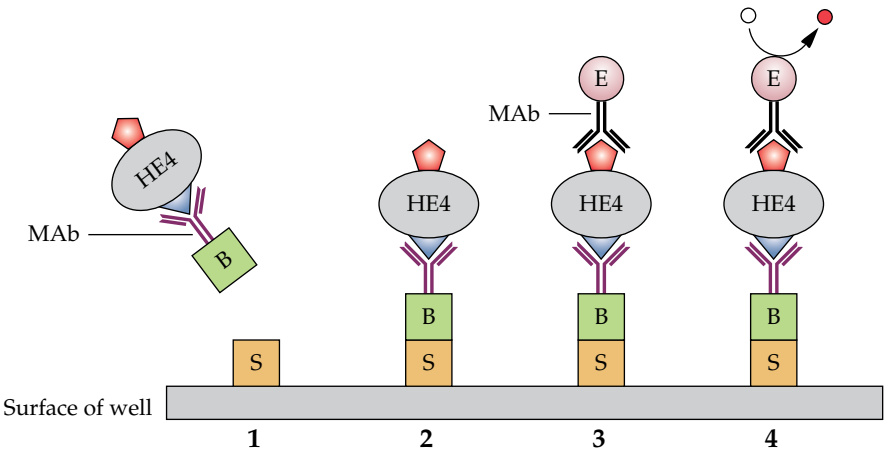
12 days after fertilization, after which, protein levels continue to rise during the first 20 weeks of pregnancy. False-negative results may occur if the test is performed too early in pregnancy.

Measuring Disease-Associated Proteins by Sandwich ELISA

Diagnosis of ovarian cancer. Ovarian cancer is a devastating disease that kills over 15,000 women a year in the United States. More than 22,000 new cases are diagnosed each year, most at an advanced stage when the survival rate is less than 20%. A useful biomarker for ovarian cancer is a protein that is secreted specifically by ovarian tumors, and not by normal tissue or other tumor types, and that can be detected at low levels in the blood or urine, enabling early, and noninvasive, detection. Using cDNA microarrays and quantitative polymerase chain reaction (PCR) (see chapter 2), human epididymis protein 4 (HE4) was identified as a biomarker that was expressed in ovarian carcinomas but not normal tissue or benign ovarian tumors. Mouse monoclonal antibodies were generated against two different HE4 epitopes and used to develop an ELISA. In an initial blinded study (researchers were not informed of the source of the patient samples), HE4 was found to be elevated in sera from patients with early and late-stage ovarian cancer compared to sera from healthy women and from women with benign ovarian tumors.

The U.S. FDA has approved the use of an HE4 sandwich ELISA to monitor the recurrence or progression of ovarian cancer in women who are being treated for the disease. Patient samples are incubated with monoclonal antibodies generated against one HE4 epitope. The anti-HE4 monoclonal antibodies were attached to a molecule of biotin (i.e., biotinylated) that has a very high affinity for the protein streptavidin, which coats the wells of a microtiter plate (Fig. 4.8). A second monoclonal antibody that binds to a different HE4 epitope and is conjugated to the enzyme horseradish peroxidase is added. After production of a colored product by the

Figure 4.8 Sandwich ELISA to monitor progression or recurrence of ovarian cancer. (1) Patient samples are incubated with biotinylated (B) monoclonal antibodies (brown MAb) generated against one HE4 epitope (blue). (2) Biotin, together with the MAb–HE4 complex, binds with high affinity to streptavidin (S) that coats the wells. (3) A second antibody (black MAb) that binds to a different HE4 epitope (red) is added, and after generation of a colored product by horseradish peroxidase (E) conjugated to the antibody (4), HE4 is quantified spectrophotometrically.



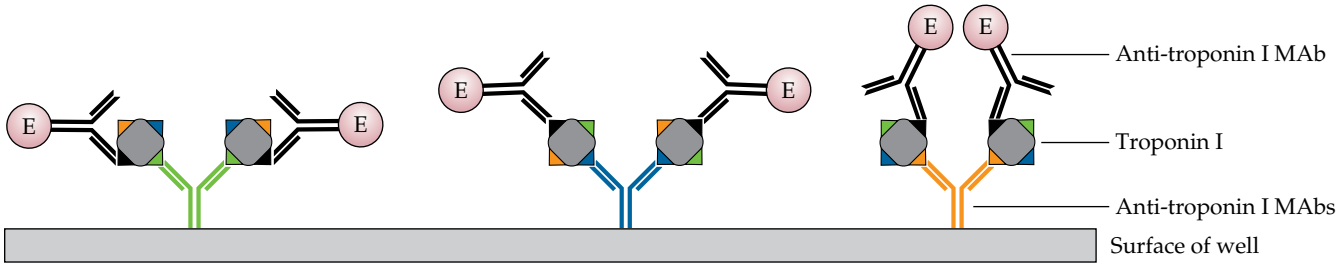
horseradish peroxidase, HE4 is quantified spectrophotometrically. Using this assay, HE4 was found to be present at elevated levels in more than 75% of patients with ovarian cancer, compared to 5% of healthy women and 13% of individuals with other nonmalignant conditions. An increase in HE4 levels of more than 25% is considered significant to suggest recurrence or disease progression, while a decrease of this magnitude suggests a positive response to treatment.

Diagnosis of acute myocardial infarction. ELISAs have been developed to diagnose many cancers and are also used to detect protein biomarkers of other human diseases. Elevated levels of cardiac protein troponin I may be indicative of a myocardial infarction (heart attack) and can be measured rapidly in patients presenting to an emergency department with chest pain. Blood levels of troponin I are normally very low but rise quickly after the onset of acute myocardial infarction and are quantified by an ELISA that employs four monoclonal antibodies against different troponin I epitopes (Fig. 4.9). Three of the antibodies are immobilized in the wells of a microtiter plate to capture troponin I in patient serum samples. The fourth antibody is conjugated to horseradish peroxidase to detect the bound antigen following addition of tetramethylbenzidine (Fig. 4.6). Elevated troponin I levels may be symptomatic of other health problems (such as advanced renal failure or SARS-CoV-2 infection), and therefore a positive diagnosis for myocardial infarction includes evidence of low blood oxygen levels. Timely detection of a myocardial infarction is imperative for a good outcome, and very recently a diagnostic test was reported to detect elevated cardiac troponin I levels in saliva samples in 10 minutes (compared to about 2 hours for the blood test). The saliva test, which is in the early stages of development, includes steps to remove inhibitors from the saliva and to concentrate samples to increase sensitivity.

Diagnosing Autoimmune Diseases by an Indirect ELISA

Autoimmune diseases occur when the body's immune system does not recognize normal cellular molecules and structures as "self" but, rather, produces antibodies (autoantibodies) that destroy those targets. Several autoimmune diseases have been identified, including celiac disease, type I diabetes, lupus erythematosus, and rheumatoid arthritis. Indirect ELISAs have been

Figure 4.9 Sandwich ELISA to diagnose acute myocardial infarction. Three different antitroponin I monoclonal antibodies are immobilized in a microtiter plate. Each monoclonal antibody captures a specific epitope (colored triangles) of troponin I present in a patient's serum or saliva sample. Bound troponin I is detected by a monoclonal antibody (black) that recognizes a fourth epitope and is conjugated to an enzyme (E) for detection. Levels of troponin I increase at the onset of a heart attack.



developed to diagnose some autoimmune diseases. In these assays, autoantibodies produced against a self-protein are detected in patient blood samples.

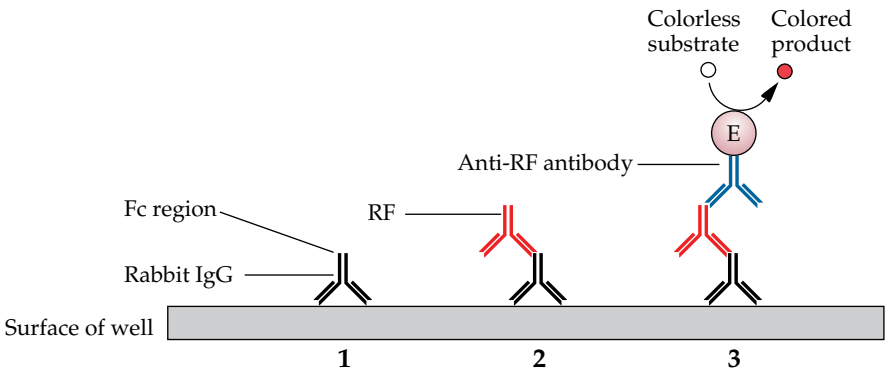
Rheumatoid arthritis is an autoimmune disease that results in chronic, systemic inflammation of the joints, mainly the synovial (flexible) joints. The synovial membrane that lines these joints secretes a fluid that lubricates the articulating bones. Initially, an inflammatory response occurs in the synovial membranes in the small joints of the hands and feet; it then occurs in larger joints, causing accumulation of excess fluid and damage to the joints. Systemic inflammation may damage organs such as the heart and lungs. The disease affects about 1% of the population of the United States.

Rheumatoid factor is an autoantibody that targets the Fc region of IgG antibodies and contributes to rheumatoid arthritis. It is commonly used as a diagnostic biomarker to differentiate rheumatoid arthritis from other forms of arthritis and other inflammatory conditions. Several indirect ELISAs have been developed to detect the presence of rheumatoid factor in patient blood samples. In one test, IgG molecules from a rabbit are the standardized antigen bound to the surface of a multiwell plate, and diluted patient sera are applied to the wells (Fig. 4.10). After incubation to allow binding of rheumatoid factor in the serum to the rabbit IgG molecules and washing to remove unbound molecules, an anti-human IgG antibody is added. The anti-human IgG antibody specifically targets rheumatoid factor (and not rabbit IgG) and is conjugated to an enzyme such as horseradish peroxidase for quantitative colorimetric detection. Other ELISAs have been developed that target rheumatoid factor IgM and IgA that can assist in accurate diagnosis of rheumatoid arthritis. Early diagnosis is important to prevent irreversible joint damage, as rheumatoid arthritis may be managed in the early stages by administering antirheumatic drugs.

Detecting Infectious Diseases by an Indirect ELISA

Indirect ELISAs are commonly employed to diagnose a broad range of infectious diseases caused by viruses, bacteria, fungi, and parasites with high sensitivity and specificity. Measurement of pathogen-specific antibodies is useful to confirm an infection in patients who are experiencing

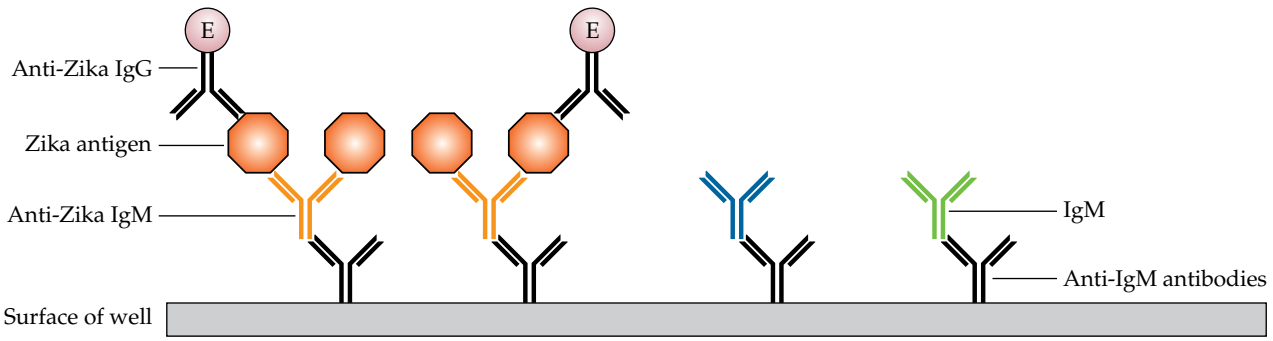
Figure 4.10 Indirect ELISA for diagnosis of the autoimmune disease rheumatoid arthritis. (1) Rabbit IgG molecules are bound to the surface of a microtiter plate, and diluted patient serum is applied. (2) If rheumatoid factor (RF) is present in the serum, it binds to the Fc region of the rabbit IgG molecules. (3) After a washing step to remove unbound molecules, an anti-RF antibody conjugated to an enzyme (E) is applied for colorimetric detection.



characteristic symptoms, to detect past exposures after the pathogen has been cleared, as antibodies generally last much longer than molecules from the infecting agent, or to identify asymptomatic infections during outbreaks to more accurately assess infection rates.

Detection of a viral infection. In 2019, the U.S. FDA approved a diagnostic test to detect antibodies produced by the immune system in response to Zika virus infection. Zika virus is primarily transmitted to humans through the bite of an infected *Aedes* mosquito and may also be sexually transmitted by infected individuals. Although many infected individuals are asymptomatic, infections during pregnancy have been linked to microcephaly (a condition where an infant's head is much smaller than expected) in infants. An outbreak first appeared in Brazil in 2015, and limited mosquito-borne transmission of Zika virus was reported in the United States in 2016. The indirect ELISA detects IgM antibodies (indicative of an early immune response) specifically against Zika virus envelope glycoproteins and not those of other similar flaviviruses. Blood samples from patients with clinical symptoms of infection (rash, fever, fatigue, joint pain) or with other epidemiological risk factors (e.g., travel to regions with outbreaks) are applied to wells of a microtiter plate coated with human anti-IgM polyclonal antibodies to capture any IgM antibodies present in the samples (Fig. 4.11). Next, the Zika virus antigen consisting of viral envelope glycoproteins is applied to specifically detect bound anti-Zika IgM. Finally, a secondary, Zika virus-specific monoclonal antibody conjugated to horseradish peroxidase is added. Incubation with tetramethylbenzidine yields a blue-colored product (Fig. 4.6) that indicates patient samples containing anti-Zika IgM. If IgM antibodies targeting Zika virus antigens are not present, the added antigen and secondary antibody do not bind to the wells and are washed away. The timing of the diagnostic test is important because anti-Zika IgM antibodies may take up to 4 days after the onset of symptoms to develop and persist for about 12 weeks. Although a treatment for Zika virus infection is not currently available, a positive test result can inform pregnant women of the potential risk of neurological damage to the fetus.

Figure 4.11 ELISA to detect IgM antibodies indicative of a Zika virus infection. Anti-IgM polyclonal antibodies are immobilized in the wells of a microtiter plate to capture any IgM molecules present in a patient sample. To distinguish IgM elicited by Zika virus infection from other IgMs in the sample, purified Zika virus envelope glycoproteins (Zika antigen) are added to the wells. Binding of the Zika antigen is detected using anti-Zika IgG conjugated to an enzyme (E) for production of a colored product.



Measuring neutralizing antibodies. During an infection, antibodies may be elicited against several different epitopes presented on the surface of a pathogen. A subset of the antibodies binds to a region of a surface protein that mediates adhesion of the pathogen to a host cell receptor during the initial stages of infection. These antibodies block the interaction between the pathogen and host cell and thereby prevent infection; that is, they induce sterilizing immunity. They are known as neutralizing antibodies, and their measurement in the sera of immunized individuals can be used to evaluate the efficacy of a vaccine. Conventional tests for neutralizing antibodies determine the ability of patient sera to block pathogen infection of cultured host cells. However, such neutralization assays require handling of live pathogens, which is time-consuming, often requires specialized biocontainment facilities, and may not be practical during large-scale pandemics. An indirect ELISA was developed to quantify neutralizing antibodies produced against the SARS-CoV-2 virus that obviates the need for live virus or host cells.

The receptor-binding domain of the SARS-CoV-2 spike protein binds to the angiotensin-converting enzyme 2 (ACE2) receptor protein on the surface of human cells (Fig. 4.12A). Neutralizing antibodies produced against the virus bind to an epitope on the receptor-binding domain and may be quantified based on their ability to inhibit the interaction between the receptor-binding domain and the ACE2 receptor. In one type of ELISA, purified receptor-binding domain protein conjugated to horseradish peroxidase bound to ACE2 receptors immobilized on a solid support, and the interaction was measured by the conversion of tetramethylbenzidine to a blue-colored product (Fig. 4.12B). However, when the receptor-binding domain conjugate was first incubated with patient serum, neutralizing antibodies present in the serum blocked the interaction between the receptor-binding domain and the immobilized ACE2 receptors (Fig. 4.12C). Unbound receptor-binding domain-antibody complexes were washed away, and a decrease in colored product was measured. Dose-dependent inhibition was observed following incubation with sera from patients infected with the SARS-CoV-2 but not with sera from healthy controls. Detection of neutralizing antibodies elicited by a novel pathogen such as SARS-CoV-2 can be used to assess the level and longevity of protective immunity in recovered patients and in individuals and populations following immunization with newly developed vaccines.

Protein Arrays To Detect Polygenic Diseases

An ELISA typically measures a single target protein; however, for some diagnoses, it may be more informative to measure multiple target proteins in a single assay. Analysis of proteomes is commonly used in research to identify and quantify protein changes in diseased tissue versus normal tissue and is useful for diagnosis of polygenic diseases (resulting from mutations in more than one gene) such as breast cancer, Alzheimer's disease, type I diabetes, and cardiovascular disease. Protein microarrays are multiplex immunoassays that can detect multiple biomarkers in a clinical sample (chapter 2). Biomarkers may be subsets of proteins in complex clinical samples such as biopsied tumor tissue. Currently, commercially available protein microarrays are most commonly used to detect

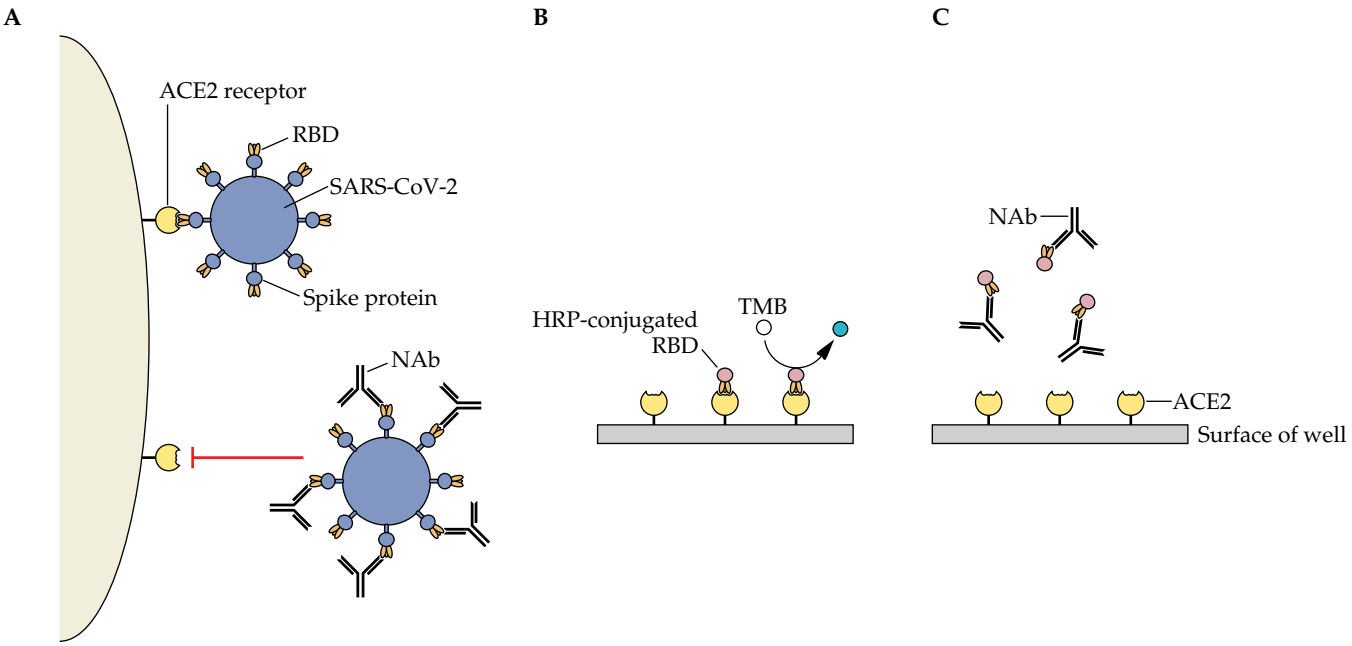


Figure 4.12 ELISA to quantify SARS-CoV-2 neutralizing antibodies present in serum. **(A)** The envelope of the SARS-CoV-2 virus contains a spike protein with a receptor-binding domain (RBD) that binds to angiotensin-converting enzyme 2 (ACE2) receptor protein on the surface of human cells during the initial stage of infection. During an infection or following immunization, neutralizing antibodies (NAbs) are produced that bind to the RBD and prevent interaction between the spike protein and ACE2. **(B)** Purified RBD binds to ACE2 immobilized in the wells of a microtiter plate and is detected by conversion of tetramethylbenzidine (TMB) to a blue product by horseradish peroxidase (HRP), which is conjugated to RBD. **(C)** Levels of NAbs in a patient's serum are measured by their ability to block binding of RBD to ACE2. Data from Tan et al., *Nat. Biotechnol*. **38:**1073–1078, 2020.

antibodies in serum from patients suffering from allergies, autoimmune diseases, or infections.

Up to 40% of the global population suffers from type I allergies, which are IgE-mediated immediate hypersensitivity reactions, such as asthma, hay fever, and eczema. In susceptible individuals, the first exposure to an allergen elicits the production of high levels of IgE antibodies that bind to mast cells. In this way, the individual becomes sensitized to the allergen. On subsequent exposures, the allergen reacts with mast cell-bound IgE and causes bridging between adjacent IgE molecules, thereby stimulating the release of inflammatory mediators such as histamine, leukotrienes, prostaglandins, and tumor necrosis factor. These chemicals stimulate smooth muscle contraction and dilation of capillaries that cause edema (swelling), itching, and development of a rash. Severe reactions can lead to systemic anaphylaxis that may be fatal if not treated quickly. Diagnostic allergen microarrays detect IgEs produced against common allergens, for example, pollen, food proteins, and molds. Purified allergens are arrayed in triplicate on a solid support and probed with sera from allergic patients (Fig. 4.13). Specific IgEs present in the serum bind to an immobilized allergen and are retained on the surface of the array. Bound IgEs are detected

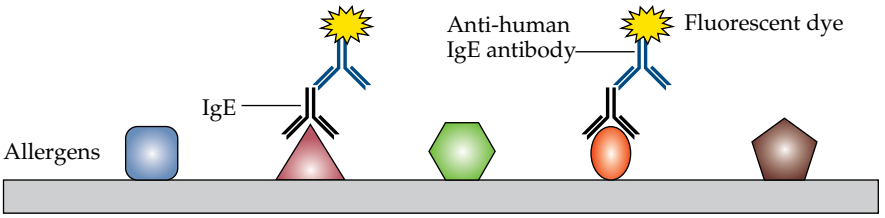
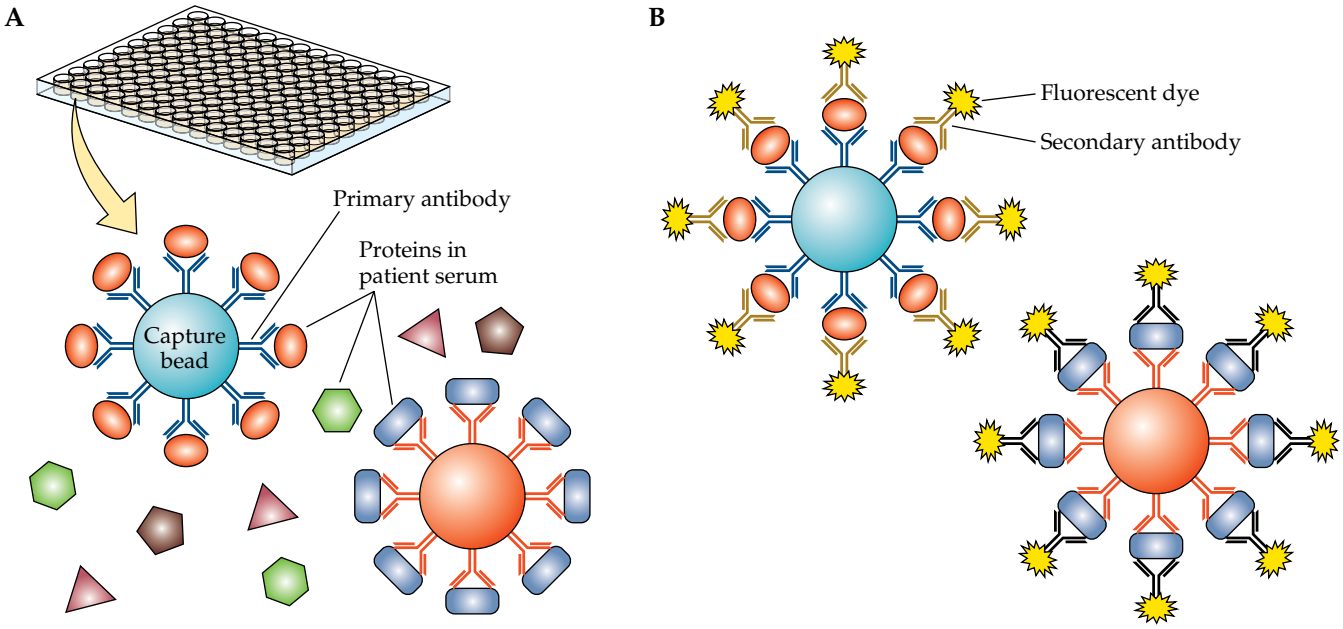


Figure 4.13 Diagnostic allergen microarray. Purified allergens are arrayed on a solid support and probed with serum from an allergic patient. Specific IgE present in the serum binds to an immobilized allergen and is detected using an anti-human IgE antibody conjugated to a fluorescent dye.

using an anti-human IgE antibody, often labeled with a fluorescent dye. The array format enables the detection of hundreds of different allergens in a single assay.

Another type of protein microarray measures levels of 50 different protein biomarkers of heart disease in a patient serum sample. Increased or decreased levels of these proteins are associated with increased risk of heart disease. For example, adiponectin is a protein hormone secreted by adipocytes (fat cells) that regulates glucose and lipid metabolism. Reduced

Figure 4.14 Liquid bead assay to detect protein biomarkers associated with increased risk of cardiovascular disease. **(A)** Patient serum is applied to a well of a microtiter plate containing capture beads bound to primary antibodies directed against all of the protein biomarkers. Capture beads of different colors distinguish primary antibodies that bind to different protein biomarkers. Target proteins in the sample bind to their cognate antibodies. **(B)** For detection of the bound proteins, secondary antibodies conjugated to a fluorescent molecule are added to the well. Each bead, and any captured proteins, is analyzed by passing by a laser that measures fluorescence; fluorescence intensity corresponds to the amount of target protein present. The presence of specific protein biomarkers is determined by the color of the capture bead to which they are bound.



levels of this hormone are found in individuals with increased risk of myocardial infarction. Elevated levels of the protein plasminogen activation inhibitor 1 are present in patients with myocardial infarction. This protein inhibits tissue plasminogen activator, a key enzyme in the breakdown of fibrin (fibrinolysis) in blood clots that is synthesized by endothelial cells, platelets, and hepatocytes. A liquid bead array was developed to quantify levels of these and other protein biomarkers in the serum samples. In this system, all of the different protein biomarkers are detected in a single well of a multiwell plate (Fig. 4.14), which enables rapid analysis of many patient samples simultaneously. Primary antibodies directed against each protein biomarker are attached to tiny capture beads that may be distinguished by either size, color, or shape. A serum sample is applied to a well, and specific proteins in the sample bind to their cognate antibody. For detection of the bound proteins, secondary antibodies conjugated to a fluorescent molecule are added to the well. Each bead, and any captured proteins, is analyzed by passing by a laser that measures fluorescence emitted from bound secondary antibodies and distinguishes the different capture beads by their size, color, or shape. Antibody- or other ligand-conjugated microbeads in liquid or arrayed on a solid support provide a rapid and cost-effective means of high-throughput screening for multiple disease-associated proteins. In addition to diagnosis of heart disease, they have been used for early detection of multiple sclerosis and Alzheimer's disease, measurement of cytokines associated with inflammatory diseases, diagnosis of hepatitis B virus infection, and detection of breast, ovarian, colorectal, prostate, and melanoma tumor markers.

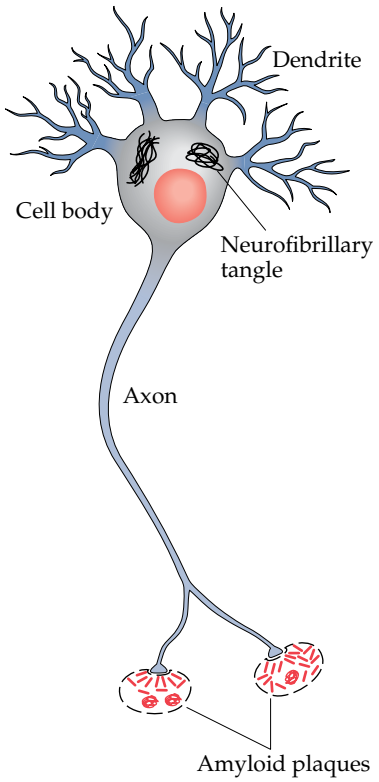
Immunoassays for Protein Conformation-Specific Disorders

Several human neurological disorders arise as a consequence of protein misfolding that leads to protein aggregation and cell death. Parkinson's disease, Alzheimer's disease, Huntington's disease, and prion diseases are examples of protein conformational disorders. Alzheimer's disease is a degenerative brain disorder that is characterized by the progressive loss of abstract thinking and memory, personality change, language disturbances, and a slowing of physical capabilities. Clinical diagnosis of Alzheimer's disease is poor, although 1% of the population between 60 and 65 years old and 30% of the population over 80 years old may develop it. Two hallmarks of Alzheimer's disease found in the brain are (i) neurofibrillary tangles of the cytoskeletal protein tau that accumulate within nerve cells and (ii) dense extracellular aggregates of insoluble proteins called amyloid plaques that develop at the ends of inflamed nerves (Fig. 4.15).

The principal protein of an amyloid plaque is a small protein called Aβ (β-amyloid protein). The Aβ protein ranges in length from 39 to 42 amino acid residues; the Aβ40 and Aβ42 forms are the main variants. All Aβ proteins are derived from the β-amyloid precursor protein (APP) by proteolytic cleavage. Abnormal cleavage of APP results in production of Aβ40 and Aβ42 and alters protein folding, causing exposed regions of the protein to self-interact; hence, the proteins aggregate and form amyloid plaques.

Diagnosis of Alzheimer's disease using immunological methods may exploit the development of antibodies that differentiate between the

Figure 4.15 Neurofibrillary tangles and amyloid plaques associated with neurons in the brain are hallmarks of Alzheimer's disease.



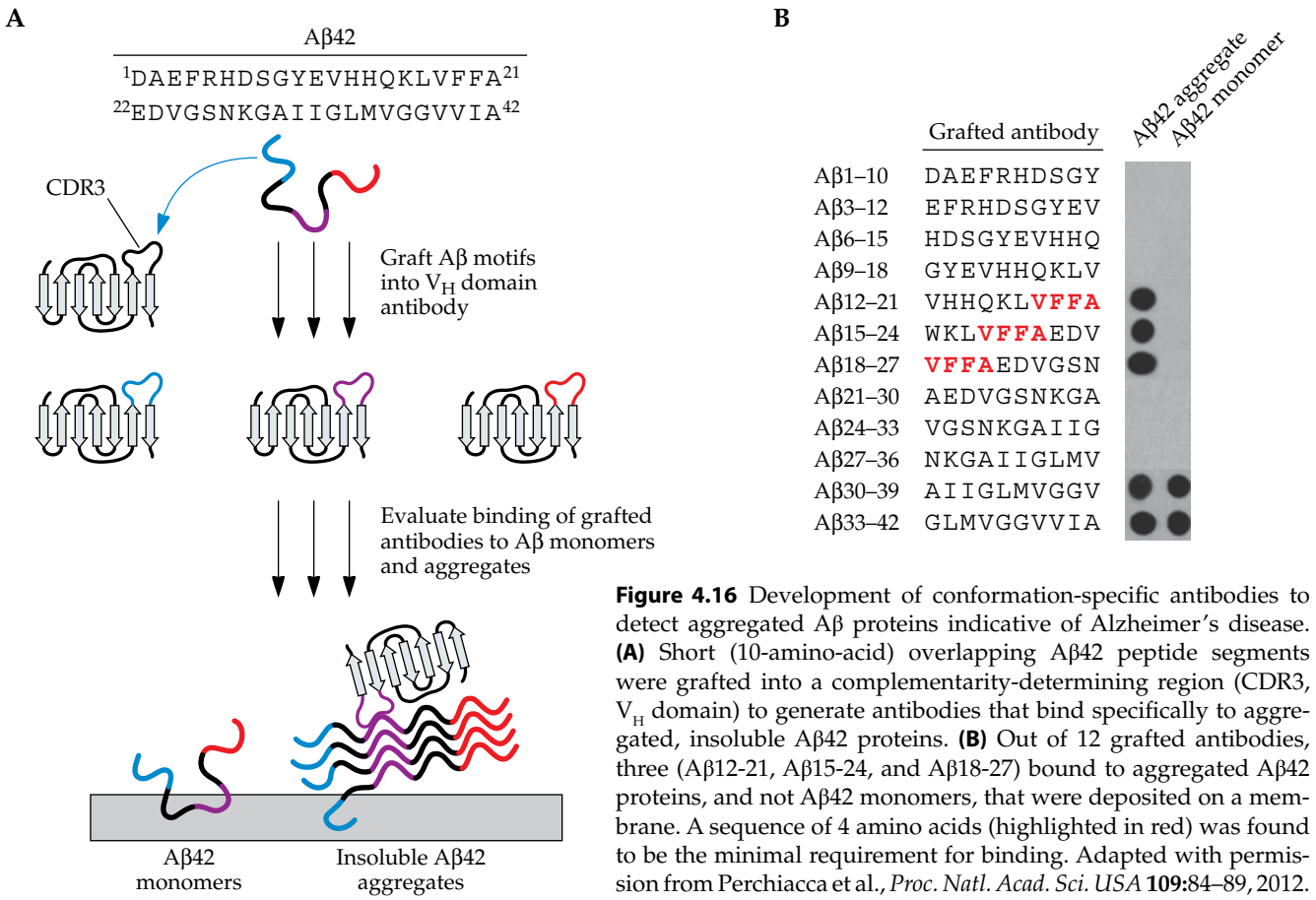


Figure 4.16 Development of conformation-specific antibodies to detect aggregated Aβ proteins indicative of Alzheimer's disease. **(A)** Short (10-amino-acid) overlapping Aβ42 peptide segments were grafted into a complementarity-determining region (CDR3, V_H domain) to generate antibodies that bind specifically to aggregated, insoluble Aβ42 proteins. **(B)** Out of 12 grafted antibodies, three (Aβ12-21, Aβ15-24, and Aβ18-27) bound to aggregated Aβ42 proteins, and not Aβ42 monomers, that were deposited on a membrane. A sequence of 4 amino acids (highlighted in red) was found to be the minimal requirement for binding. Adapted with permission from Perchiacca et al., *Proc. Natl. Acad. Sci. USA* **109:**84–89, 2012.

conformations of the disease-associated Aβ proteins and normal APP. To develop conformation-specific antibodies, researchers used an approach that mimics protein aggregation during the disease process, that is, the self-interaction of the Aβ proteins. They grafted DNA fragments encoding short motifs (10 amino acids) from the Aβ42 protein into the coding sequence of a complementarity-determining region (CDR3, V_H domain) of an antibody to produce antibodies that would recognize aggregated Aβ42 via the grafted motif (Fig. 4.16A). Out of 12 antibodies, with different but overlapping Aβ42 sequences, 3 antibodies bound specifically to insoluble Aβ42 aggregates but not Aβ42 monomers (Fig. 4.16B). A motif of four amino acids, common to all three positive antibodies, was found to be the minimal sequence required for binding. To determine whether binding of the grafted antibodies is mediated by interactions between the Aβ motif on the antibody and the same motif in the aggregated Aβ proteins, each grafted antibody was bound to immobilized Aβ42 aggregates, and then a second antibody was applied. Binding of the second antibody was reduced only when it recognized a motif that overlapped with that grafted on the first antibody. This relatively simple approach may be used to generate conformation-specific antibodies against other proteins that not only aid in accurate disease diagnosis but also may be used in therapeutic strategies to target these proteins.

DNA-Based Diagnostic Approaches

DNA-based diagnostic tests detect the presence of specific nucleotide sequences, for example, mutations associated with a genetic disease or sequences indicative of the presence of a pathogen. They are highly sensitive and specific and can detect variations among individuals in a single nucleotide at a specific position in the genome (single nucleotide polymorphism; SNP). The ability to diagnose diseases in humans at the genetic level makes it possible to determine the cause of an illness and to predict whether individuals or their offspring are predisposed to the disease. Because a DNA-based test does not require expression of a gene, in contrast to the diagnostic detection of proteins, DNA analysis can be used for the identification of asymptomatic carriers of hereditary disorders, for prenatal diagnosis of serious genetic conditions, and for early diagnosis before the onset of symptoms. DNA sequence-specific diagnostic approaches include hybridization of a unique DNA probe to a complementary target sequence, target sequence amplification by PCR, and microarray analysis to detect multiple sequences in a single sample. Whole-genome sequencing can facilitate the development of strategies to prevent, diagnose, and treat diseases that are tailored to our individual genotypes.

Hybridization Probes

Hybridization is the formation of hydrogen bonds between two complementary strands of nucleic acids. A diagnostic test involving DNA hybridization utilizes a DNA probe to detect a complementary target DNA sequence. The probe is labeled with a reporter molecule that indicates hybridization between the target and probe DNA. Typically, the cells in a sample such as infected or biopsied tissue are lysed, and the genomic DNA in the lysed cells is denatured by treatment with a strong alkali to generate single-stranded target DNA. Then, an excess of labeled oligonucleotide probe is added under appropriate conditions of temperature and ionic strength to promote base-pairing between the probe and the target DNA. Unbound probe DNA is removed by washing, and hybridization is detected by measuring the activity of the reporter molecule attached to the probe. A hybridization probe used in a diagnostic assay must hybridize exclusively to the selected target nucleic acid sequence. False positives (i.e., detection of a signal in the absence of the target sequence) and false negatives (i.e., no detection of a signal when the target is present) severely reduce the utility of a diagnostic test and can have serious consequences, for a patient, for example. Oligonucleotide probes are usually less than 100 nucleotides in length, although they may be longer, and are labeled either with an enzyme that produces a color change when it acts on certain substrates or with a fluorescent dye.

Hybridization To Identify Bacteria

Many DNA diagnostic probes have been developed for the detection of various strains of bacteria, viruses, and parasites. For example, probes have been developed for the diagnosis of human bacterial infections caused by *Legionella pneumophila* (a causative agent of pneumonia), *Salmonella enterica* serovar Typhi (food poisoning), enterotoxigenic

Escherichia coli (gastroenteritis), and *Neisseria gonorrhoeae* (gonorrhea, a sexually transmitted infection). In principle, nearly all pathogenic organisms can be detected by this procedure.

The composition of microbes, in particular the balance of bacteria of the phyla *Firmicutes*, *Bacteroidetes*, and *Proteobacteria*, in our intestinal tract affects our health. Transient or chronic alterations (dysbiosis) in the microbial community are associated with intestinal disorders such as irritable bowel syndrome and inflammatory bowel diseases. One study aimed to develop a diagnostic test for intestinal dysbiosis using hybridization probes to profile bacteria in fecal samples. A total of 54 different oligonucleotide probes were designed to hybridize to the variable regions of the bacterial 16S rRNA gene (Fig. 4.17A), targeting bacteria from 10 bacterial taxonomic groups and 36 different genera with high specificity. The probe set was based on profiles compiled from intestinal microbiome research and optimized to distinguish between samples from healthy individuals and those with irritable bowel syndrome. The diagnostic test entails extraction of genomic DNA from fecal samples, PCR amplification of a 1,180-bp region of the 16S rRNA gene comprising seven variable regions using a pair of common primers, and then hybridization of the probes to the PCR products to generate a bacterial profile (Fig. 4.17B). The performance of the test was evaluated using 287 fecal samples collected from healthy volunteers and patients previously diagnosed with irritable bowel syndrome or inflammatory bowel disease based on clinical symptoms. The hybridization probe-based test identified 73% of the patients with irritable bowel syndrome and 74% of the patients with inflammatory bowel disease as dysbiotic. Patients with inflammatory bowel disease showed a greater degree of dysbiosis than those with irritable bowel syndrome, reflecting differences in the severity of the disorders. Interestingly, 16% of healthy controls had intestinal bacterial profiles characteristic of dysbiosis. Several bacterial groups, including bacilli, proteobacteria, actinobacteria, *Shigella/Escherichia*, and *Ruminococcus gnavus*, were more abundant in patients with irritable bowel syndrome than in healthy individuals, while proteobacteria, and particularly *Shigella/Escherichia*, were more abundant and *Faecalibacterium prausnitzii*, clostridia, *Bacteroides*, and *Prevotella* were less abundant in the group with inflammatory bowel disease (Fig. 4.17C); *F. prausnitzii* is known to reduce intestinal inflammation. Individual profiles changed very little over a weekly sampling period of 14 weeks. This easy-to-use test may help to diagnose bowel disorders and to monitor the success of therapeutic regimens such as fecal microbiota transplantation, although testing on a larger scale that includes individuals from different geographical regions is required to establish its broader efficacy.

Allele-Specific Hybridization

In addition to infectious diseases, hybridization probes are widely used to detect specific disease-associated alleles. Monogenic diseases are caused by mutations in a single gene; however, any one of several alterations to the normal nucleotide sequence of a gene may be responsible. This is exemplified by cystic fibrosis, a lethal autosomal recessive disorder that affects approximately 1 in every 3,000 to 4,000 live births in populations of European descent (the disorder is less common in other populations).

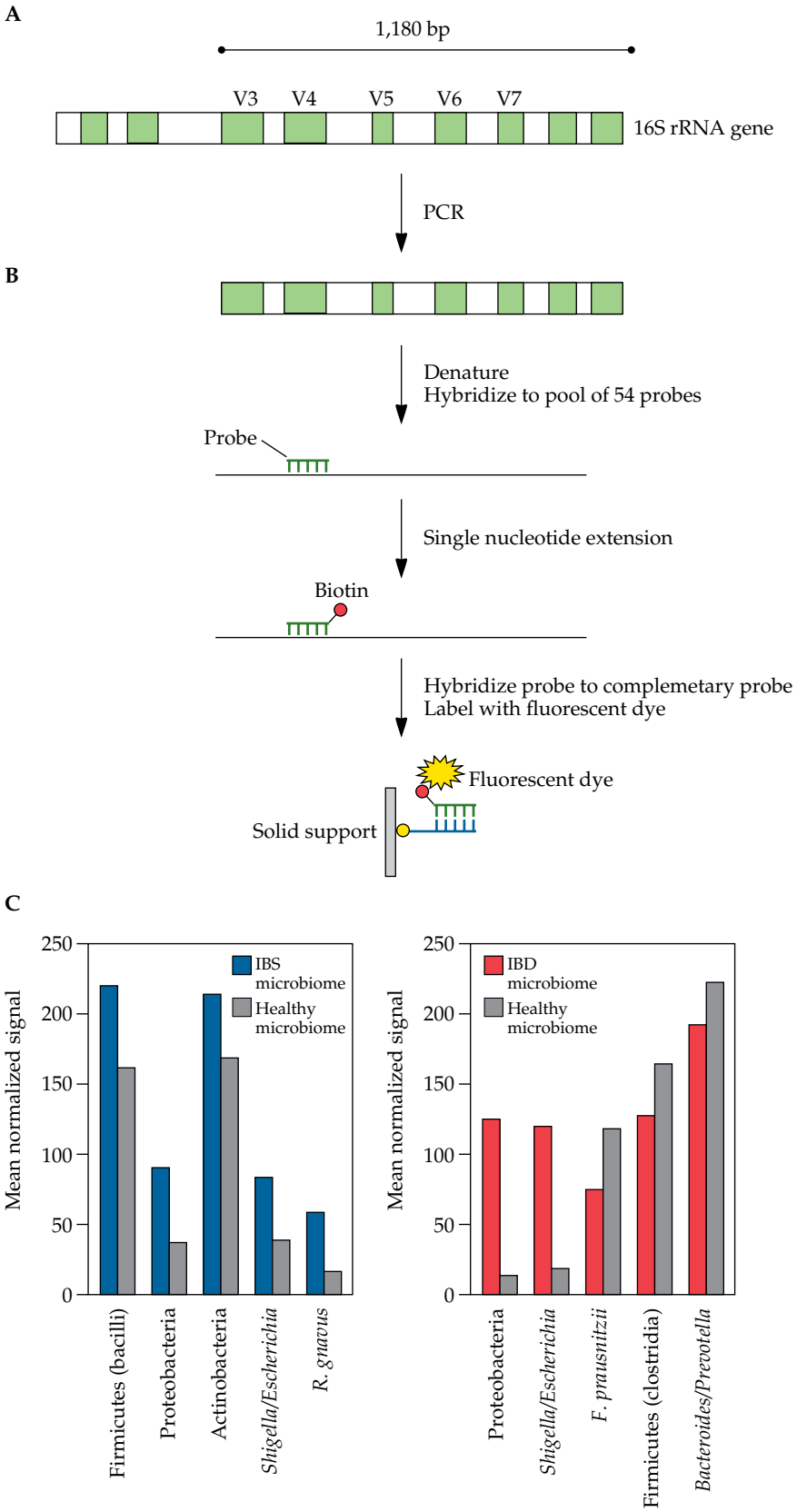


Figure 4.17 Diagnostic test for dysbiosis of intestinal microbiota. **(A)** The 1,400-bp 16S rRNA gene contains regions of conserved (white) and variable (green) sequence among bacteria. Variable regions V3 to V7 were targeted by 54 different oligonucleotide probes to assess the abundance of 10 taxonomic groups and 36 bacterial genera in the intestinal microbiome. **(B)** DNA was extracted from fecal samples, and a 1,180-bp region of the 16S rRNA gene containing variable regions V3 to V7 was amplified by PCR using primers that anneal to conserved regions. The probes were hybridized to target sequences in the PCR products (a single probe hybridized to V4 is shown in green) and were extended by a single dideoxyribonucleotide conjugated to biotin (red circle) by DNA polymerase. Probes were released from the target sequences and captured on a solid support by complementary oligonucleotides (blue) that were labeled with an optical barcode (yellow circle) for identification. A fluorescent dye (phycoerythrin) conjugated to streptavidin, which has a high affinity for biotin, was added to measure probe abundance. **(C)** The abundance of each probe, indicated by the mean normalized fluorescence signal intensity, reflects the abundance of each target bacterial group in the fecal samples from healthy individuals and individuals with irritable bowel syndrome (IBS) and inflammatory bowel disease (IBD). Data from Casén et al., *Aliment. Pharmacol. Ther.* **42:**71–83, 2015.

Mutations in a single gene, the cystic fibrosis transmembrane conductance regulator (CFTR) gene, result in defects in chloride ion transport. As a consequence, the mucus of lung and other mucosal tissues is thick and viscous, obstructing the respiratory, digestive, and reproduction system functions (see chapter 5). Every state in the United States now routinely screens newborns for cystic fibrosis. An ELISA is used to detect higher-than-normal levels of immunoreactive trypsinogen in blood. This protein is produced by the pancreas and leaks into the blood of infants with cystic fibrosis as a result of pancreatic damage. Premature babies and babies from stressful births may also have elevated levels of immunoreactive trypsinogen, and in these cases, genetic tests are employed for confirmation.

About 1,900 different mutations are reported to occur in the CFTR genes of patients with cystic fibrosis. Screening individuals who may be at risk for cystic fibrosis for such a large number of possible mutations is a daunting task. Fortunately, some of the mutations that cause cystic fibrosis are much more common than others (Table 4.2). The most common mutation is an in-frame deletion of three nucleotides in exon 10 of the CFTR gene that leads to loss of the amino acid phenylalanine at codon 508 (ΔF508). Over 90% of cystic fibrosis patients carry at least one ΔF508 allele, and nearly 50% of cystic fibrosis patients are individuals who are homozygous for ΔF508. It is estimated that about 160 different mutations account for 96% to 97% of cystic fibrosis alleles.

Allele-specific hybridization is commonly used in genetic tests for cystic fibrosis. With this technique, an individual's CFTR gene is amplified by PCR and then hybridized to labeled oligonucleotide probes for the mutant

Detection of Sickle-Cell β^S-Globin Allele by Hybridization with Synthetic Oligonucleotides

milestone

The development of techniques in the early 1980s for rapid, efficient, and relatively inexpensive chemical synthesis of DNA oligonucleotides opened up the possibility of using radiolabeled oligonucleotides as hybridization probes for the detection of a variety of DNAs, including mutations in human genes. At the time, very few human genes or even cDNAs had been isolated. Thus, it was not always easy to find a homologous hybridization probe for a particular human gene. Moreover, even if the relevant cDNAs had been isolated, it was not possible to discriminate by DNA hybridization between wild-type and mutant human genes that had only a single-base-pair difference when the hybridization probe was large (≥100 bp). However, Conner et al. (*Proc. Natl. Acad. Sci. USA* **80**:278–282, 1983) synthesized specific oligonucleotides that could recognize either the wild-type or the mutant DNA. Moreover, oligonucleotides complementary to the sequences of both the wild type and the mutant were synthesized so that it could be determined whether a person was heterozygous, as well as whether a person was homozygous. For sickle-cell anemia, two 19-base-long oligonucleotides were used, one complementary to the normal β-globin gene (β^A) and the other complementary to the sickle-cell gene (β^S). DNA from normal individuals ($\beta^A\beta^A$) hybridized only with the β^A probe, DNA from individuals with sickle-cell anemia ($\beta^S\beta^S$) hybridized only with the β^S probe, and DNA from heterozygous individuals ($\beta^A\beta^S$) hybridized with both probes. This model system was the first demonstration of the feasibility of determining genotypes by DNA hybridization and opened up the possibility of detecting a range of human genetic disorders by hybridization with oligonucleotide probes. This possibility has been realized with the development of a large number of DNA-based gene mutation tests. Although the original strategy has been largely supplanted by newer techniques, such as PCR and OLA, this work was important in establishing that single DNA base-pair mutations could be accurately and easily detected.

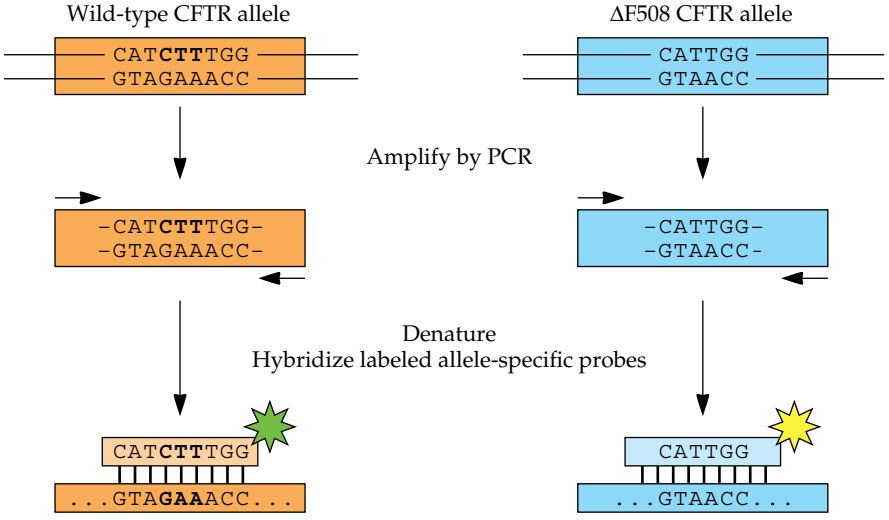
Table 4.2 The most common mutations of the CFTR protein that lead to cystic fibrosis[a]

Mutation designation[b]	Amino acid change to the CFTR protein
ΔF508	Deletion of phenylalanine at position 508
G542X	Replacement of glycine at position 542 by a stop codon
W1282X	Replacement of tryptophan at position 1282 by a stop codon
N1303K	Replacement of asparagine at position 1303 by lysine
1717-1G > A	Replacement of glycine by alanine at the last nucleotide in the intron proceeding nucleotide 1717 in the cDNA
R553X	Replacement of arginine at position 553 by a stop codon
I148T	Replacement of isoleucine at position 148 by threonine
3120 +1G > A	Replacement of glycine by alanine at the first nucleotide in the intron following nucleotide 3120 in the cDNA

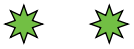
[a]Data from Eshaque and Dixon, *Biotechnol. Adv.* **24:**86–93, 2006.
[b]Amino acids are numbered starting at the N-terminal end of the protein.

Figure 4.18 Allele-specific hybridization to screen for cystic fibrosis. **(A)** The CFTR gene is amplified by PCR, and then the PCR products are denatured and incubated with oligonucleotide probes that specifically hybridize to wild-type or mutant (usually ΔF508) alleles (note that only part of the sequence of the CFTR gene and probe are shown). After washing to remove unbound probe, hybridization is detected by measuring fluorescence emitted by the fluorophore attached to the probe. **(B)** Healthy individuals are homozygous for the wild-type allele, while those affected by the disease carry two mutant alleles. Cystic fibrosis carriers are heterozygous for the CFTR alleles.

A



B

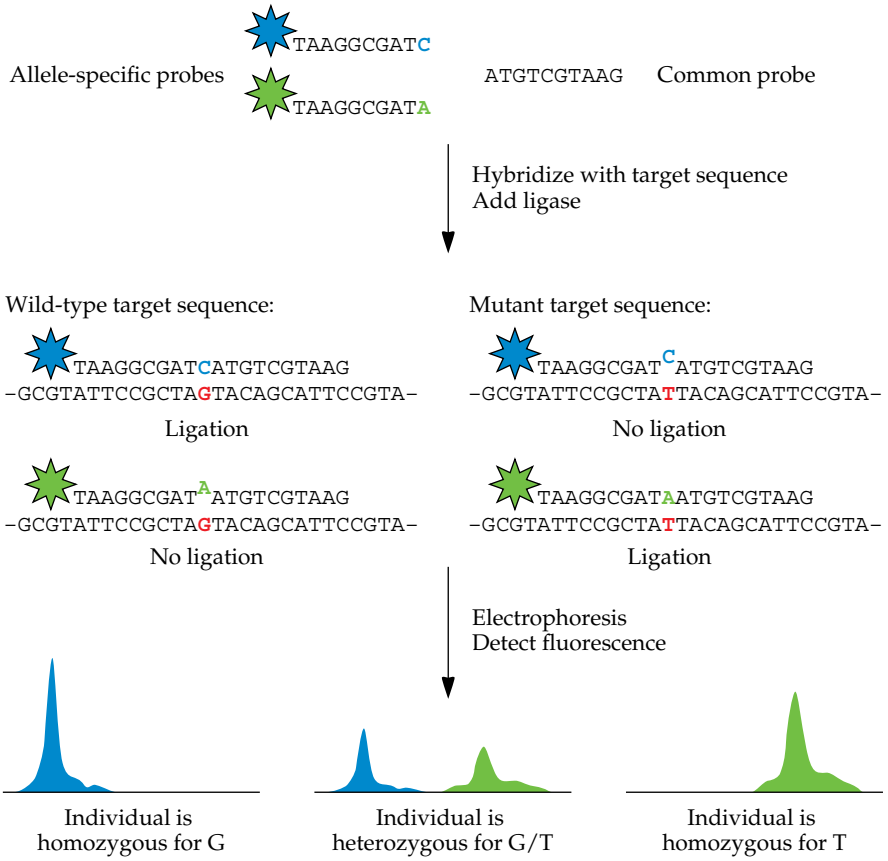
Source of blood sample	Hybridization test results
Normal individual	
Cystic fibrosis carrier	
Cystic fibrosis-affected individual	

(e.g., ΔF508) and wild-type genes separately (Fig. 4.18A). In this way, it is possible to distinguish between healthy individuals with two wild-type alleles, cystic fibrosis carriers with one mutant and one wild-type allele, and cystic fibrosis-affected individuals with two mutant alleles (Fig. 4.18B). Diagnostic kits are commercially available that test patient blood samples for the presence of a panel of common CFTR mutations as recommended by the American College of Medical Genetics.

Oligonucleotide Ligation Assay

The oligonucleotide ligation assay (OLA) is also commonly used to detect SNPs known to be associated with human diseases with a high degree of accuracy. In this diagnostic assay, two short oligonucleotide probes (~50 nucleotides) are designed to anneal to adjacent sequences within a gene that encompasses the polymorphic nucleotide (Fig. 4.19). Importantly, one of the probes (allele-specific probe) has, as its last base at the 3′ end, the nucleotide that is complementary to the polymorphic nucleotide. The second probe (common probe) is complementary to the sequence immediately downstream of the polymorphic nucleotide. When these two probes are hybridized with the target gene (which has been

Figure 4.19 OLA. Ligation of the allele-specific and common probes occurs only when the allele-specific probe is perfectly complementary to the target sequence. Longer ligated probes may be separated from shorter unligated probes by electrophoresis, and then probe fluorescence can be detected to distinguish homozygous wild-type, homozygous mutant, and heterozygous individuals.



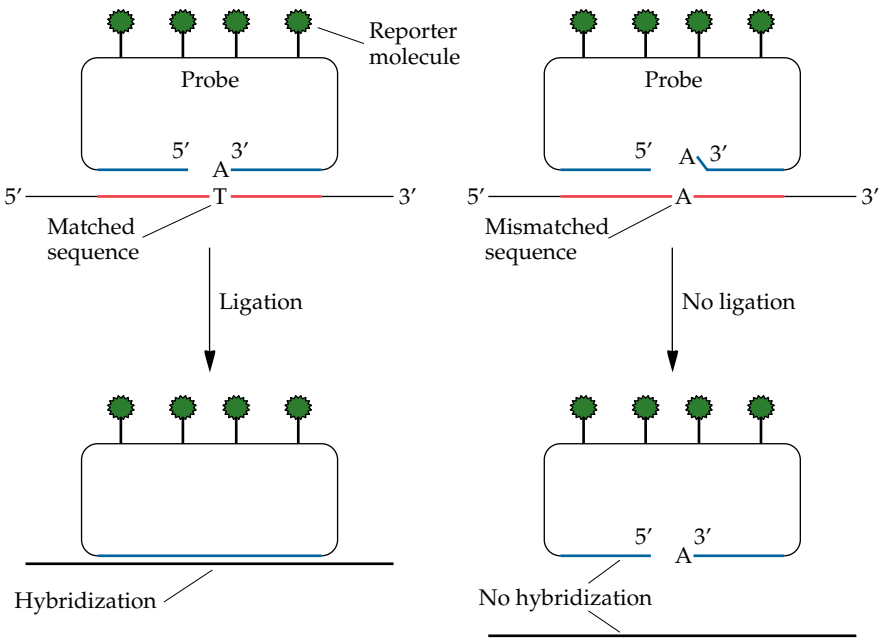
amplified by PCR), base-pairing occurs between the allele-specific probe and the target sequence, including the 3′ nucleotide if the complementary polymorphic nucleotide is present, and the common probe binds immediately downstream. DNA ligase, added to the reaction, covalently joins the two probes. However, if the nucleotide at the 3′ end of the allele-specific probe is mismatched, it will not base-pair with the polymorphic nucleotide in the target DNA sequence, although the common probe will be perfectly aligned. As a consequence of the single-nucleotide misalignment, DNA ligase cannot join the two probes. In short, OLA is designed to distinguish between two possibilities: ligation if both probes are perfectly matched to the target sequence and no ligation if the allele-specific probe carries a mismatched nucleotide. Other oligonucleotide probes may be designed to detect different SNPs in the same gene.

To determine whether ligation has occurred, one of the two probes is labeled with a reporter molecule. For example, the allele-specific probe may be labeled at the 5′ end with a fluorescent dye that can be detected by laser excitation and fluorescence emission. Probes may be labeled with different fluorescent molecules corresponding to different SNPs. If an individual is homozygous for a particular allele, either two copies of a normal gene or two copies of a disease-associated SNP, a positive signal will be detected only from one of the fluorescent probes. Heterozygous individuals will yield positive signals from two different probes. Overall, the OLA system is rapid, sensitive, highly specific, and amenable to automation.

Padlock Probes

Diagnostic assays for SNPs that use padlock probes are very similar to those that use OLA probes, except that the former utilizes only one probe rather than two as used in the OLA procedure. A padlock probe is

Figure 4.20 The function of a padlock probe. When the bases at the 5′ and 3′ ends of the probes are completely paired to the target DNA (red), ligation can take place. When there is a single-base mismatch at the 3′ end of the probe, ligation cannot occur, and the probe assumes a conformation that does not allow hybridization.

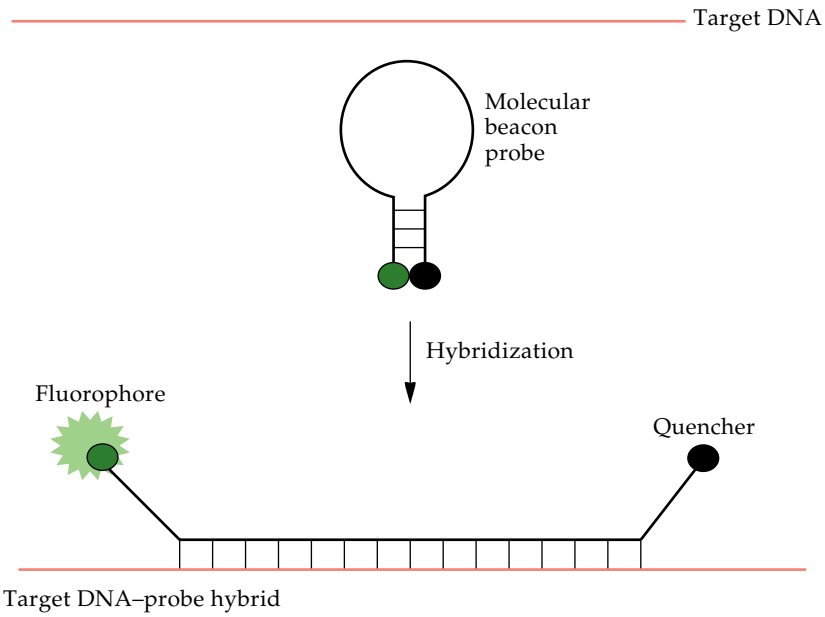


an oligonucleotide that is complementary at its 5′ and 3′ ends, but not in its middle region, to a target sequence (Fig. 4.20). When a padlock probe hybridizes to its target sequence, the 5′ and 3′ ends of the probe come into close proximity to each other. If the ends of the probe are exactly complementary to the target sequence, after hybridization the double-stranded DNA can be joined together by DNA ligase. If there is a mismatch between the target and probe, ligation does not occur. The requirement for both ends of the probe to bind perfectly to the target sequence for ligation to occur ensures a high specificity of detection and therefore the ability to easily detect SNPs. Following the ligation reaction, the probe–target hybrid can be detected by the activity of reporter molecules attached to the middle (linker) portion of the probe. To increase the sensitivity of the assay, the ligated, circular probes can be amplified by rolling-circle amplification using *phi29* DNA polymerase before labeling with a fluorescent dye.

Molecular Beacons

Molecular beacon probes are also used to detect specific sequences of nucleic acids (Fig. 4.21). A typical molecular beacon probe is 25 nucleotides long. The 15 nucleotides in the middle are complementary to the target DNA and are designed so that this single-stranded molecule does not form a structure in which these nucleotides base-pair with one another. However, the 5 nucleotides at each end are complementary to each other and not to the target DNA. A fluorescent molecule (fluorophore) is attached to the 5′ end, and a nonfluorescent molecule

Figure 4.21 Hybridization of a molecular beacon probe to target DNA. In the absence of a target DNA, the complementary sequences at the ends of the molecular beacon base-pair, and the quencher prevents fluorescence of the fluorophore. If the target sequence is present, it hybridizes to the molecular beacon, thereby separating the fluorophore and quencher moieties. As a result, the fluorescence of the fluorophore is not quenched.



(quencher) that can absorb the energy emitted by the fluorophore before it fluoresces is attached to the 3′ end. In solution at room temperature, the conformation of the molecular beacon ensures that the fluorophore and quencher are close to one another, and the fluorophore is quenched (does not fluoresce). On the other hand, when the 15 middle nucleotides of the molecular beacon probe hybridize to a target DNA or RNA sequence, the fluorophore and quencher are separated from each other and the fluorophore is not quenched, i.e., it fluoresces. With this procedure, care must be taken to maintain the reaction mixture at near-ambient temperatures, since high temperatures can break the bonds between the complementary nucleotides in the intrastrand (hairpin) stem of the molecular beacon, with the result that the molecule fluoresces. For this procedure to be effective, all 15 nucleotides in the molecular beacon probe must be perfectly complementary to the target DNA or RNA. The sensitivity of this procedure can be improved dramatically if the target DNA is first amplified by PCR.

A number of variations of the basic molecular beacon protocol have been developed. For example, combinations of molecular beacon probes may be used simultaneously, provided that each one is complementary to a different target DNA and contains a different-color light-emitting fluorophore (Fig. 4.22). For example, when the goal is to determine the genotype of an individual, two different molecular beacon probes are added to a biological sample, such as blood, that contains DNA from that individual. The first molecular beacon probe is labeled with one of the fluorophores (e.g., fluorescein), and all 15 probe nucleotides are exactly complementary to the wild-type sequence (Fig. 4.23). A one-nucleotide difference is sufficient to prevent hybridization. The second molecular beacon probe is labeled with a different fluorophore (e.g., Texas red), and the 15 probe nucleotides are complementary to the sequence from the mutant form (Fig. 4.23). Following hybridization, the appearance of green fluorescence indicates a homozygous normal genotype, red fluorescence indicates a homozygous mutant genotype, and both green and red fluorescence indicates a heterozygous genotype.

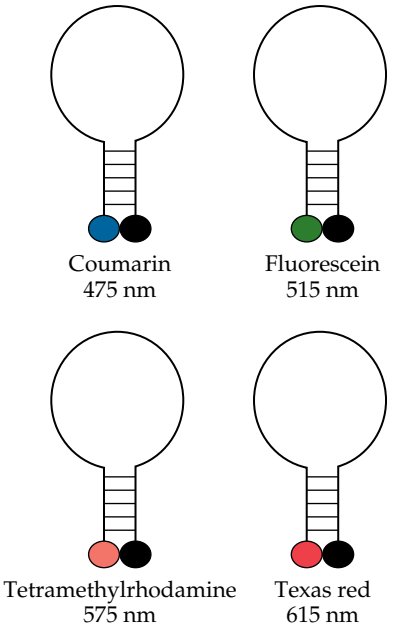


Figure 4.22 Molecular-beacon probes with different fluorophores. The name of the fluorophore and the wavelength of its maximum emission are indicated.

PCR-Based Detection Methods

Many nucleic acid diagnostic tests are based on PCR (see chapter 2). The advantages of PCR-based tests include (i) specificity that enables detection of a particular nucleotide sequence in complex samples, (ii) sensitivity that enables detection of low-abundance targets, (iii) an amplification step that generates substantial amounts of a target sequence for additional analyses such as hybridization or sequencing, (iv) rapid analysis (usually completed in 1 to 2 h or less), (v) multiplexing that enables identification of multiple targets in a single sample, and (vi) low cost. Conventional PCR to detect a specific sequence in a sample and real-time quantitative PCR to determine the abundance of a target sequence have been used extensively in diagnostic laboratories over the last 20 years.

Allele-Specific PCR

Allele-specific PCR (also referred to as a PCR amplification refractory mutation system) screens for known SNPs. Different forward PCR primers are used to distinguish among alleles that differ by a single nucleotide, for example, between healthy and disease-associated alleles. One forward

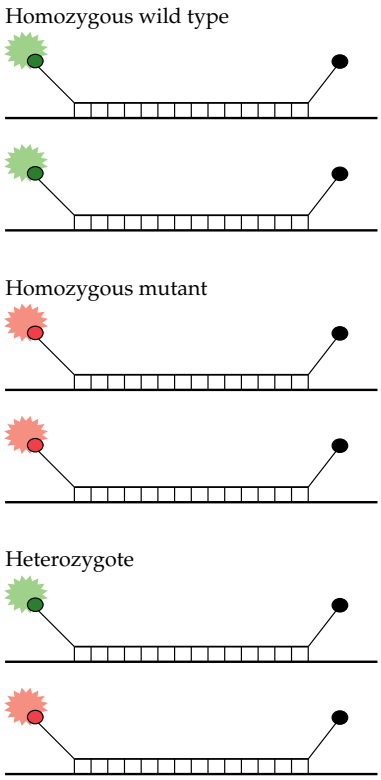
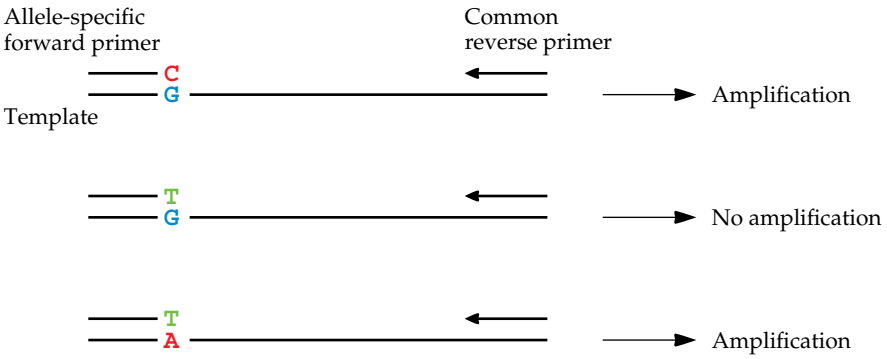


Figure 4.23 Use of two different molecular beacon probes, each with a different fluorophore (green and red circles), to distinguish between different genotypes. When the probe is bound to target DNA, the fluorescence of the fluorophore is no longer quenched.

primer is exactly complementary to the normal DNA sequence, and another forward primer anneals to a variant sequence containing the disease-associated SNP (Fig. 4.24). The primers are usually designed to place the polymorphic nucleotide at the 3′-terminal end of the primer because most polymerases used for PCR do not extend 3′ mismatched primers efficiently. The reverse primer is complementary to the opposite strand and is common to all reactions. Each reaction contains only one of the allele-specific forward primers and the common reverse primer, together with the patient DNA sample, a thermostable DNA polymerase, and all four deoxyribonucleotides. PCR amplification occurs only when a forward primer is present that is exactly complementary to the target sequence in the patient sample; mismatches between primer and template DNA prevent primer annealing and therefore primer extension during DNA synthesis. An advantage of this method, compared to allele-specific hybridization, described above, is that the amplification and diagnostic steps are combined.

One variation of allele-specific PCR is known as competitive oligopriming, in which two SNP-specific forward primers are included in a single reaction. To discriminate between the PCR products, each forward primer is labeled with a different fluorescent dye. For example, one forward primer that is exactly complementary to a normal allele may be labeled at its 5′ end with rhodamine (which fluoresces red), while another forward primer that is complementary to the disease-associated allele may be labeled at its 5′ end with fluorescein (which fluoresces green) (Fig. 4.25). In both cases, the reverse primer that is complementary to the opposite strand is unlabeled. PCR amplification occurs only when a forward primer is exactly complementary to the target DNA; therefore, the presence of these two forward primers in the same reaction mixture will result in the amplification of either the normal or disease-associated DNA sequence or both, depending on which alleles are present. If an individual is homozygous for the normal allele, after PCR and removal of unincorporated primers, the reaction mixture will fluoresce red; if he or she is homozygous for the disease-associated allele, the reaction mixture will fluoresce green; and if he or she is heterozygous, the reaction mixture will fluoresce yellow. This

Figure 4.24 Allele-specific PCR to detect SNPs. Amplification occurs only when the forward and reverse PCR primers perfectly match the target sequence. Different allele-specific primers that carry an SNP-specific nucleotide at the 3′ end are employed to distinguish among alleles that differ by a single nucleotide.



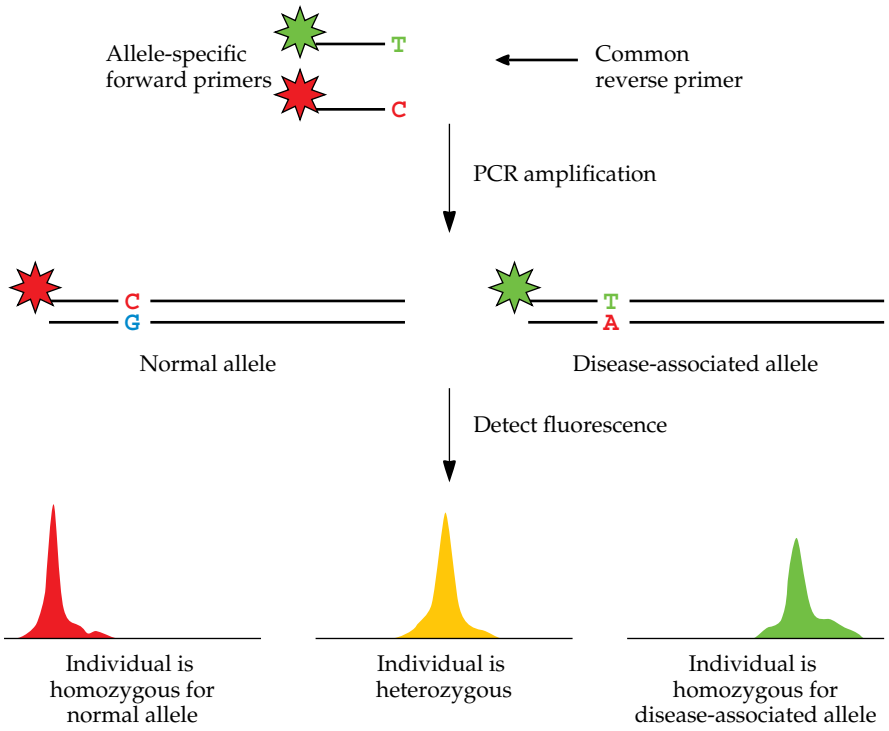


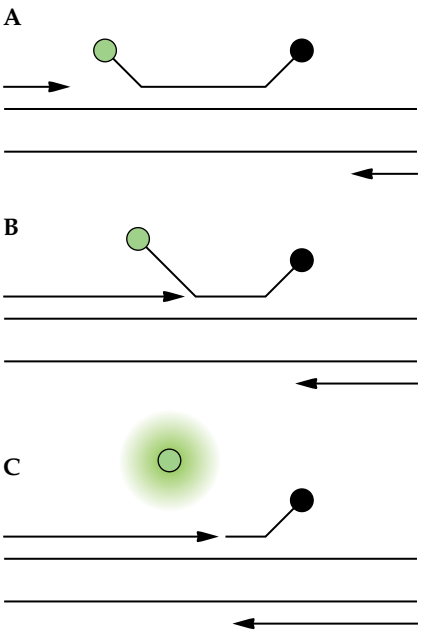
Figure 4.25 Competitive oligopriming to detect SNPs. Two allele-specific forward primers that differ in the nucleotide at the 3′ end are included in a single reaction together with a reverse primer that anneals to a sequence that is common among variant alleles. PCR amplification occurs only when a forward primer is perfectly complementary to the target DNA. To discriminate between PCR products corresponding to normal and disease-associated alleles, each forward primer is labeled with a different fluorescent dye. In the example shown here, PCR products from individuals who are homozygous for the normal allele fluoresce red, those from individuals who are homozygous for the disease-associated allele fluoresce green, and those from individuals who are heterozygous fluoresce yellow.

assay can be automated and adapted for any single-nucleotide target site of any gene that has been sequenced.

TaqMan PCR for SNP Genotyping

The TaqMan PCR protocol is used to screen individuals for the presence of SNPs that are indicative of any of a variety of genetic diseases. It is based on the 5′ nuclease activity of *Taq* polymerase, which is commonly used to amplify DNA in PCR applications. To simultaneously detect normal and disease-associated alleles, two TaqMan probes are utilized. Each probe is exactly complementary to either the normal or the disease-associated DNA sequence, and each probe has a different fluorescent dye attached to its 5′ end (Fig. 4.26). Intact probes, whether unbound or bound to template DNA, do not fluoresce, because of the presence of a quencher molecule at the 3′ end of the probe. PCR primers anneal to sequences that flank the probe hybridization site, and as PCR amplification proceeds, the TaqMan probe is displaced by the growing DNA strand. The 5′ nuclease activity of the *Taq* polymerase degrades the 5′ end of the TaqMan probe, thereby releasing the

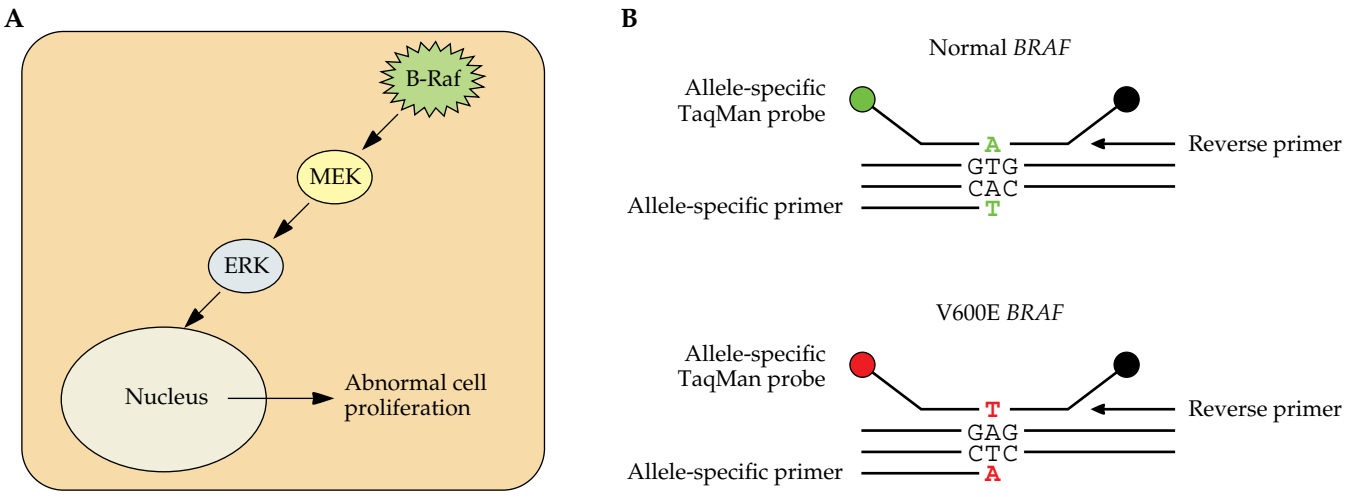
Figure 4.26 TaqMan PCR. **(A)** A TaqMan probe that is complementary to the wild-type DNA is added prior to PCR amplification of the DNA sequence. The probe contains a fluorescent dye attached to its 5′ end (green circle) and a quencher attached to its 3′ end (black circle). When the probe is intact, the quencher interacts with the fluorophore, quenching its fluorescence. The PCR primers are indicated by arrows. **(B)** In the extension phase of the PCR, the TaqMan probe is displaced by the growing DNA strand. **(C)** Subsequently, the 5′ fluorescent dye is cleaved from the probe by the 5′ nuclease activity of the *Taq* polymerase, leading to a dramatic increase in the fluorescence of the reporter dye (shown as a starburst of color from the dye).



fluorescent dye and removing it from the proximity of the quencher molecule. Thus, only TaqMan probes that were previously bound to target DNA will be degraded and subsequently fluoresce. Any mismatched probes that are not complementary to the SNP in the region where the TaqMan probe binds will be displaced but not cleaved, so they will not fluoresce. By monitoring the fluorescence at two different wavelengths (one for each TaqMan probe), it is possible to distinguish individuals that are homozygous for each SNP (either normal or disease associated) and heterozygous (carrying one normal and one disease-associated allele). This technique may be used to detect more than two different SNPs at the same time as long as the fluorescent dyes attached to each SNP-specific probe have well-separated, nonoverlapping fluorescence maxima.

A TaqMan diagnostic assay has been developed to screen individuals for specific mutations in the *BRAF* gene that are associated with melanoma. Melanoma is a type of skin cancer that is caused by changes in melanocytes that produce the skin pigment melanin. More than 75,000 cases are diagnosed each year in the United States. More than half of these are associated with mutations in the *BRAF* gene, which encodes B-Raf, a serine/threonine protein kinase that regulates a signal transduction pathway (BRAF–MEK–extracellular signal-regulated kinase [ERK] pathway) that controls cell division, differentiation, and apoptosis in response to growth factors (Fig. 4.27A). In 90% of the *BRAF* mutations associated with

Figure 4.27 TaqMan PCR to detect *BRAF* alleles associated with melanoma. **(A)** Mutations in *BRAF* may lead to constitutive kinase activity of B-Raf and, consequently, constitutive activation of the MEK-ERK signaling pathway that controls cell proliferation. **(B)** A TaqMan PCR assay to detect the V600E *BRAF* allele commonly associated with cutaneous melanoma uses an allele-specific PCR primer and a reverse PCR primer that binds to a sequence present in both wild-type and mutant *BRAF*. An allele-specific TaqMan probe, labeled with a fluorescent dye and a quencher, binds to the sequence that spans the mutation but in the DNA strand opposite to that bound to the forward primer. As the PCR proceeds, the probe is displaced and cleaved by *Taq* polymerase as it extends the strand from the reverse primer, thereby separating the fluorophore and quencher. An allele-specific PCR primer and a TaqMan probe labeled with a different fluorescent dye bind specifically to the normal *BRAF* sequence. The presence of mutant and/or normal *BRAF* is determined by monitoring the fluorescence of the cleaved probe(s).



cutaneous melanomas, glutamic acid, a negatively charged amino acid, has been replaced with the hydrophobic amino acid valine in codon 600 (V600E; GTG→GAG). This change alters the conformation of the activation segment of the protein, leading to constitutive kinase activity of B-Raf and uncontrolled proliferation of melanocytes. A TaqMan PCR assay to detect the V600E *BRAF* mutation uses a forward primer that specifically binds to the mutant *BRAF* sequence and a reverse primer that binds to a sequence present in both wild-type and mutant *BRAF* (Fig. 4.27B). A fluorescently labeled TaqMan probe also anneals to the sequence that spans the mutation but in the DNA strand opposite to that bound to the forward primer. Another probe, labeled with a different fluorescent dye, binds specifically to the wild-type sequence. As PCR proceeds, the probe is displaced and cleaved by *Taq* polymerase as it extends the strand from the reverse primer. The presence of mutant and/or wild-type *BRAF* is determined by monitoring the fluorescence of the released probe(s).

The mutation can be detected both in tumor tissue samples and in DNA from blood plasma. The screen assists in identifying patients that have a high probability of responding well to treatment with the drug vemurafenib, a B-Raf kinase inhibitor. Studies have shown that tumors with the V600E mutation are more sensitive to vemurafenib and that treatment with vemurafenib reduced the risk of death by 56% and decreased tumor size in 52% of melanoma patients with this allele compared to those who received chemotherapy. In 2011, the U.S. FDA approved the use of vemurafenib and the test for the V600E mutation for diagnosis and treatment of melanoma associated with this allele.

Real-Time PCR To Diagnose Infectious Diseases

Most nucleic acid diagnostic tests for etiological agents of infectious disease are based on PCR. PCR circumvents the requirement to grow the pathogens, which is time-consuming for viruses and slow-growing or fastidious bacteria. Not only does this technique enable identification of a viral, bacterial, or fungal pathogen, but also it can reveal specific characteristics of that pathogen, for example, the presence of antibiotic resistance genes that can be used to determine the best course of treatment. A successful PCR assay to detect microbes in human-derived samples must be able to distinguish among the vast number of commensal species that exist in the human gastrointestinal, genitourinary, and upper respiratory tracts and on skin.

Real-time PCR has all of these benefits of a PCR assay and, in addition, can quantify the pathogen causing an infection. In a real-time PCR, the double-stranded DNA product is bound by a fluorescent molecule, and the fluorescence is measured after each amplification cycle to quantify the amount of product, such as a pathogen sequence, in real time (described in chapter 2). Post-PCR processing is not required, and therefore, the time to pathogen identification is shorter than for conventional PCR. One application for this technology in infectious disease diagnosis is the detection of methicillin-resistant *Staphylococcus aureus* (MRSA) in clinical samples such as nasal swabs. MRSA is an increasing problem worldwide. Infections with these multidrug-resistant bacteria are usually associated with hospitals and other health care facilities, although community-associated infections are becoming increasingly common. In the United States, health care-associated MRSA infections account for 90,000 deaths and $4.5 billion in health care costs annually; thus, there is strong motivation to develop

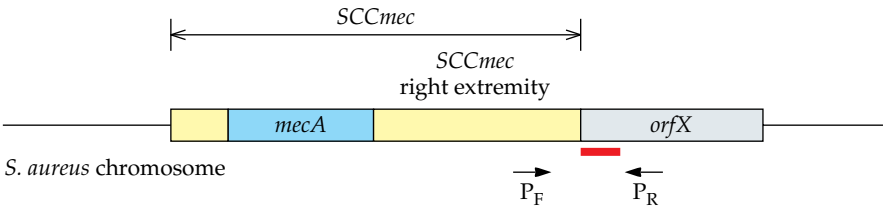
rapid, accurate screening tests to prevent transmission of these bacterial pathogens among hospitalized patients.

Methicillin resistance in *S. aureus* is due to the *mecA* gene encoding a cell wall penicillin-binding protein (PBP 2a) that protects *S. aureus* from the bacteriocidal effects of methicillin and structurally related β-lactam antibiotics. The gene is encoded within a genomic region (pathogenicity island) known as the staphylococcal cassette chromosome *mec* (SCC*mec*), which was likely acquired by some *S. aureus* strains by horizontal gene transfer (Fig. 4.28). SCC*mec* is inserted in the *S. aureus orfX* gene, and the SCC*mec*–*orfX* junction is a target for some MRSA diagnostic tests that employ real-time PCR. A pair of PCR primers is designed such that the forward primer anneals to the right extremity of SCC*mec* and the reverse primer anneals to a sequence within *orfX*. Seven different forward primers are employed; each binds specifically to a variant sequence of the right extremity of SCC*mec*. This enables discrimination of the seven possible subtypes of MRSA that differ in the sequence of the SCC*mec* right extremity. Each of the amplified DNA targets is detected with a molecular beacon probe labeled with a different fluorescent dye (Fig. 4.21). The fluorescent signal emitted by the hybridized molecular beacon is measured at the end of the annealing step in each cycle of the PCR, and the amount of fluorescence corresponds to the amount of MRSA present in the sample. Results are obtained in about 1 hour, which is significantly faster than culture-dependent detection methods that require 2 to 3 days and enables rapid intervention to treat the patient and prevent the spread of MRSA.

When a pathogen cannot be detected using available primers, broad-range PCR may be used. This approach employs universal primers that anneal to a common nucleotide sequence, for example, the 16S rRNA gene present in all bacteria. Following amplification, products are sequenced for pathogen identification by comparison of the sequence to a database such as the Ribosomal Database Project. This method is more suited to determine infections in samples of low complexity, for example, meningitis, pneumonia, and endocarditis, which occur in tissues that are normally sterile (meninges, lower lungs, and heart, respectively).

In addition to its use in the measurement of pathogenic agents in the environment, a variant of real-time PCR may be used to quantify the levels

Figure 4.28 Real-time PCR assay to detect MRSA. PCR primers are designed to anneal to the right extremity of the *SCCmec* pathogenicity island (P_F) and to a sequence in *orfX* (P_R). Primers that bind to variant sequences in the right extremity enable identification of the specific MRSA strain. At each cycle in the PCR, the amplification products are quantified by measuring the fluorescence emitted from a hybridized molecular beacon probe (probe binding site is indicated in red) that is labeled with a fluorescent dye. From Huletsky et al., *J. Clin. Microbiol*. **42:**1875–1884, 2004.

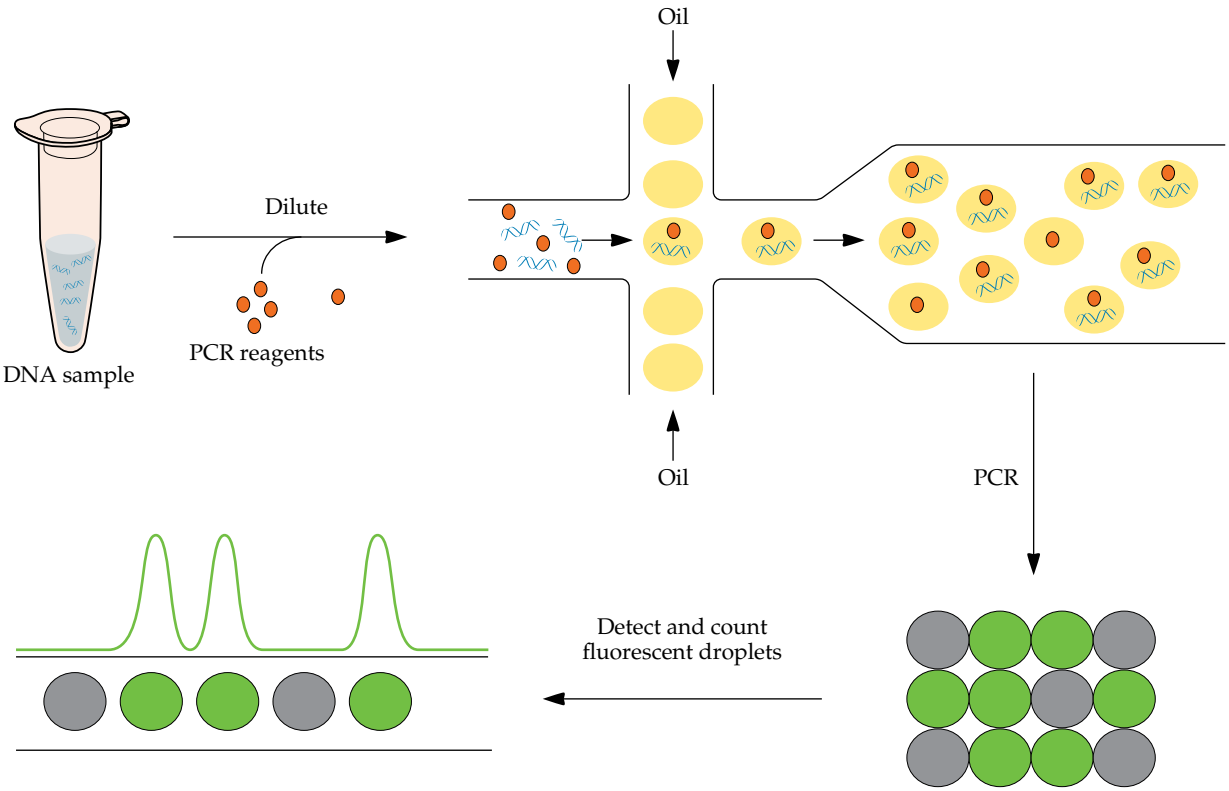


of a variety of mRNAs in different eukaryotic tissues or prokaryotic cells. In these instances, the initial target is RNA and not DNA, and therefore, a reverse transcription (RT) step is needed to generate cDNA before the real-time PCR (see chapter 2).

Droplet Digital PCR To Quantify Low-Abundance Sequences

Although real-time PCR is widely used for infectious disease diagnosis, it is less accurate for quantification of pathogens that are present at low levels, such as during latent infections with *Mycobacterium tuberculosis* or human immunodeficiency virus. Droplet digital PCR is a sensitive and precise method for quantification of target DNA, especially when it is present in low concentrations. The principles of amplification and amplicon detection for droplet digital PCR are similar to those for real-time PCR; however, initial sample preparation is different. Template DNA in a sample is diluted and encapsulated in 20,000 nanoliter-sized oil droplets (Fig. 4.29). Each oil droplet is an individual reaction system that contains PCR primers, DNA polymerase, deoxyribonucleotides, and other reaction components for amplification of a target sequence and fluorescent

Figure 4.29 Droplet digital PCR for absolute quantification of a target sequence. Extracted DNA is mixed with PCR primers, fluorescent probes or dyes, a thermostable DNA polymerase, and deoxyribonucleotides and partitioned into thousands of oil droplets. Each droplet contains zero, one, or more molecules of template DNA and all reagents required for amplification of a target sequence in a PCR reaction. At the end of the reaction, the droplets pass by a laser that records the fluorescence of each. The number of droplets that are positive for PCR products (green) and the number that are negative (gray) provide the concentration of target sequences in a sample.



probes or dyes for detection of PCR products. The PCR proceeds for 30 to 40 cycles, similar to conventional PCR, and then fluorescence intensity is measured for each droplet as it passes by a detector. The target quantity is determined simply by counting the number of droplets exhibiting a positive signal (containing at least one copy of template DNA) and negative signal (containing no template DNA), obviating the need to determine threshold cycles (C_T) and generate standard curves as required for quantification by real-time PCR (see chapter 2). Because the DNA molecules are distributed randomly into droplets, positive droplets may contain one, two, or more templates, and a statistical model (Poisson model) is applied to take this into consideration when calculating abundance. In addition to detecting low levels of microbes in patient or environmental samples, droplet digital PCR has been used to detect rare cancer-associated mutations, variations in gene copy number associated with neurogenetic disorders, and mRNA from genes that are expressed at low levels, such as stem cell biomarkers.

Loop-Mediated Isothermal Amplification for Point-of-Care Diagnostics

PCR-based diagnostic assays require specialized equipment to rapidly cycle the reactions through various temperatures for template denaturation, primer annealing, and DNA synthesis, and for real-time PCR, to measure the fluorescence intensity of probes and dyes. The tests are performed by skilled personnel in medical laboratories and, given the need to transport samples, typically have a turnaround time of days. They are therefore not amenable for some applications, in particular for point-of-care diagnostics (near-patient testing done outside a medical laboratory). Loop-mediated isothermal amplification (LAMP) is a simplified alternative to PCR in which the entire reaction is carried out at a single temperature, and therefore a specialized thermal cycler is not required and results are obtained in a shorter time. Key to this, during amplification, the template DNA is not heat-denatured; rather, a special DNA polymerase, such as phi29 polymerase from the *Bacillus subtilis* bacteriophage phi29 or *Bst* polymerase from *Geobacillus stearothermophilus*, displaces the nontemplate strand during DNA synthesis. For a LAMP diagnostic assay, at least four oligonucleotide primers are needed that are specific for the target sequence. One of the two strands of the DNA molecules in a sample serves as the initial template for synthesis of the complementary strand from one of the primers (inner primer) (Fig. 4.30A). In the next step, the newly synthesized strand is displaced during another round of DNA synthesis from the original template, extending from an upstream (outer) primer (Fig. 4.30B). The released strand is then a template for DNA synthesis (Fig. 4.30C), resulting in a product that forms a stem-loop at each end due to intrastrand complementarity of sequences that were derived from the inner primers (Fig. 4.30D). These dumbbell-shaped DNA molecules are amplified from the inner primers which bind to the loops to initiate strand displacement and DNA synthesis. Additional loop-binding primers may be employed to accelerate the reaction. A very large number of double-stranded DNA molecules are generated that can be detected by eye using visible dyes. For example, as a by-product of DNA synthesis, a large number of protons are produced that acidify the reaction solution and can be detected by a change in the color of a pH-sensitive dye. Because multiple primers are used to amplify a single template, the reaction products are long

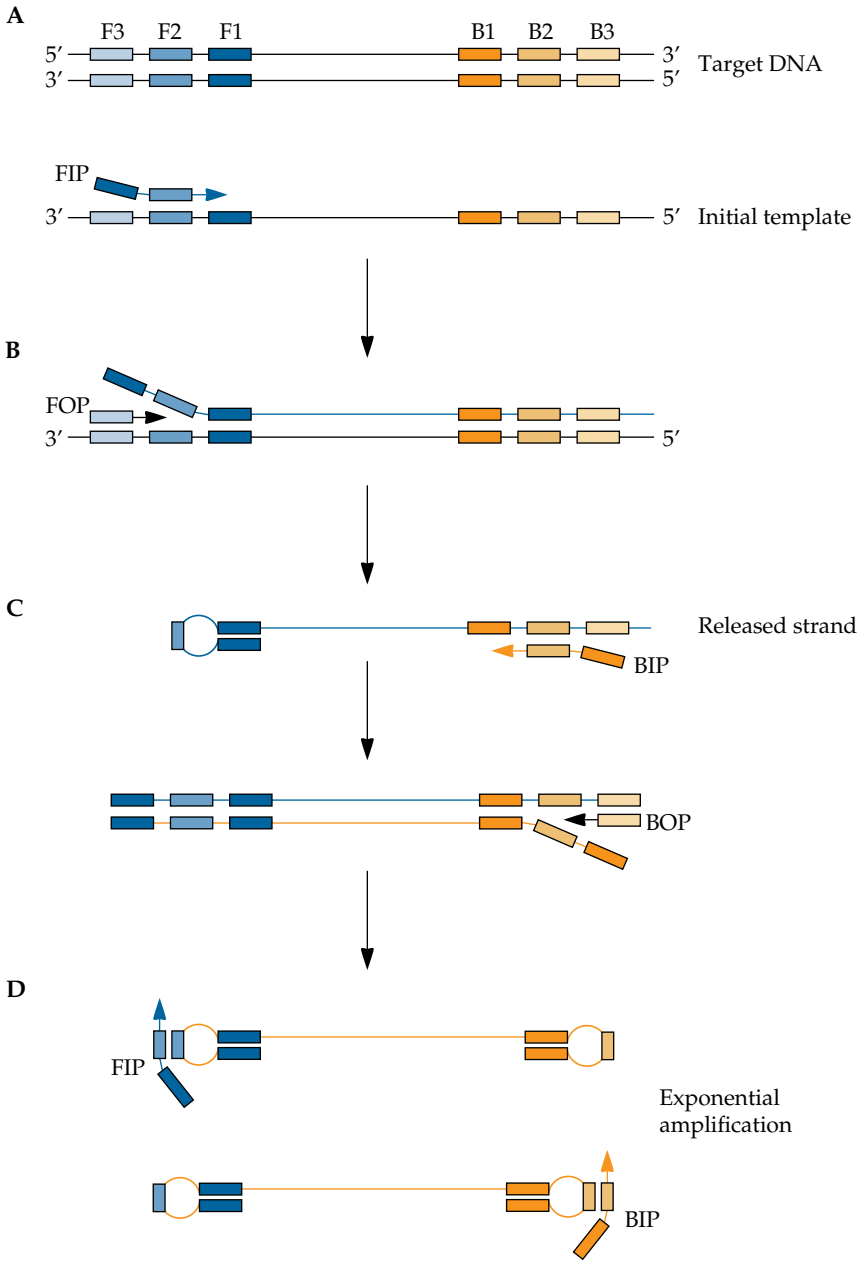


Figure 4.30 LAMP assay. **(A)** One strand of the DNA molecule containing the target sequence serves as the initial template for synthesis of a complementary strand from a forward internal primer (FIP) that anneals to binding site F2 and contains a sequence complementary to F1 (dark blue box). **(B)** The initial template strand is again a template for DNA synthesis. As DNA is synthesized from a forward outer primer (FOP) that anneals to site F3 in the template molecule, the DNA polymerase displaces the previously synthesized (blue) strand. **(C)** The released strand then serves as a template for synthesis of DNA extending from a backward internal primer (BIP) that anneals to site B2 and contains a sequence complementary to B1 (dark orange box). The newly synthesized (orange) strand is displaced during subsequent extension of a backward outer primer (BOP) that anneals to site B3 and forms a dumbbell-shaped molecule due to hybridization of complementary sequences at the ends of the strand. **(D)** The dumbbell-shaped molecules are amplified from primers FIP and BIP. Additional primers (not shown) that bind within the loop region may be employed to accelerate amplification. Data from Nagamine et al., *Mol. Cell. Probes*. **16:**223–229, 2002.

concatamers of target DNA of various sizes and therefore may not be suitable for some downstream applications, such as cloning.

An example of an application for LAMP is the diagnosis of SARS-CoV-2 during the 2020–2021 pandemic, when rapid, easy-to-use, low-cost tests were imperative to quickly identify infected individuals who then isolated to prevent transmission of the virus. In one LAMP test for SARS-CoV-2, a region within the nucleocapsid gene in the viral genome was targeted for amplification. Nasopharyngeal swab samples were introduced directly into the reaction tube. SARS-CoV-2 is an RNA virus, and therefore a reverse transcriptase reaction step was included in the LAMP assay to generate cDNA from the RNA genome. The target sequence was amplified using inner, outer, and loop primers specific for the nucleocapsid gene. The pH indicator phenol red was included in the reaction to detect a reduction in pH by a change in color from red to yellow as DNA synthesis proceeded. A positive result was obtained in 30 minutes at 65°C for samples containing at least 100 copies of target RNA. The assay is therefore sufficiently sensitive to identify individuals with moderate or high levels of SARS-CoV-2; efforts are under way to improve its performance.

Forensic Analysis Using PCR-Amplified Microsatellites

The DNA from a biological sample left at the scene of a crime can be analyzed and compared with the DNAs of likely suspects. A match between evidence and a particular individual is helpful to the prosecution. In addition, DNA comparisons have been used to determine whether individuals have been wrongly convicted of a crime. In other instances, DNA analyses help determine paternity and identify victims of disasters and terrorist attacks. Distinguishing individuals with DNA analysis is called DNA fingerprinting.

One approach for determining DNA relationships among humans relies on microsatellite DNA. Microsatellites are short tandem repeat sequences that occur throughout the human genome. The lengths of the repeated sequences range from 2 to 5 bp, and the numbers of repeats in a microsatellite range from about 5 to 50 (Fig. 4.31A). A microsatellite DNA sequence at a specific chromosome location can have different lengths in different individuals. This variability is due to either a gain or a loss of tandem repeats, probably during DNA replication. These changes do not have any biological effects because microsatellite DNA does not encode any proteins. Unrelated individuals generally have microsatellites that differ in length, but children inherit one set of microsatellite DNA sequences from each parent.

For forensic DNA fingerprinting, several microsatellites, usually tetra- or pentanucleotide repeats, are analyzed. The amount of DNA obtained from blood, semen, hair, skin, or other tissues found at the crime scene is usually very small, and therefore PCR is required to amplify the microsatellite DNA. Each microsatellite is amplified using PCR primers that specifically anneal to unique flanking sequences (Fig. 4.31A). The PCR products are separated as they migrate through a gel in an electric field (gel electrophoresis) based on their size, and the microsatellite pattern is visualized using a fluorescent dye.

A microsatellite DNA pattern (fingerprint) represents the repertoire of the lengths of these sequences in an individual (Fig. 4.31B). Because of the extensive variability in human microsatellite DNA sequences, the chance of

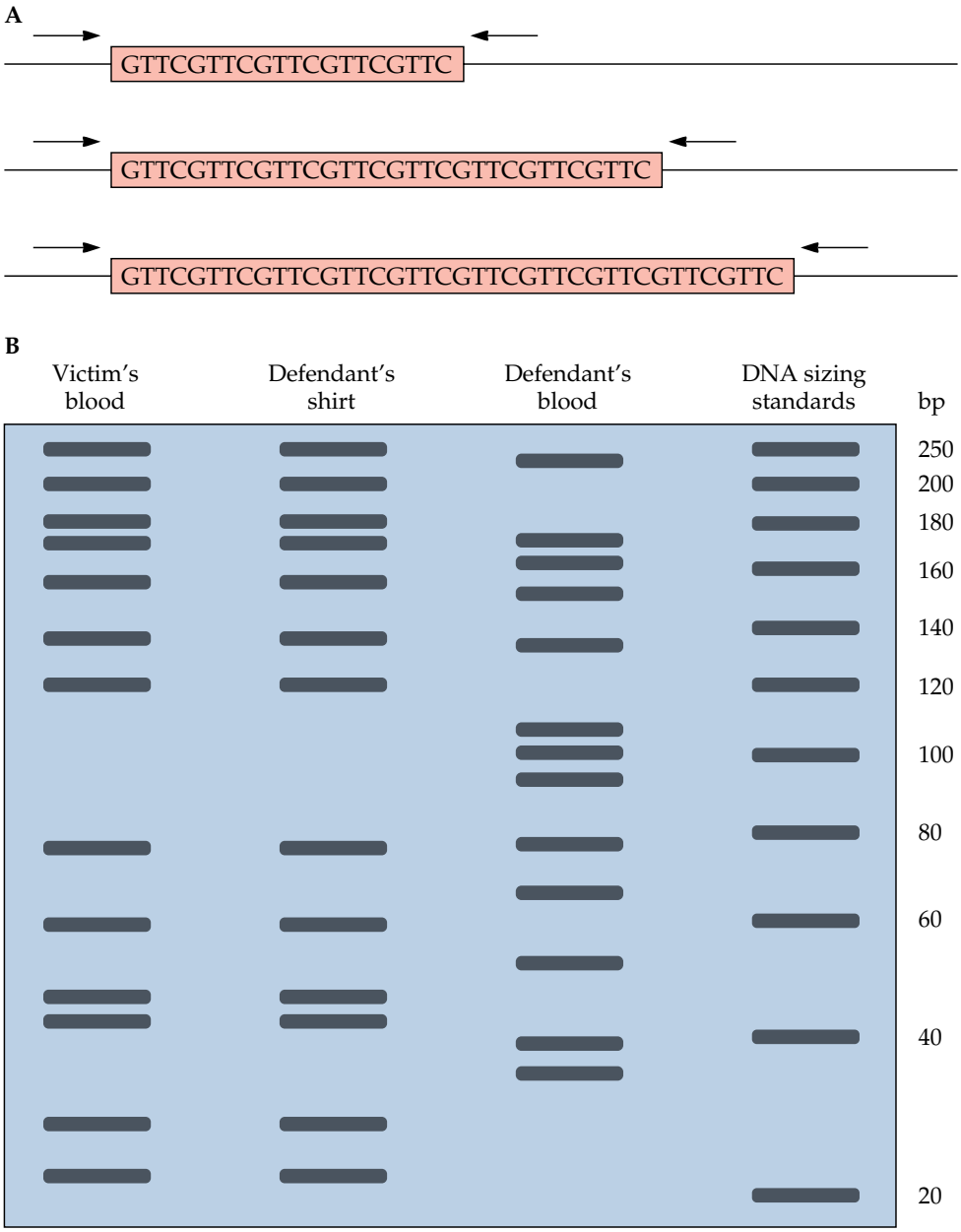


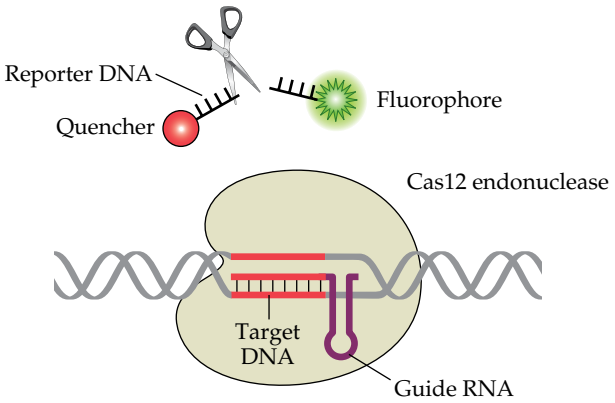
Figure 4.31 Forensic analysis of microsatellite DNA. **(A)** A tetranucleotide (GTTC) microsatellite repeated 5, 8, and 10 times in tandem. Arrows indicate annealing sites for PCR primers that flank the microsatellite and are used to amplify the microsatellite. **(B)** DNA fingerprints generated by amplification of 13 different microsatellites in blood samples from the victim, the defendant's shirt, and the defendant. PCR products were resolved by gel electrophoresis. The banding pattern of the microsatellite DNA from the blood on the defendant's shirt is identical to the victim's DNA fingerprint and different from the defendant's DNA fingerprint. The sizing of the DNA molecules in the bands is estimated by comparison with the positions of the sizing standards.

finding two individuals in the population with the same pattern of 13 microsatellites is less than 1 in 10^6. Therefore, an individual's DNA banding patterns based on microsatellite DNA sequences are almost as unique as their fingerprints. Many countries store forensic DNA fingerprints in databases that can aid in solving crimes. In the United States, the Combined DNA Index System (CODIS) was established in 1994 and is maintained by the Federal Bureau of Investigation. The original CODIS Core Loci consisted of 13 microsatellites; however, this was increased to 20 in 2017 to reduce the chances of false matches.

CRISPR-Cas-Based Diagnostic Assays

Rapid, specific, and sensitive recognition of a target DNA is a cornerstone of a nucleic acid diagnostic test. In the CRISPR-Cas system, a Cas endonuclease is directed to a target sequence in a genome by a guide RNA (see chapter 2). The specificity of this system is due to the ability of the guide RNA to hybridize to a complementary DNA sequence and to discriminate among single-base-pair differences. While genome editing typically employs the relatively simple Cas9 nuclease from the bacterium *Streptococcus pyogenes*, other Cas endonucleases that have properties suitable for use in molecular diagnostics have been isolated from various bacteria. Cas12 from a *Lachnospiraceae* bacterium binds to a target DNA sequence determined by guide RNA and is activated to cleave nearby DNA regardless of its sequence. In a diagnostic assay, the nonspecific collateral activity of Cas12 is exploited to cleave a reporter DNA. The reporter DNA is a single-stranded oligonucleotide that, similar to the TaqMan probes, is covalently bound to a fluorophore on one end and a quencher on the other end that absorbs energy emitted by the fluorophore when the molecule is intact. Recognition of a specific DNA sequence by the guide RNA activates the Cas12 endonuclease to cleave the reporter DNA which was included in the reaction (Fig. 4.32). Fluorescence of the released fluorophore is measured, and fluorescence intensity indicates the amount of target sequence

Figure 4.32 CRISPR-Cas system as a diagnostic tool. A guide RNA binds to a complementary strand of a target DNA and activates Cas12 endonuclease to indiscriminately cleave single-stranded DNA. As a diagnostic tool, a reporter DNA is included in an assay with the DNA sample, a target-specific guide RNA, and Cas12. Target-dependent activation of Cas12 results in cleavage of the reporter DNA and thereby separation of the fluorophore from the quencher, which are both covalently bound to the reporter DNA. Fluorescence indicates the presence of the target sequence.



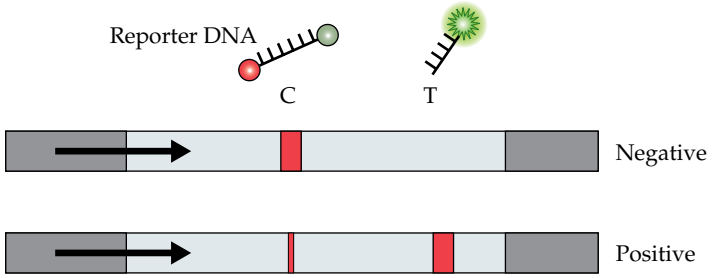


Figure 4.33 CRISPR-Cas-based lateral flow test strip. A DNA sample is mixed with a specific guide RNA, Cas12 endonuclease, and a reporter DNA and then applied to one end of the lateral flow strip. The molecules migrate through the test strip, and if the target sequence is not present in the sample DNA, the reporter DNA remains intact and is captured at the first control line (C). A colored signal is generated at this position, indicating a negative result. If the target sequence is present, it is bound by the guide RNA, which activates Cas12 to cleave the reporter DNA. The cleaved reporter DNA migrates to the second, test line (T), where it produces a colored signal that indicates a positive result.

present in a sample. The sensitivity of the assay can be increased by first amplifying the target sequence by a method such as LAMP (Fig. 4.30). Colorimetric probes have been developed for lateral flow test strips to detect the presence of a target sequence by eye as a colored band (Fig. 4.33); including the LAMP assay, results are achieved within 40 minutes compared to 2 hours for real-time PCR. Several target sequences may be detected in a single sample by including multiple probes, each with a different fluorophore. If the target is an RNA sequence, for example, in the RNA genome of a virus, then the Cas13 endonuclease (e.g., from *Listeria seeligeri*) is employed, as it recognizes and cleaves RNA.

One interesting application for this technology is the determination of object provenance (origin). The objects are deliberately labeled with nonviable bacterial spores that have been engineered with a unique 28-bp DNA sequence (barcode) integrated into their genome. The barcode does not encode a protein. An advantage of using spores, such as those of the bacterium *Bacillus subtilis* or the yeast *Saccharomyces cerevisiae*, is their ability to persist in the environment for long periods of time. Auxotrophic strains that require a nutritional supplement for growth are used to prevent proliferation, and the spores are rendered unable to germinate by deletion of essential germination genes. Objects carrying the spores are identified by the DNA barcode using a guide RNA in conjunction with Cas12- or Cas13-mediated collateral cleavage of a reporter DNA. The origins of various products can be determined in this way; for example, the source of a food could be traced during an outbreak of foodborne illness, or brand-name goods could be distinguished from counterfeit products that infringe on a trademark. The spores can also be inoculated onto surfaces such as soil or carpet to identify objects or persons to which the spores were transferred as they passed through.

DNA Microarrays

While hybridization probes and nucleic acid amplification tests are used extensively to detect mutations in single genes, microarray analysis (see chapter 2) enables simultaneous detection of mutations at multiple loci

associated with complex genetic diseases. Arrays have been developed for a variety of purposes. AmpliChip CYP450, a microarray to predict the response to certain drugs, and Mammaprint, a microarray to detect breast cancer, were the first diagnostic arrays to be approved by the U.S. FDA, in 2005 and 2007, respectively. Companies are now marketing microarray-based genomic tests directly to consumers to determine ancestry and risk for specific health conditions.

Detection of Multiple Alleles

Cytochrome P450 oxidases are a group of enzymes that regulate the metabolism of approximately 25% of prescription drugs, including antidepressants, antipsychotics, beta blockers, and opiates (Table 4.3). Different alleles of two cytochrome P450 genes, *CYP2D6* and *CYP2C19*, are associated with differential drug metabolism. Over 80 different alleles of *CYP2D6* and three different alleles of *CYP2C19* have been discovered in humans and are linked to different levels of cytochrome P450 enzyme activity. Based on genotype, individuals are classified into four groups: poor, intermediate, extensive, and ultrarapid metabolizers. Individuals exhibiting ultrarapid drug metabolism may have multiple (3 to 13) copies of the *CYP2D6* gene and remove the drug rapidly. Thus, the dosage of a drug may have to be increased in patients with this genetic makeup. Poor metabolizers, on the other hand, may have two inactive *CYP2D6* alleles, and therefore, the drug may accumulate, leading to undesirable side effects, or drugs that require activation *in vivo* may be ineffective. For example, tamoxifen, a therapy for breast cancer, is administered in an inactive form and is converted into an active form (endoxifen) by several enzymes, including the CYP2D6 enzyme. This drug is unlikely to be effective in individuals who are poor or intermediate metabolizers, such as those with inactive or low-activity *CYP2D6* alleles. Thus, by identifying the alleles that an individual possesses, it may be possible to predict how that person will metabolize a particular drug. The AmpliChip CYP450 microarray detects 33 different *CYP2D6* alleles, including deletions and duplications, and 3 different *CYP2C19* alleles. Genomic DNA is extracted from blood samples or buccal (cheek) swabs, PCR amplified, fragmented, labeled, and hybridized to the array to identify the alleles present. The results facilitate drug selection, optimization of drug dosage, and prediction of adverse drug effects.

Table 4.3 Drugs metabolized by CYP2D6 and CYP2C19[a]

Cytochrome P450	Antidepressants	Beta blockers	Antipsychotics	Proton pump inhibitors	Antiepileptics	Others
CYP2D6	Amitriptyline Clomipramine Desipramine Paroxetine Imipramine Venlafaxine	Carvedilol Metoprolol Propafenone Timolol	Haloperidol Risperidone Thioridazine			Atomoxetine Codeine Dextromethorphan Flecainide Mexiletine Ondansetron Tamoxifen Tramadol
CYP2C19	Amitriptyline Clomipramine			Omeprazole Lansoprazole	Diazepam Phenytoin Pantoprazole	Cyclophosphamide Progesterone Phenobarbitone

[a]Data from Li et al., *Curr. Genomics* **9:**466–474, 2008.

Detection of Epigenetic Markers

Epigenetic modifications refer to the covalent modification of DNA or histone proteins associated with DNA. Specific bases in DNA may be methylated, and acetyl, methyl, or phosphate groups, or the small proteins ubiquitin and small ubiquitin-like modifier (SUMO) protein, may be covalently attached to specific amino acids in histone proteins. Epigenetic changes alter gene expression and cellular phenotype, are stable, and may be transmitted to offspring (i.e., they are heritable), although they do not change the genotype of a cell. DNA methylation almost always involves the addition of a methyl group to a cystosine residue in a CpG dinucleotide, generating 5-methylcytosine (Fig. 4.34). The addition is catalyzed by a DNA methyltransferase. About 75% of the 28 million CpG dinucleotides in the mammalian genome are methylated. DNA methylation is a normal process that serves a regulatory function. It downregulates gene expression in response to environmental conditions, either by inhibiting binding of transcription factors or by recruiting specific methyl-CpG-binding domain proteins that induce the formation of heterochromatin (tightly packed regions of DNA). DNA methylation plays an important role in animal development, for example, by silencing genes on the X chromosome or on one of the two copies of some autosomal genes, depending on which parent it was inherited from.

Altered DNA methylation patterns are associated with a number of human diseases, especially cancers. In tumors, levels of genome methylation generally decrease; however, the promoters of some genes exhibit increased methylation of normally unmethylated CpG dinucleotides. Methylation of core promoter sequences inactivates expression of these genes, some of which encode tumor suppressors (Fig. 4.35). Thus, epigenetic modification can have an effect similar to that of a mutation or deletion. However, because DNA methylation is an epigenetic alteration, it is, at least in theory, reversible and therefore may be a target for demethylation drugs to treat human disease.

Aberrant methylation seems to be nonrandom, and CpG dinucleotides at different loci have different propensities for methylation. These factors,

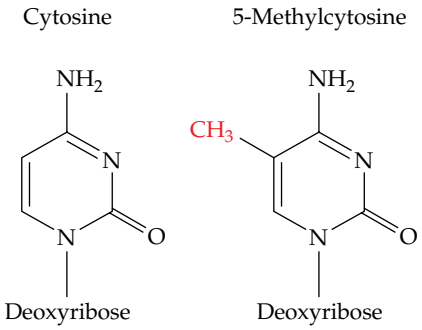
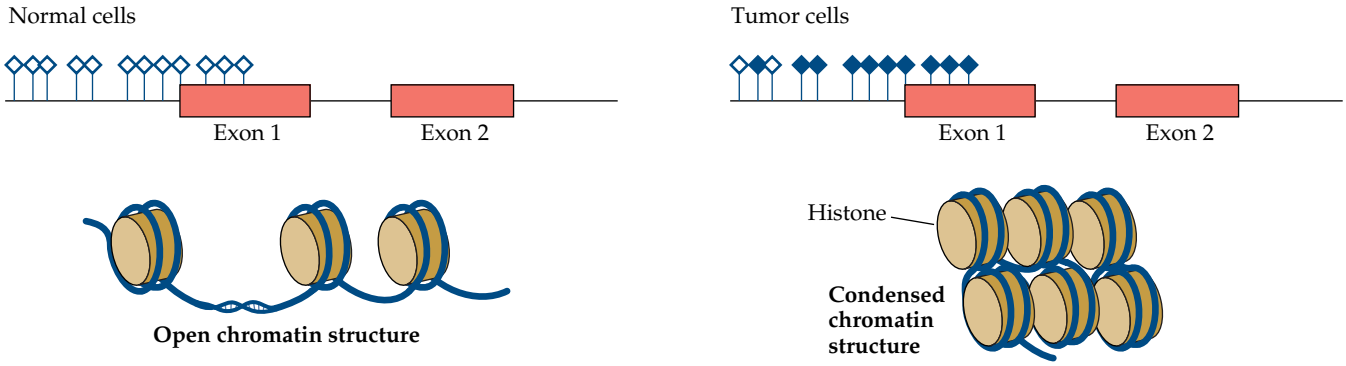


Figure 4.34 DNA methylation. A methyl group is covalently bound to a cytosine residue in CpG dinucleotides, generating 5-methylcytosine.

Figure 4.35 Hypermethylation of DNA in promoter regions decreases gene expression. In tumors, levels of genome methylation generally decrease; however, the promoters of some genes exhibit increased methylation of normally unmethylated CpG dinucleotides. Methylated CpG dinucleotides are associated with heterochromatin and transcriptional silencing. Expression of tumor suppressor genes may be inhibited. Open diamonds, unmethylated CpG; closed diamonds, methylated CpG. Data from Lao and Grady, *Nat. Rev. Gastroenterol. Hepatol.* **8:**686–700, 2011.



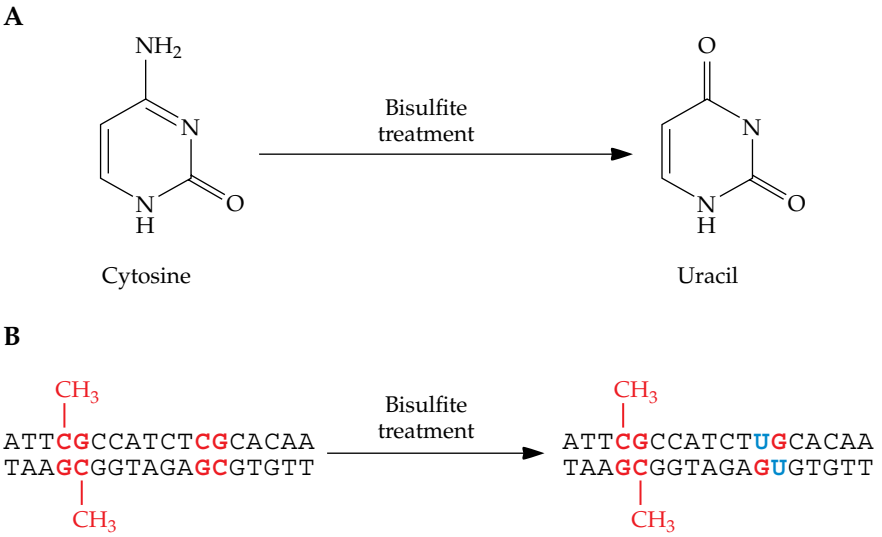


Figure 4.36 Sodium bisulfite converts cytosine to uracil **(A)**. Unmethylated cytosines in CpG dinucleotides are converted to uracil, while methylated cytosines are protected from deamination **(B)**.

as well as the chemical stability of methylcytosine, make methylation patterns useful as biomarkers of disease. Genome-wide or locus-specific screening may be used to detect aberrant DNA methylation. Strategies to detect methylated sequences employ (i) methylation-sensitive restriction endonucleases that bind to DNA only when CpG dinucleotides in their recognition site are methylated (or only when they are unmethylated), (ii) 5-methylcytosine-specific antibodies to capture methylated fragments that can be identified using a microarray or by DNA sequencing, and (iii) chemical deamination of nonmethylated cytosines to uracils using sodium bisulfite; methylated cytosines are protected against deamination (Fig. 4.36). The last method converts differences in methylation patterns into differences in sequence (i.e., C to U) that can be identified by PCR, sequencing, or microarrays.

In one diagnostic approach to detect DNA methylation profiles associated with specific cancers, genomic DNA extracted from tumors is first treated with sodium bisulfite and then hybridized to a microarray that detects 450,000 CpG sites in the entire human genome. To increase the amount of sample DNA applied to the array, the bisulfite-treated DNA is amplified in an isothermal reaction using random primers. During the amplification step, uracils converted from unmethylated cytosines during bisulfite treatment are treated the same as thymines. The amplification products are enzymatically fragmented (to 300 to 600 bp) and applied to the microarray, which consists of 50-nucleotide probes (50-mers) attached to silica beads that each detect a single CpG site. The probe hybridizes to a specific CpG site on a DNA fragment such that the 3′ nucleotide of the probe is complementary to the base immediately upstream of the query site (Fig. 4.37). After washing away the unhybridized fragments, DNA polymerase extends the probes by a single, chain-terminating dideoxynucleotide (Fig. 2.28) that is complementary to either a cytosine that corresponds to a deamination-resistant,

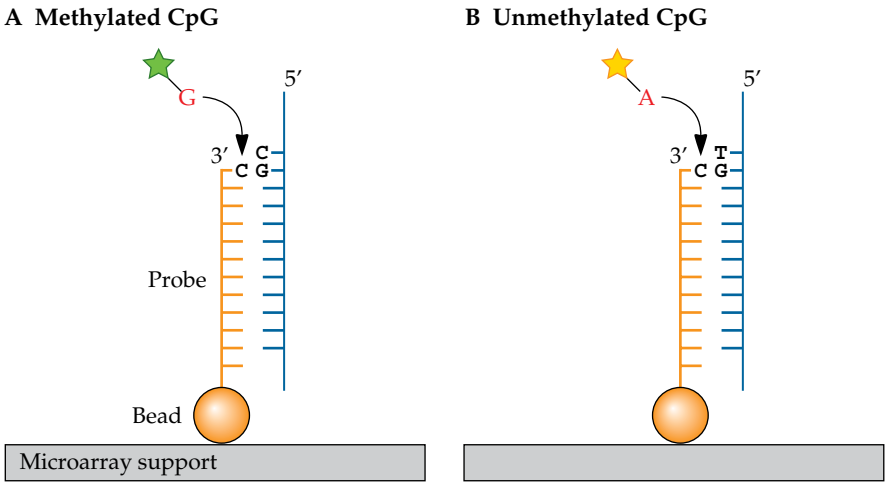


Figure 4.37 Microarray analysis of methylated DNA. Oligonucleotide probes are covalently linked to silica beads (orange circles) and immobilized on a solid support. Each of the thousands of beads in a microarray carries many copies of a probe that detect a single CpG site, although only one is shown here for simplicity. Bisulfite-treated sample DNA (blue) is amplified, fragmented, and applied to the microarray. A probe hybridizes to a specific DNA fragment such that the 3′ nucleotide of the probe is complementary to the base immediately adjacent to the query site and is extended by a single nucleotide by DNA polymerase. If the query site contains a cytosine from a methylated CpG, then a dideoxyguanosine triphosphate, labeled with an identifying fluorophore (green star), is inserted **(A)**. If the query site contains a thymine derived from a bisulfite-converted, unmethylated CpG, then a dideoxyadenosine triphosphate, labeled with a different fluorophore (yellow star), is inserted **(B)**. A laser measures fluorescence to reveal the identity of the inserted nucleotide and the methylation status of the CpG site in the sample. All of the CpG sites targeted by the probes of the microarray are interrogated at the same time.

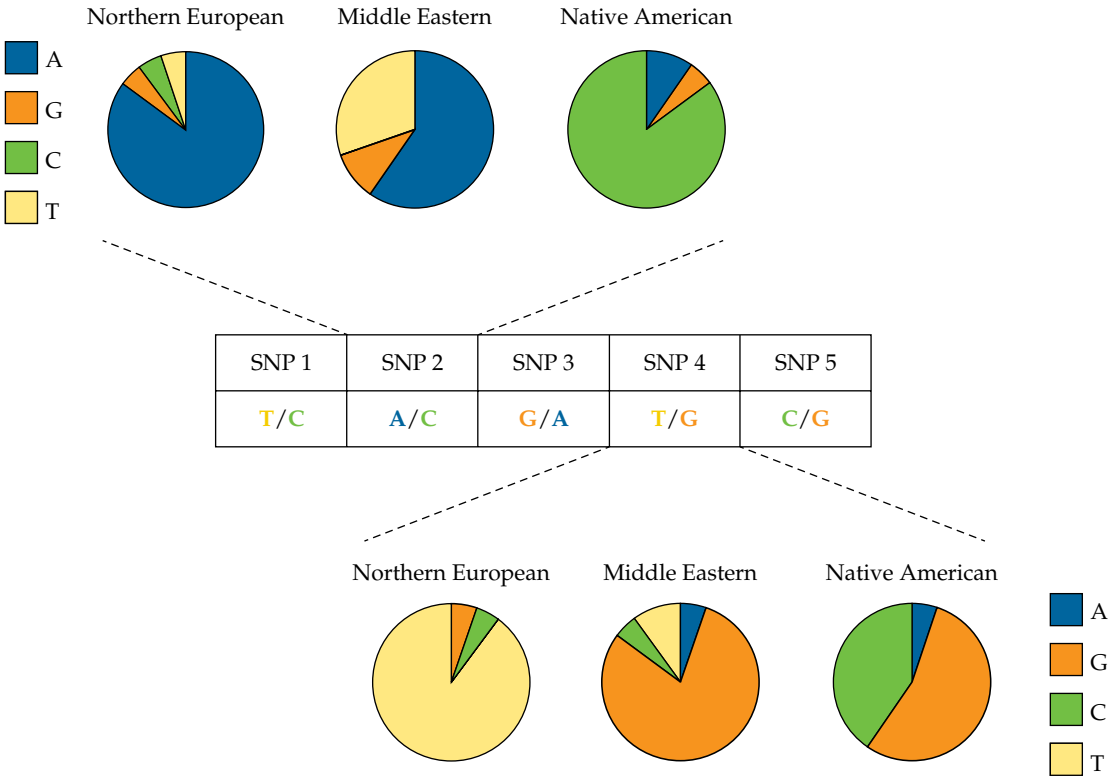
methylated cytosine in the original genomic DNA (Fig. 4.37A) or a thymine that was derived from a bisulfite-converted, unmethylated cytosine (Fig. 4.37B). Each of the two possible complementary nucleotides (either G or A) is labeled with a different fluorescent dye that enables its identification using a laser.

Ancestry Determination

Most of the 4 million to 5 million SNPs in our genome are inherited. It is therefore possible to infer ancestry by examining a number of different SNPs in an individual's genome and comparing the pattern of the SNPs to those of other individuals within and among populations. For ancestry determination, SNPs in three different types of DNA can be examined: autosomal DNA (which includes all of a person's DNA except for the X and Y chromosomes and mitochondrial DNA), which originates from a combination of a person's parents' DNA, paternal DNA (i.e., the Y chromosome), which is passed on from father to son, and maternal DNA (i.e., mitochondrial DNA), which is passed on from a mother to all of her children. Since mitochondrial DNA changes only very little over many generations, characterization of mitochondrial DNA is an ideal means of tracking migrations over many hundreds of generations of human genetic history.

Several SNP microarrays have been developed to analyze an individual's ancestry. Genomic DNA is typically extracted from buccal swabs or saliva samples, fragmented, and hybridized to a microarray that detects SNPs at 600,000 or more sites in the human genome, most commonly in autosomal DNA, although some microarrays also detect SNPs in mitochondrial and Y chromosomal DNA. The targeted SNPs are selected to capture most of the common genetic variation in populations throughout the world. In one type of microarray, a SNP is identified by extension of arrayed oligonucleotide probes with a single labeled nucleotide that is complementary to the SNP. The frequency of each of the SNPs identified in an individual's genome in various ethnic groups provides an estimation of the ethnic background or ancestry of that individual (Fig. 4.38). For example, this type of testing may indicate that an individual is genetically 70% Northern European, 17% Middle Eastern, and 13% Native American. Some tests can detect the presence of Neanderthal DNA, which

Figure 4.38 Estimating ancestry by allele frequencies. The two alleles at each of five SNPs in an individual's genome are shown in this example. The two alleles at SNP 2 are A and C. A's are present at this site at a high frequency in Northern European populations and to a lesser extent in Middle Eastern populations, while C's are common at this site in Native American populations. At SNP 4, the two alleles are T and G, which are present at high frequencies in Northern European and Middle Eastern populations, respectively. The alleles at each of more than 600,000 SNPs in the genome are compared to the frequencies of those alleles in reference populations to estimate the ethnicity of the individual.



makes up about 2% of European and Asian genomes and about 0.5% of African genomes.

Several companies market genetic ancestry kits; each test costs around $100. In the few years that they have been available, more than 25 million people have purchased these tests. Some geneticists, anthropologists, and epidemiologists have publicly expressed concern that naive laypersons, anxious for definitive information about their personal genetic ancestry, often misinterpret the results of these tests. For example, the fact that a particular SNP or haplotype is quite common in certain populations does not mean that its presence is a definitive diagnostic of membership in those populations. There is often high genetic diversity within populations, and gene flow can readily occur between populations. Thus, a particular allele could have been inherited from a population in which it is less common rather than from the population in which it appears to be "diagnostic," leading some individuals to mistakenly believe that they are genetically part of a particular ethnic or racial group.

Animal Species Determination

Microarrays that detect SNPs are also gaining attention for their potential applications in wildlife forensics to identify illegal trade in animals or their parts. For example, microarrays that identify SNPs in various fish species can be used to identify fish that were caught illegally or labeled deceptively in supermarkets. Illegal fishing includes harvesting more fish than quotas in a region allow, removing fish from prohibited waters, or intentionally mislabeling the species or origins of marketed fish. Overexploitation of fish populations is a $23 billion/year problem worldwide and represents 20% of the harvest, which threatens the sustainability of many fisheries. Fish of the same species from different waters often look very similar; however, SNP analysis can distinguish not only among different fish species but also among different populations of the same fish species.

To develop a SNP microarray to determine the source of a fish or fish products, the RNA extracted from eight individual fish from each of four populations of common sole, European hake, and Atlantic herring in the Mediterranean, Baltic, and North Seas was sequenced and mapped, and 1,536 polymorphic sites (the capacity of an array) were initially selected for each species. Of these SNPs, 427 for sole, 395 for hake, and 281 for herring were sufficiently variable to enable reliable discrimination among populations (tested on 27, 19, and 18 different populations, respectively). As few as 13 polymorphic sites in the hake genome were sufficient to reveal the origin of a fish with 98% accuracy. Distinction of fish populations using SNPs is more reliable than using microsatellites, and the microarray format facilitates low-cost, high-throughput processing of samples that will have important implications in fishery regulation and conservation. This technology can be expanded to stem the illegal traffic in remains of other wild animals and enforce international conventions designed to prevent this trade.

Whole-Genome Sequencing To Assess Genetic Disease Risk

The low cost, high quality, and increased speed of next-generation sequencing technologies (see chapter 2) have made accessible the sequencing of individual human genomes to assess risk for various

diseases. When combined with family medical histories, analysis of genome sequences can be a powerful diagnostic tool. One investigation assessed the potential of whole-genome sequencing to screen for age-related chronic diseases in healthy, asymptomatic adults. The genomes of 209 volunteers with a median age of 55 years were sequenced, and single nucleotide variants (~22 million) and insertions-deletions (~6 million) were identified. In all, 310 unique disease-associated alleles in 231 genes were identified using the American College of Medical Genetics criteria. These included autosomal recessive, autosomal dominant, and X-linked recessive alleles. Age-related chronic disease-associated alleles were identified in 19% of study participants that predicted increased risk for neoplasms, cardiovascular disease, diabetes, chronic liver disease, and neurological disorders. The findings were supported by family medical histories and results of routine clinical tests, including diagnostic whole-body magnetic resonance imaging, global metabolomics, and heart monitoring (e.g., echocardiography and cardiac rhythm monitoring). Identification of individuals at risk through whole-genome sequencing may reduce premature mortality that is associated with age-related chronic diseases.

Interpretation of whole-genome sequences for diagnostic purposes is limited by our understanding of the alleles associated with human diseases, although this is rapidly expanding. Currently, more than 190,000 disease-causing mutations have been identified in 6,770 genes. Thus, the absence of a disease-associated allele does not necessarily indicate reduced risk. On the other hand, interpretation of genome sequences may lead to false-positive conclusions (i.e., the diagnosis of a disease that is not present). Although the cost of sequencing a human genome is now low enough to be accessible to many individuals, few clinical personnel have the skills to interpret the resultant data. Genome sequences among individuals are highly polymorphic, and there is a large amount of information to process, interpret, and relay to patients, both immediately following sequence acquisition, and later as our ability to interpret genome sequences improves. There is the potential to reveal risks for diseases for which there are no treatments, for example, for individuals who possess alleles associated with Huntington's disease. Patients must be counseled about the medical, emotional, and social risks and benefits of genomic testing, as well as understand the scope and limitations of the predictive ability of genome sequence data.

Detecting RNA Signatures of Disease

There are several advantages of detecting and quantifying RNA biomarkers of disease. While DNA sequences can indicate the presence of a disease-associated allele, not all individuals with the allele will develop the disease; that is, expression (penetrance) of the trait may be less than 100%. The degree to which a trait is expressed is influenced by the presence of other modifying alleles or genes for polygenic diseases, epigenetic modifiers, and environmental factors. Diseased or infected tissues exhibit characteristic changes in RNA profiles. Moreover, specific mRNAs are generally expressed in many copies and therefore can often be detected without an amplification step. Many tumors secrete substantial amounts of RNA within vesicles or associated with RNA-binding proteins, which

can be detected in blood samples, obviating the need to perform a more invasive tissue biopsy. Current efforts are cataloguing extracellular RNAs in healthy and diseased tissues. While protein profiles are characteristic of diseased tissues, small, noncoding RNAs play an important role in regulating protein production, and alterations in their abundance are indicative of cell dysregulation.

RNA signatures of disease may be determined by RNA sequencing, quantitative reverse transcriptase PCR, microarrays, and hybridization probes. RNA sequencing has the advantages of broad quantitative range, especially high sensitivity for detection of rare RNAs and the ability to detect single nucleotide and RNA splice variants (see chapter 2). It is often employed for discovery of differentially expressed RNA, although data analysis is a challenge for clinical applications. Quantitative reverse transcriptase PCR, microarrays, and hybridization probes, which all measure predefined RNA sequences, are currently more rapid and cost-effective for routine diagnostic assays.

Detection of Disease-Associated Changes in Gene Expression

Assessing Prostate Cancer Aggressiveness by PCR

Prostate cancer is currently detected by elevated levels of prostate-specific antigen in an ELISA. In many men, the tumor grows slowly and remains localized, while in others the cancer spreads rapidly and requires a more aggressive and immediate treatment. The diagnostic test used in routine prostate cancer screening does not distinguish between these outcomes. Prostate cancer prognosis can be assessed by measuring the expression of cell cycle progression genes to determine how rapidly cancer cells are multiplying. In a quantitative reverse transcriptase PCR assay, the mRNA levels from 31 cell cycle progression genes are measured. RNA is extracted from tumor cells obtained by needle biopsy or prostatectomy, converted to cDNA using reverse transcriptase, and quantified by TaqMan PCR using gene-specific probes (described previously in this chapter). Levels of transcription of each gene are determined from the C_T value (see chapter 2), normalized to the levels of several housekeeping genes, and then averaged over all of the cell cycle progression genes to obtain a tumor aggressiveness score ranging from 0 to 10. Each unit increase in the score represents a 2-fold increase in expression levels and an increase in risk of disease progression. The results of the test together with other clinicopathological factors are used to estimate 10-year prostate cancer-specific metastasis and mortality risk and to determine an appropriate course of treatment ranging from regular monitoring of individuals with low tumor aggressiveness scores to prostate surgery or hormone and/or radiation therapy for those with higher scores.

Determining Breast Cancer Prognosis Using Microarrays

In addition to applications for microarrays in genotyping as described above for the AmpliChip CYP450 microarray, this tool has also been used to determine and detect RNA profiles associated with disease (see chapter 2). For example, the clinical microarray MammaPrint measures the expression of 70 different genes in breast tumors. Breast cancer differs among patients in terms of risk of metastasis and response

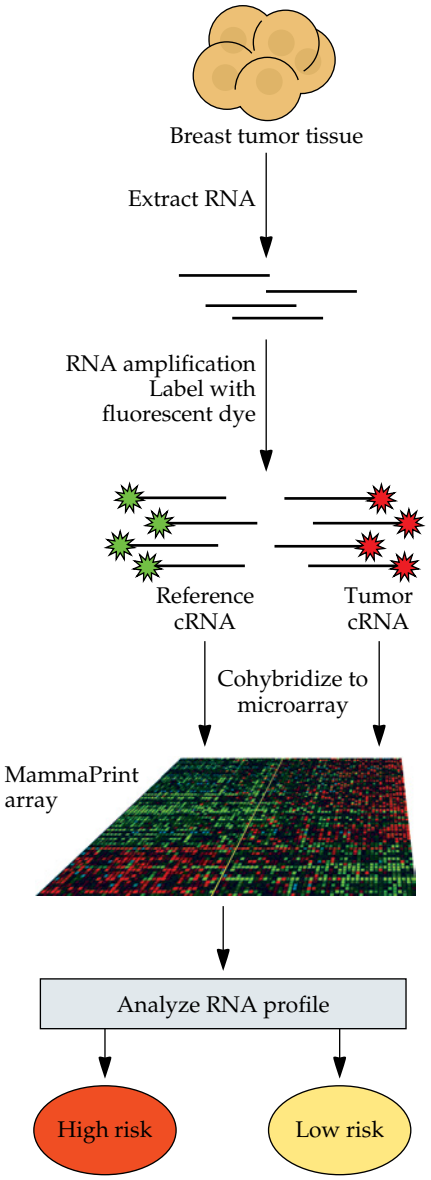


Figure 4.39 The MammaPrint diagnostic array assesses risk for recurrence or metastasis of breast cancer. The microarray contains sequences that target mRNA expressed from 70 different breast cancer prognosis genes. RNA is extracted from patient tumor tissue, amplified using T7 RNA polymerase to generate complementary RNA (cRNA), and then labeled with a fluorescent dye (e.g., Cy5). Levels of expression of each gene are determined by hybridizing labeled cRNA to the microarray and measuring fluorescence intensity relative to reference cRNA; reference cRNA consists of a pool of RNA from 100 breast tumors (from equal numbers of patients with good and poor outcomes). Tumor RNA expression profiles stratify women into two groups: those with a low risk for breast cancer recurrence and those with a high risk for recurrence.

to treatment. Prior to development of a diagnostic microarray, disease prognosis was based primarily on tumor size and lymph node status; however, these are not accurate predictors of long-term outcome. The goal of the MammaPrint diagnostic array is to determine if a woman with breast cancer that has not spread to lymph nodes is at risk for recurrence or metastasis (Fig. 4.39). Women with low risk have a 10% chance of having the cancer return within 10 years without further treatment, while those with high risk have a 29% chance of recurrence in the same period.

To develop the MammaPrint microarray, RNA from 98 breast tumors was extracted and hybridized to microarrays carrying 25,000 human gene sequences. Two distinct gene expression profiles were apparent; these correlated with disease progression, that is, with either progressive, metastatic disease (poor prognosis) or lack of metastasis 5 years after primary diagnosis (good prognosis). Elevated expression levels of genes involved in cell cycle regulation, invasion, metastasis, and angiogenesis provide an RNA signature for poor prognosis, and these women are recommended for chemotherapy. Without this test, many patients are unnecessarily subjected to cytotoxic chemotherapy; 70% would have survived without this aggressive treatment. If the microarray results indicate a good prognosis, a less aggressive hormonal therapy such as tamoxifen may be prescribed. Tamoxifen blocks the activity of the hormone estrogen that breast tumors need to grow.

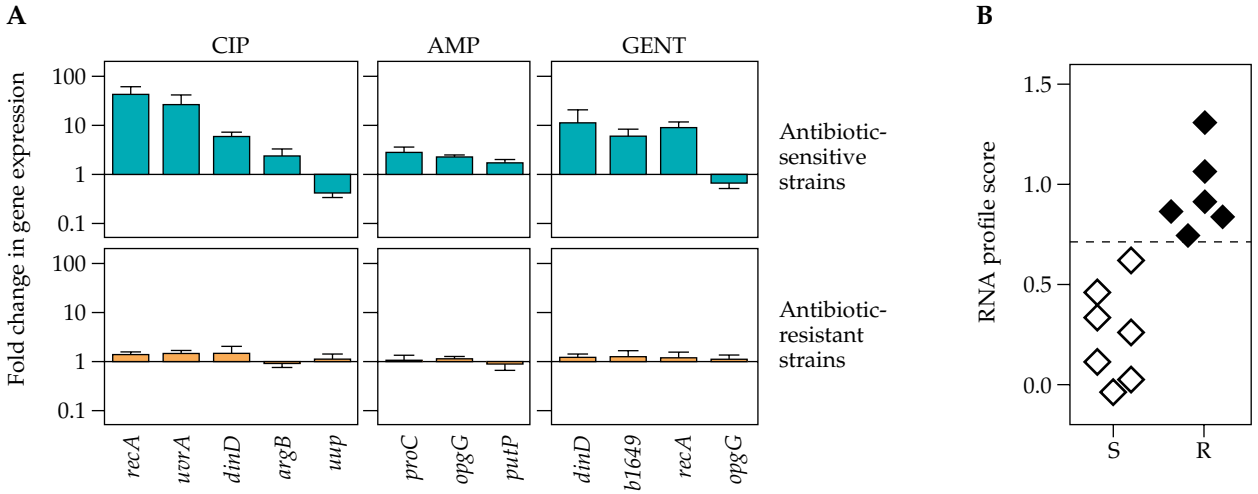
Detection of RNA Signatures of Antibiotic Resistance in Bacteria

Effective treatments for microbial infections are often difficult to predict due to the prevalence of antibiotic-resistant strains of pathogens. For some slow-growing pathogens, for example, *Mycobacterium tuberculosis*, current clinical diagnostic methods that require culturing of the pathogen to determine antibiotic sensitivities can take weeks to months. In the meantime, patients may be treated with ineffective drugs. Furthermore, DNA-based approaches such as hybridization or PCR to detect specific antibiotic resistance genes in bacterial genomes are limited by the available knowledge of these gene sequences, which represents only a fraction of such genes. A possible solution is the use of specific RNA signatures not only for identification of pathogens in clinical samples but also for rapid determination of their antibiotic susceptibilities. Rather than detection of specific antibiotic resistance genes, which is limited to known sequences, this approach detects RNA profiles that are associated

with an antibiotic susceptibility response. Exposure of susceptible strains to an antibiotic induces a characteristic transcriptional response that is antibiotic specific. Direct detection of RNA using fluorescent hybridization probes therefore offers a rapid, sensitive, and specific method for distinguishing antibiotic-resistant from antibiotic-sensitive strains in clinical samples.

RNA profiles associated with susceptibility to the antibiotics ciprofloxacin, ampicillin, and gentamicin were determined following a brief exposure of susceptible and resistant laboratory strains of *E. coli*. Crude cell lysates from cultured bacteria were hybridized to a pool of oligonucleotide probes that target specific RNA sequences; each gene-specific probe was tagged with a different fluorescent dye. While expression of some genes increased in response to several different antibiotics, a small set of genes exhibited antibiotic-specific expression profiles (Fig. 4.40A). The transcriptional responses were observed in the antibiotic-sensitive strains but not in the resistant strains. To test the efficacy of this approach on clinical samples, urine was obtained from patients suspected of having urinary tract infections. The urine specimens were exposed briefly (30 min) to ciprofloxacin, cells in the urine were lysed rapidly, and transcriptional profiles were determined by hybridization of probes directly to RNA. RNA signatures

Figure 4.40 RNA signatures of antibiotic susceptibility. **(A)** Antibiotic-sensitive and -resistant strains of *E. coli* were exposed to the antibiotics ciprofloxacin (CIP), ampicillin (AMP), and gentamicin (GENT) for a short time (10 min). Levels of expression of several genes were measured by hybridizing RNA from antibiotic-treated and untreated cultures to oligonucleotide probes labeled with fluorescent dyes (a subset of antibiotic-responsive genes are labeled on the *x* axis). Changes in the levels of expression of each gene were determined by comparing RNA levels in antibiotic-treated and untreated cultures. Significant changes in gene expression were observed in the antibiotic-sensitive strains but not in the antibiotic-resistant strains. **(B)** Ciprofloxacin-resistant (R) and -sensitive (S) strains of *E. coli* were differentiated in *E. coli*-positive urine samples from 13 patients based on RNA profile scores. For each patient sample, a single RNA profile score was calculated from the transcript levels of the individual antibiotic-responsive genes present in *E. coli*. A low score indicates a susceptible strain, while a high score indicates a resistant strain. From Barczak et al., *Proc. Natl. Acad. Sci. USA* **109:**6217–6222, 2012.

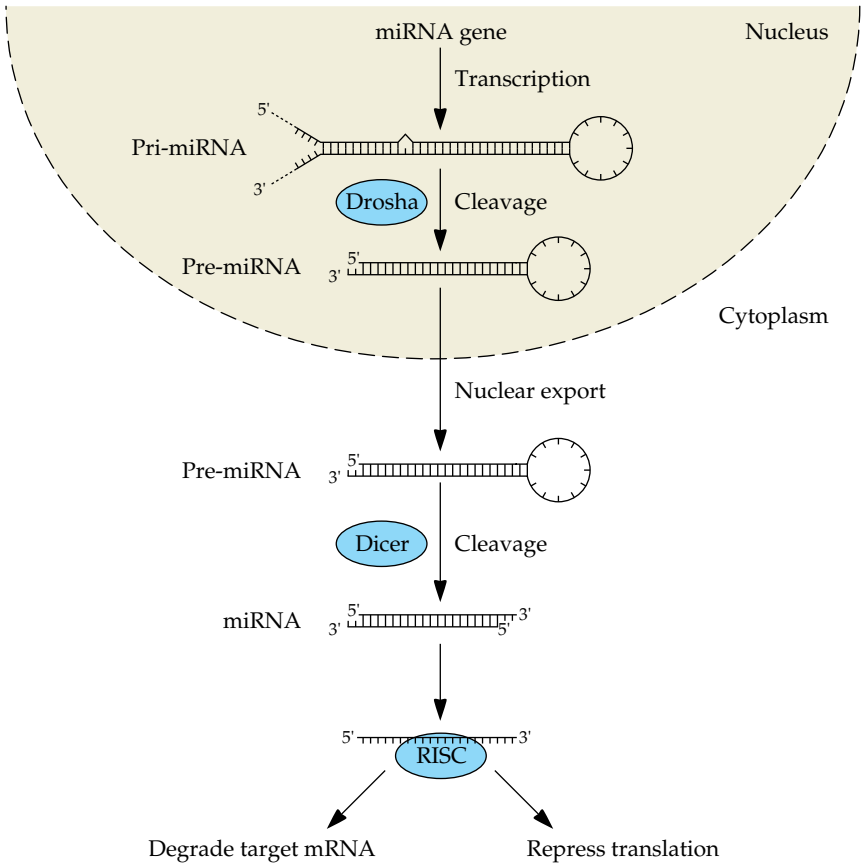


enabled identification of *E. coli* in the samples and discrimination of ciprofloxacin-susceptible and -resistant strains as confirmed by culture-dependent diagnostic methods (Fig. 4.40B). Although similar results could be achieved using PCR-based approaches, direct detection of antibiotic-susceptible specific RNAs via probe hybridization avoids having to purify and amplify the nucleic acids.

Detection of miRNA Signatures of Disease

MicroRNAs (miRNAs) are small (~20-nucleotide) natural RNA molecules that regulate production of specific proteins by preventing translation of mRNA. They are produced from longer primary miRNA molecules that are folded into stem-loop structures and cleaved by the RNases Drosha and Dicer into smaller double-stranded miRNA (Fig. 4.41). The large nuclease complex RNA-induced silencing complex (RISC) separates the strands of the double-stranded miRNA, and the resultant single-stranded RNA products, together with RISC, bind to homologous sequences on

Figure 4.41 Production and activity of miRNA. Double-stranded miRNA is produced from primary miRNA (pri-miRNA) that is folded into a stem-loop structure by intramolecular complementary base-pairing and then cleaved by RNases Drosha and Dicer. RISC mediates binding of single-stranded miRNA to complementary sequences in target mRNAs, which prevents production of the encoded proteins.



mRNA molecules. The nuclease component of RISC then degrades the mRNA, thereby preventing synthesis of the encoded protein. In some cases where the miRNA is imperfectly complementary to its target mRNA, the mRNA is not destroyed but, rather, translation is repressed.

Microarray Analysis of Cancer miRNA Signatures

Some miRNAs play a role in initiation and progression of several human cancers, and therefore, changes in their expression could provide an early indicator of these diseases. Abnormal expression of some miRNAs as a consequence of mutations, epigenetic modifications, or alterations in miRNA processing has been detected in several different tumor types. These miRNAs target mRNA involved in tumorigenesis, such as from Ras oncogenes, antiapoptotic genes, and transcription factors.

B-cell chronic lymphocytic leukemia is a common form of adult leukemia that can progress to a more aggressive lymphoma or leukemia. Affected individuals produce an abnormally high number of lymphocytes (white blood cells). However, these do not function normally; that is, they do not help reduce infections. Moreover, they impair the function of other normal blood cells such as healthy lymphocytes, erythrocytes, and platelets. A microarray containing oligonucleotide probes corresponding to 190 human miRNAs was used to determine miRNA expression profiles in 144 patients with chronic lymphocytic leukemia. Expression levels of 13 miRNAs were found to correlate with disease development (prognosis and progression). For example, levels of two miRNAs (*miR-16-1* and *miR-15a*) were lower in patients with a good prognosis and slow disease progression, while lower levels of another miRNA (*miR-29*) were associated with a poor prognosis. The target gene for *miR-29* is the oncogene *TCL1*, which is overexpressed in patients with more aggressive chronic lymphocytic leukemia. These miRNAs may be early predictors of disease development in patients with this malignancy.

Identification of Alzheimer's Disease miRNAs Using Hybridization Probes

Alzheimer's disease is a progressive neurodegenerative brain disease that manifests as dementia decades after the onset of pathogenesis. Neuronal plaques of aggregated amyloid β peptides and neurofibrillary tangles of hyperphosphorylated tau proteins are hallmarks of Alzheimer's disease (Fig. 4.14). Dysregulated immune responses, including sustained activation of brain macrophage (microglia) and chronic brain inflammation, contribute to the disease, and extracellular vesicles containing cytokines and miRNAs released from brain cells may perturb intercellular signaling and increase neuroinflammation. The levels of specific miRNAs circulating in extracellular vesicles are the basis of a blood test to assess Alzheimer's disease progression. To develop the test, extracellular vesicles were obtained by centrifugation of patients' blood samples, total RNA was extracted, and each of 800 different miRNAs was measured using two oligonucleotide hybridization probes (Fig. 4.42). One biotin-labeled capture probe contains 50 nucleotides that bind to a complementary sequence of an miRNA. The second, reporter probe binds to a 50-nucleotide complementary sequence on the miRNA immediately adjacent to the capture probe. Each miRNA-specific reporter probe is labeled with a

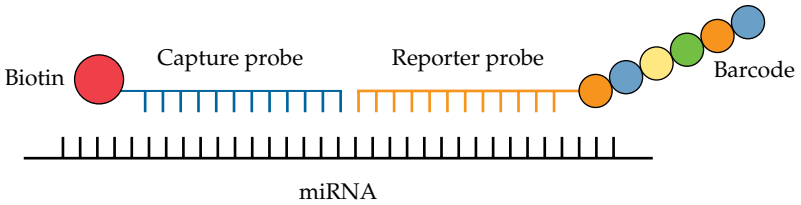
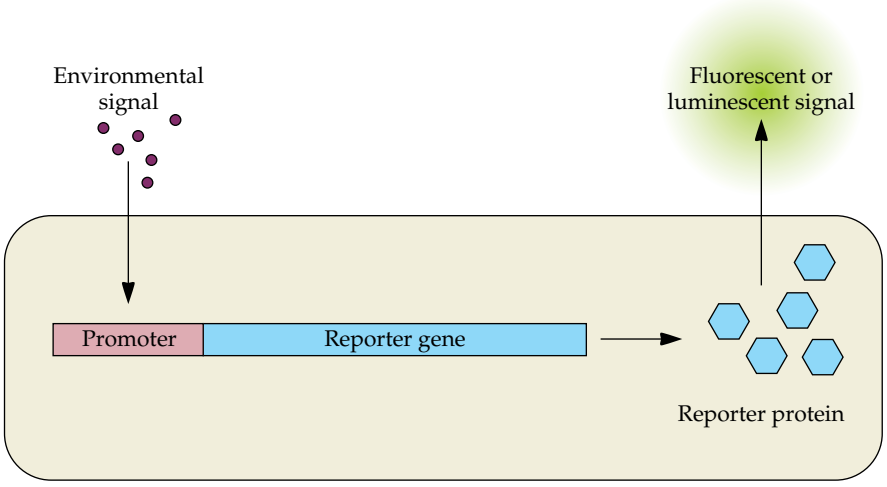


Figure 4.42 Hybridization probes to quantify RNA. In one technology, two oligonucleotide probes hybridize to each of the target RNAs, such as miRNAs, in a sample. The reporter probe binds to a complementary sequence on the RNA and is labeled with a barcode that has a unique combination of fluorophores to identify and quantify the bound RNA. The capture probe binds to an adjacent sequence on the target RNA and is labeled with a biotin molecule so that the probe-target RNA complex can be captured on a streptavidin-coated slide. The captured complexes are scanned with a laser to identify and quantify the RNA molecules.

barcode that has a unique fluorescent color combination to identify the bound miRNA. The capture and reporter probes are mixed with the extracted RNA, and hybridization occurs in solution. After unhybridized probes are washed away, the miRNA-probe complexes are captured on a streptavidin-coated slide by the binding of streptavidin to the biotin molecules on the capture probes, and the fluorescent barcodes are read to count individual miRNA molecules. Lower levels of five miRNAs (*let-7g*, *miR126-3*, *miR142-3*, *miR146a*, and *miR223-3*) correlated with severity of disease, which was confirmed by quantitative reverse transcriptase PCR. Downregulation of these miRNAs was previously shown to be associated with an increase in some proinflammatory cytokines and inflammatory responses of macrophage. Quantification of specific miRNAs in circulating extracellular vesicles may serve as a noninvasive test to monitor Alzheimer's disease progression. The capture-reporter probe technology is available as a commercial assay to measure miRNA and protein-coding gene expression in Alzheimer's disease, cancers, and other diseases.

Figure 4.43 A bacterial cell, engineered to respond to a particular environmental compound, produces a detectable fluorescent or bioluminescent signal.



Biofluorescent and Bioluminescent Systems

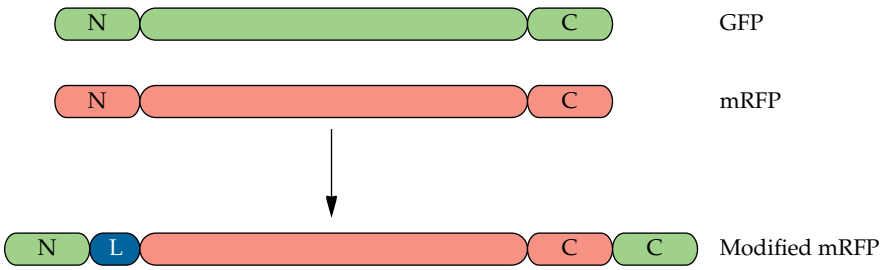
Proteins that naturally fluoresce or luminesce, or that can be easily induced to do so, may be used as biological reporters in a variety of ways. For example, genes encoding these bioreporter proteins may be used to engineer cells to produce a measurable signal in response to a particular chemical or physical agent in their environment. In one version of this system, a gene for a fluorescent or luminescent protein is placed under the control of a promoter that responds to certain environmental signals so that when the promoter is activated, a fluorescent or luminescent signal is produced (Fig. 4.43).

Fluorescent Proteins

The 238-amino-acid-long photoprotein green fluorescent protein, isolated from the jellyfish *Aequorea victoria*, fluoresces green when it is exposed to ultraviolet light. Expression of green fluorescent protein by genetically engineered cells allows intact living cells to be monitored in real time. The use of this reporter molecule has revolutionized fluorescence microscopy. Among its many uses, researchers have used green fluorescent protein to monitor tumor cells in gene therapy protocols, to assess the responses of specific cell types to various therapeutic drugs and treatments, to monitor protein–protein interactions, and to monitor the fates of individual proteins in different therapies.

Following the discovery and subsequent successful employment of green fluorescent protein, scientists began to search, both in nature and by directed mutagenesis, for other colored fluorescent proteins. Having multiple colored fluorescent proteins would enable several biological processes to be monitored at the same time. For the practical use of these proteins, it is essential that they be as stable and as bright as possible. One problem with many of the naturally occurring colored fluorescent proteins is that they often have a tendency to form homodimers or homotetramers. Such multimeric structures can adversely influence the subcellular localization of the proteins, potentially leading to intracellular aggregation and other artifacts. One group of researchers isolated a gene for a red fluorescent protein from *Discosoma* coral and, by means of multiple random mutations, generated a mutant that existed exclusively as a monomer instead of as a tetramer. With each iterative cycle of random

Figure 4.44 Construction of a modified monomeric red fluorescent protein (modified mRFP). The regions of the gene encoding the N and C termini of the green fluorescent protein (GFP) were spliced onto the gene for mRFP following the removal of the portion of the mRFP gene encoding the N terminus and the addition of an oligonucleotide encoding a short linker peptide (L).



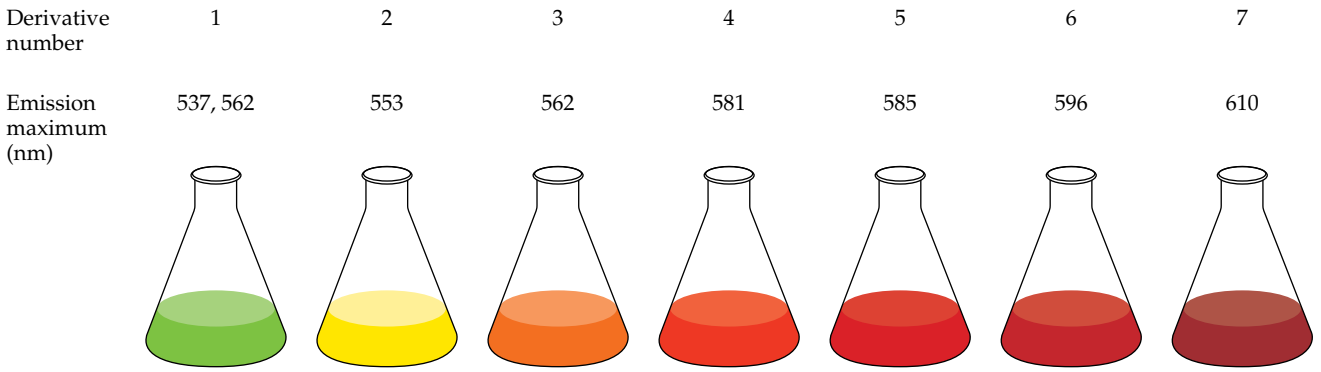


Figure 4.45 Various colored monomeric fluorescent proteins derived from monomeric red fluorescent protein showing their emission wavelengths and color maxima following excitation. The colors have been called, from derivative 1 through 7, honeydew, banana, orange, tomato, tangerine, strawberry, and cherry, respectively. Data from Shaner et al., *Nat. Biotechnol.* **22:**1567–1572, 2004.

mutagenesis, proteins that yielded red fluorescence that was both bright and stable were selected. The production of monomeric red fluorescent protein required 33 mutations. This success notwithstanding, monomeric red fluorescent protein had several drawbacks compared to the native tetrameric form of the protein, including decreased brightness and reduced photostability.

To address some of the above-mentioned problems with monomeric red fluorescent protein, as well as to expand the repertoire of available fluorescent proteins, additional rounds of mutagenesis were performed. First, DNA encoding seven amino acids from the N terminus of green fluorescent protein was added to the gene for red fluorescent protein (Fig. 4.44). Then, DNA for the six amino acids from the green fluorescent protein C terminus was added to the gene for the red fluorescent protein. This construct then became the starting point for several additional rounds of random mutagenesis and directed evolution. Eventually, seven different monomeric colored fluorescent proteins were produced (Fig. 4.45). There is no one best colored fluorescent protein. Some are brighter than others, some are more photostable, and some are more sensitive to changes in pH. Thus, various fluorescent proteins may be used for different applications. Using two or three of these proteins, it is possible to label several different cellular components or cell types at once, thereby increasing the utility of this approach.

Luciferase

The luciferase enzyme, which catalyzes a light-emitting reaction, is produced by a variety of different organisms, including bacteria, algae, fungi, jellyfish, insects, shrimp, and squid. Luciferase genes from bacteria are termed *lux* genes, while those from other organisms—the most widely studied and utilized being the firefly—are termed *luc* genes. The *lux* system is encoded by five genes, *luxCDABE*, and produces a peak of light at 490 nm. In some applications, all five *lux* genes are utilized as a means of monitoring the presence and concentrations of

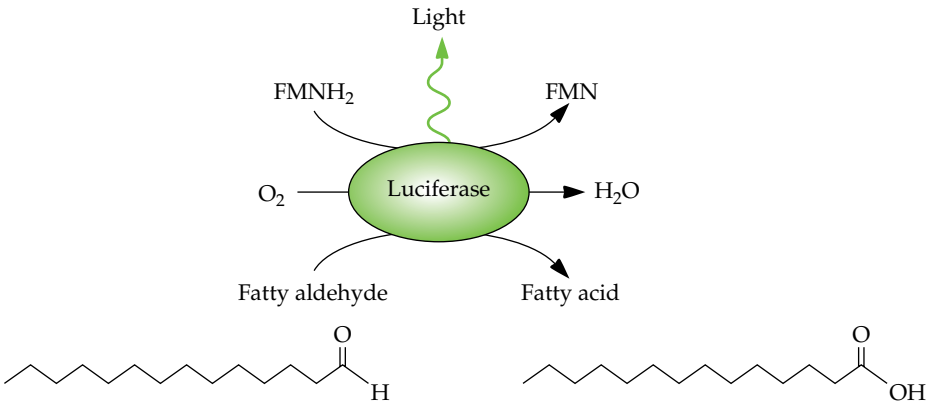


Figure 4.46 Bioluminescence reaction catalyzed by bacterial luciferase. Luciferase, encoded by *luxAB*, catalyzes the oxidation of a long-chain fatty aldehyde and flavin mononucleotide ($FMNH_2$) concomitant with the reduction of oxygen (O_2). Energy is released from the reaction as blue-green light (λ_{max} ~490 nm). Long-chain fatty aldehyde substrates with eight or more carbons (tetradecanal is shown) are synthesized by a fatty acid reductase complex encoded by *luxCDE*.

various contaminants in the environment. When all five genes are used, the light-generating system does not require the addition of any other compounds because the fatty acid reductase complex produced from *luxCDE* catalyzes the synthesis of the aldehyde substrate for luciferase, which is encoded by *luxAB* (Fig. 4.46). Following the addition of an environmental sample to bacterial cells carrying *luxCDABE* under the control of a contaminant-responsive promoter, luciferase and its substrate are produced if the contaminant is present. A quantifiable amount of bioluminescence is produced within a period of a few minutes to no more than a few hours. In some cases, the reporter system includes only the *luxAB* genes. This system requires that the substrate decanal be added during the assay procedure.

The firefly luciferase-catalyzed reaction results in the production of light at 550 to 575 nm. The system requires the addition of the low-molecular-weight organic compound luciferin as a substrate for the light reaction and ATP. Typically, *luc* genes are used in conjunction with eukaryotic cells.

Microbial Biosensors

There is a need for methods that can easily and rapidly detect the large numbers of potentially toxic compounds that contaminate the environment, such as organic compounds, including phenol, salicylate, benzene, trichloroethylene, xylene, toluene, and ethylbenzene, or metals, including cobalt, copper, iron, lead, mercury, nickel, and zinc. Once the contaminated sites have been identified and their range has been delineated, there are a number of highly sophisticated analytical techniques available to identify and quantify specific pollutants. Bacteria that are constitutively bioluminescent (i.e., the bioluminescence does not need to be induced) are good candidates for use as microbial biosensors to detect pollutants. In the presence of pollutants, the bioluminescence decreases, providing a clear indication of the presence of the pollutants. Naturally bioluminescent bacteria, such as the marine

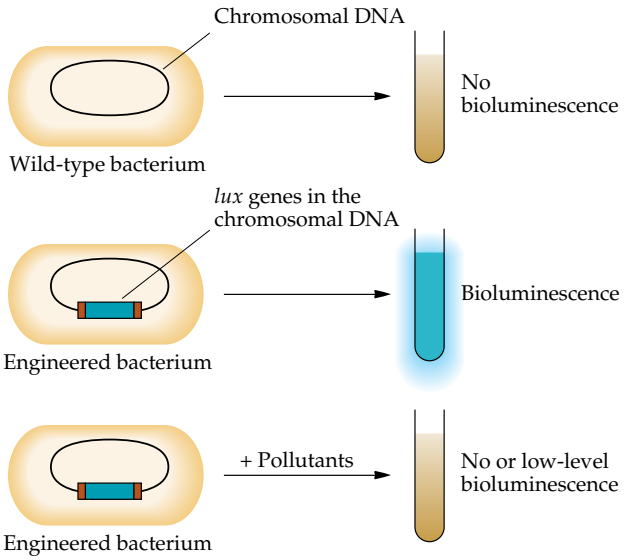


Figure 4.47 Assaying for the presence of pollutants with genetically engineered bioluminescent *P. fluorescens*.

bacterium *Vibrio fischeri*, require saline conditions and a particular pH range and are therefore not useful for testing terrestrial groundwater. To develop a suitable microbial biosensor, the *luxCDABE* genes from this bacterium may be inserted into random sites in the chromosome of a soil bacterium, such as *Pseudomonas fluorescens*. The inserted *lux* genes do not contain a transcriptional promoter, so after insertion into the chromosome of *P. fluorescens*, the only luminescent colonies are those in which the *lux* genes are inserted downstream from a constitutive *P. fluorescens* promoter (without disrupting any important bacterial genes). The cells that luminesce to the greatest extent and have a growth rate similar to that of the wild-type strain are selected for testing with various environmental pollutants. To screen for the presence of various pollutants, a water sample is mixed with a suspension of bioluminescent *P. fluorescens*, and after a 15-minute incubation together, the luminescence of the suspension is measured in a luminometer (Fig. 4.47). When a test sample contains a low to moderate level of certain pollutants, the cell luminescence is inhibited, presumably because the pollutant directly interferes with bacterial metabolism. Since this procedure is rapid, simple, and inexpensive, it is a good first screen for assessing the presence of pollution at a particular site. After a positive response with a bacterial biosensor, the actual pollutants can be determined by other methods.

In the United States, it has been estimated that there are approximately 87,000 different chemical compounds that need to be tested for estrogenic activity, steroid-like activity that can disrupt the endocrine system in vertebrates. While a number of different methods exist that could be used to test these compounds, they are too slow for this sort of large-scale screening. Therefore, scientists have developed a simple and sensitive system for the rapid initial screening of these 87,000 compounds. With this system, yeast (*Saccharomyces cerevisiae*) cells have been genetically engineered to produce measurable quantities of light in the presence of extremely low levels of estrogenic compounds (Fig. 4.48). Using this method, in which

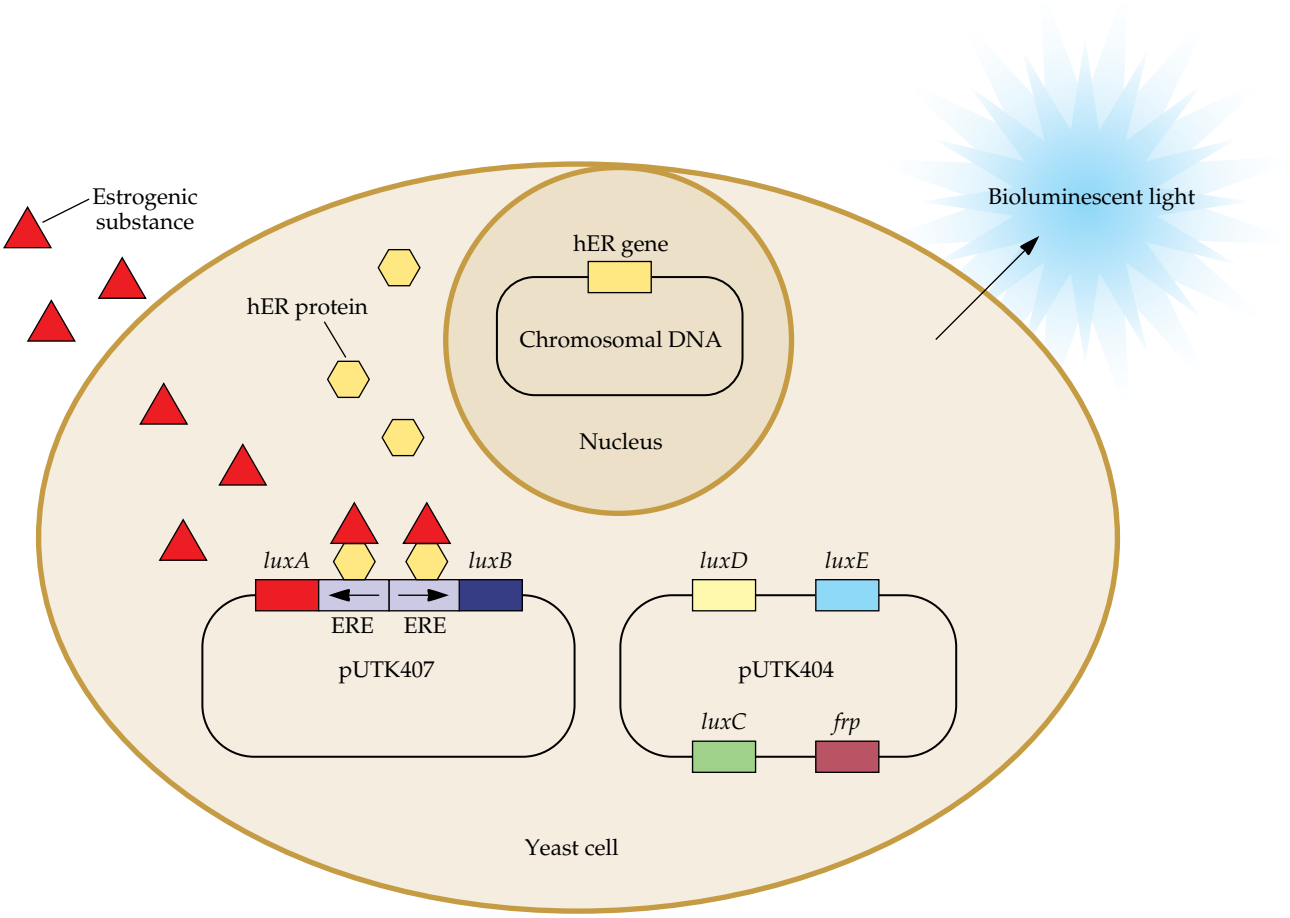


Figure 4.48 A strain of the yeast *Saccharomyces cerevisiae* that has been engineered to detect low levels of estrogenic substances in the environment. Each yeast cell contains the human estrogen receptor (hER) gene integrated into its chromosomal DNA, one plasmid (pUTK404) that contains constitutively expressed bacterial *luxCDE* genes and the flavin oxidoreductase (*frp*) gene from the bacterium *Vibrio harveyi*, and one plasmid (pUTK407) that contains *luxA* and *luxB* genes under the regulatory control of estrogen response elements (ERE). Following the binding of the hER protein to an estrogenic compound, the complex binds to an ERE and activates *luxAB* transcription. Expression of the *luxAB*, *luxCDE*, and *frp* genes leads to the production of a measurable bioluminescent signal. Adapted from Sanseverino et al., *Appl. Environ. Microbiol*. **71:**4455–4460, 2005.

luminescence is induced, to test for estrogenic compounds, light production could be detected in as little as 1 hour. After 6 hours of incubation, the assay attained maximum bioluminescence when the engineered yeast cells were exposed to as little as 5×10^{-11} M 17 β-estradiol, a common estrogen. Many estrogenic compounds required higher concentrations in order to be detected. The main drawback of this approach is that the yeast cell wall and transport system facilitate the entry of some compounds into the cell and inhibit the uptake of other compounds. This can skew the results

and in some instances may suggest that a compound is not estrogenic when it is unable to efficiently enter yeast cells. Nevertheless, this technique is likely to identify a large number of estrogenic compounds that were deemed nonestrogenic until they were tested with this protocol.

At the same time that some groups of scientists are working to develop and perfect cells as biosensors, others have focused their efforts on automating these systems. Such an automated system might include genetically engineered cells that emit blue-green light (~490 nm) in response to specific compounds, an environment that sustains the cells, a light-tight enclosure, and an integrated circuit for light detection and signal processing. While simple prototypes have been constructed, the difficulties associated with utilizing living cells remain major impediments.

summary

A major goal of a molecular diagnostic test is to detect specific biological molecules (biomarkers) associated with infectious or genetic disease in patient samples or with contamination in environmental samples. The biomarkers may be specific proteins, DNA or RNA sequences, or metabolites that are present, or present at altered levels, in body fluids or tissues of affected individuals compared to healthy individuals, or in samples from contaminated but not pristine environments. The purpose of a clinical diagnostic test is to identify or confirm the cause of a disease (etiology), to determine risk of developing the disease in asymptomatic individuals, to determine the best course of treatment, or to monitor progression of the disease or response to treatment. Therefore, the test must have a high degree of specificity (i.e., detect only the targeted molecule) and sensitivity (i.e., detect the targeted molecule even when it is present in low abundance). Importantly, a diagnostic test must also be rapid, inexpensive, and amenable to high throughput.

Molecular technologies have greatly facilitated accurate detection of biomarkers. Immunological approaches exploit antibody–antigen interactions to detect the presence of a pathogen or antibodies against a pathogen, proteins that are dysregulated or have altered conformation, or antibodies against self-proteins (autoantibodies). Two common approaches to detect protein biomarkers of disease are ELISAs and protein microarrays. An indirect ELISA detects the presence of specific antibodies (primary antibodies) in patient serum by their binding to a purified antigen immobilized on a solid support. The antibody–antigen interaction is detected by addition of a labeled secondary antibody. This assay may be used to determine if a patient has mounted an autoimmune response or an immune response to a pathogen. A sandwich ELISA measures levels of an antigen, often a protein associated with a disease or particular condition, in a patient's sample. In this assay, specific (usually monoclonal) antibodies are bound to a solid support to which a complex clinical sample is applied. If the cognate antigen is present in the sample, it is captured by the immobilized antibody and can be detected by applying a labeled antibody.

Microarrays may be used to detect multiple protein biomarkers in clinical samples from patients affected by polygenic diseases or the presence of IgE molecules produced against specific allergens in patients who suffer from allergies. For detection of protein biomarkers, primary antibodies are bound to a solid support (e.g., small capture beads) and then probed with a patient's sample. Secondary antibodies, often labeled with a fluorescent dye, are added to detect captured antigens. For detection of specific antibodies in patient serum, purified proteins or other antigens (e.g., allergens) are arrayed and then probed with a serum sample.

Nucleic acid-based diagnostic tests determine the existence of specific nucleotide sequences, including genetic mutations and sequences present in pathogens. In addition to being highly sensitive and specific, these tests can be used to assess risk for a disease before the onset of symptoms. Specific DNA or RNA sequences can be detected in patient samples by hybridization, PCR, microarray analysis, CRISPR-Cas-guided cleavage, and whole-genome or RNA sequencing. Hybridization utilizes labeled oligonucleotide probes to detect a complementary target DNA sequence that indicates the presence of a specific microorganism or a specific disease-associated allele (SNP). Oligonucleotide probes

(continued)

may also be employed to detect and quantify RNA molecules such as specific mRNAs and miRNAs that indicate dysregulation of gene expression in diseased tissue or a pathogen trait. Perfect complementarity between a target sequence that includes a polymorphic site and a probe is required for ligation of two tandem probes in an OLA and for ligation of two ends of a padlock probe. Hybridization of a CRISPR-based guide RNA to a target DNA or RNA sequence activates Cas12 or Cas13 endonuclease, respectively, to cleave a reporter probe, which then fluoresces.

Allele-specific PCR requires a perfect match between an oligonucleotide primer and a target sequence for amplification. One of the pair of PCR primers includes a nucleotide that is complementary to an SNP at its 3′ end. Mismatches prevent primer extension by the thermostable DNA polymerase. Several variations of allele-specific PCR are employed in diagnostic assays, including the TaqMan protocol, which measures the activity of a fluorescent dye in real time as it is released during primer extension. Taq polymerase, which catalyzes the synthesis of the DNA strand, possesses nuclease activity and cleaves a probe that is also bound to the template. Cleavage removes a quencher molecule that prevents fluorescence of the dye. An alternative real-time PCR protocol measures fluorescence of a molecular beacon that specifically binds to target DNA that is produced after each amplification cycle. Because the amount of double-stranded DNA doubles after each cycle, the target DNA can be quantified in a sample, such as a sample containing an infectious agent. Droplet digital PCR amplifies a single template molecule within an oil droplet and is useful to quantify very low-abundance sequences. For point-of-care diagnostic tests, target sequences may be amplified in LAMP reactions that are carried out at a single temperature and visualized by eye, which reduces equipment costs and the time to acquire a test result. When the target is an RNA molecule such as the genome of an RNA virus or transcripts from an expressed gene, the RNA is converted to cDNA using reverse transcriptase prior to PCR amplification.

Microarray analyses enable simultaneous detection of SNPs at multiple loci for polygenic diseases or determination of RNA signatures associated with a disease. For example, DNA microarrays have been approved by the U.S. FDA to detect cytochrome P450 oxidase alleles associated with drug metabolism (AmpliChip CYP450 microarray) or to measure expression of specific genes in breast tumors (MammaPrint microarray). The results from these tests can be used to determine an effective course of treatment for the disease. Microarray analysis of SNPs among individuals can also be used to determine ancestry.

One of the goals of the next-generation sequencing technologies is personalized genomic medicine, that is, the availability of genome sequencing to assist in disease diagnosis. The detection of disease-associated alleles will lead to recommendations for treatment and, where the genetic risk of a disease is detected before symptoms are present, disease prevention strategies that are tailored to individual genotypes. A major limitation of this approach is our current understanding of the genetic variants that contribute to human diseases.

In addition to detection of human genetic and infectious diseases, molecular diagnostics has applications in agriculture, environmental monitoring, and forensics. For example, microorganisms engineered to express fluorescent or luminescent proteins in response to a chemical toxin can be used to monitor pollutants in environmental samples. Forensic analysis of microsatellite DNA fingerprints in biological samples found at a crime scene can aid in solving crimes, and analysis of SNPs in animal tissues may indicate illegal possession of wild animals. The specificity, sensitivity, and amenability to high throughput have made molecular technologies integral diagnostic tools in many applications.

REFERENCES

Aharon A, **Spector P**, **Ahmad RS**, **Horrany N**, **Sabbach A**, **Brenner B**, **Aharon-Peretz J.** 2020. Extracellular vesicles of Alzheimer's disease patients as a biomarker for disease progression. *Mol Neurobiol* **57:**4156–4169.

Aqeilan RI, **Calin GA**, **Croce CM.** 2010. *miR-15a* and *miR-16-1* in cancer: discovery, function and future perspectives. *Cell Death Differ* **17:**215–220.

Banér J, **Nilsson M**, **Isaksson A**, **Mendel-Hartvig M**, **Antson DO**, **Landegren U.** 2001. More keys to padlock probes: mechanisms for high-throughput nucleic acid analysis. *Curr Opin Biotechnol* **12:**11–15.

Barczak AK, **Gomez JE**, **Kaufmann BB**, **Hinson ER**, **Cosimi L**, **Borowsky ML**, **Onderdonk AB**, **Stanley SA**, **Kaur D**, **Bryant KF**, **Knipe DM**, **Sloutsky A**, **Hung DT.** 2012. RNA signatures allow rapid identification of pathogens and antibiotic susceptibilities. *Proc Natl Acad Sci USA* **109:**6217–6222.

Bibikova M, **Lin Z**, **Zhou L**, **Chudin E**, **Garcia EW**, **Wu B**, **Doucet D**, **Thomas NJ**, **Wang Y**, **Vollmer E**, **Goldmann T**, **Seifart C**, **Jiang W**, **Barker DL**, **Chee MS**, **Floros J**, **Fan JB.** 2006. High-throughput DNA methylation profiling using universal bead arrays. *Genome Res* **16:**383–393.

Bolnick DA, **Fullwiley D**, **Duster T**, **Cooper RS**, **Fujimura JH**, **Kahn J**, **Kaufman JS**, **Marks J**, **Morning A**, **Nelson A**, **Ossorio P**, **Reardon J**, **Reverby SM**, **TallBear K.** 2007. The science and business of genetic ancestry testing. *Science* **318:**399–400.

Broughton JP, **Deng X**, **Yu G**, **Fasching CL**, **Servellita V**, **Singh J**, **Miao X**, **Streithorst JA**, **Granados A**, **Sotomayor-Gonzalez A**, **Zorn K**, **Gopez A**, **Hsu E**, **Gu W**, **Miller S**, **Pan CY**, **Guevara H**, **Wadford DA**, **Chen JS**, **Chiu CY.** 2020. CRISPR-Cas12-based detection of SARS-CoV-2. *Nat Biotechnol* **38:**870–874.

Byron SA, **Van Keuren-Jensen KR**, **Engelthaler DM**, **Carpten JD**, **Craig DW.** 2016. Translating RNA sequencing into clinical diagnostics: opportunities and challenges. *Nat Rev Genet* **17:**257–271.

Calin GA, **Ferracin M**, **Cimmino A**, **Di Leva G**, **Shimizu M**, **Wojcik SE**, **Iorio MV**, **Visone R**, **Sever NI**, **Fabbri M**, **Iuliano R**, **Palumbo T**, **Pichiorri F**, **Roldo C**, **Garzon R**, **Sevignani C**, **Rassenti L**, **Alder H**, **Volinia S**, **Liu CG**, **Kipps TJ**, **Negrini M**, **Croce CM.** 2005. MicroRNA signature associated with prognosis and progression in chronic lymphocytic leukemia. *N Engl J Med* **353:**1793–1801.

Campbell RE, **Tour O**, **Palmer AE**, **Steinbach PA**, **Baird GS**, **Zacharias DA**, **Tsien RY.** 2002. A monomeric red fluorescent protein. *Proc Natl Acad Sci USA* **99:**7877–7882.

Casén C, **Vebø HC**, **Sekelja M**, **Hegge FT**, **Karlsson MK**, **Ciemniejewska E**, **Dzankovic S**, **Frøyland C**, **Nestestog R**, **Engstrand L**, **Munkholm P**, **Nielsen OH**, **Rogler G**, **Simrén M**, **Öhman L**, **Vatn MH**, **Rudi K.** 2015. Deviations in human gut microbiota: a novel diagnostic test for determining dysbiosis in patients with IBS or IBD. *Aliment Pharmacol Ther* **42:**71–83.

Chen JS, **Ma E**, **Harrington LB**, **Da Costa M**, **Tian X**, **Palefsky JM**, **Doudna JA.** 2018. CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science* **360:**436–439.

Conner BJ, **Reyes AA**, **Morin C**, **Itakura K**, **Teplitz RL**, **Wallace RB.** 1983. Detection of sickle cell β^S-globin allele by hybridization with synthetic oligonucleotides. *Proc Natl Acad Sci USA* **80:**278–282.

Cuzick J, **Swanson GP**, **Fisher G**, **Brothman AR**, **Berney DM**, **Reid JE**, **Stat M**, **Mesher D**, **Speights VO**, **Stankiewicz E**, **Foster CS**, **Møller H**, **Scardino P**, **Warren JD**, **Park J**, **Younus A**, **Flake DD**, **Wagner S**, **Gutin A**, **Lanchbury JS**, **Stone S.** 2011. Prognostic value of an RNA expression signature derived from cell cycle proliferation genes for recurrence and death from prostate cancer: a retrospective study in two cohorts. *Lancet Oncol* **12:**245–255.

Dao Thi VL, **Herbst K**, **Boerner K**, **Meurer M**, **Kremer LPM**, **Kirrmaier D**, **Freistaedter A**, **Papagiannidis D**, **Galmozzi C**, **Stanifer ML**, **Boulant S**, **Klein S**, **Chlanda P**, **Khalid D**, **Barreto Miranda I**, **Schnitzler P**, **Kräusslich H-G**, **Knop M**, **Anders S.** 2020. A colorimetric RT-LAMP assay and LAMP-sequencing for detecting SARS-CoV-2 RNA in clinical samples. *Sci Transl Med* **12:**eabc7075.

Eshaque B, **Dixon B.** 2006. Technology platforms for molecular diagnosis of cystic fibrosis. *Biotechnol Adv* **24:**86–93.

Ferec C, **Cutting GR.** 2012. Assessing the disease-liability of mutations in CFTR. *Cold Spring Harb Perspect Med* **2:**a009480.

Grover A, **Sharma PC.** 2016. Development and use of molecular markers: past and present. *Crit Rev Biotechnol* **36:**290–302.

Hartmann M, **Roeraade J**, **Stoll D**, **Templin MF**, **Joos TO.** 2009. Protein microarrays for diagnostic assays. *Anal Bioanal Chem* **393:**1407–1416.

Hellström I, **Raycraft J**, **Hayden-Ledbetter M**, **Ledbetter JA**, **Schummer M**, **McIntosh M**, **Drescher C**, **Urban N**, **Hellström KE.** 2003. The HE4 (WFDC2) protein is a biomarker for ovarian carcinoma. *Cancer Res* **63:**3695–3700.

Huletsky A, **Giroux R**, **Rossbach V**, **Gagnon M**, **Vaillancourt M**, **Bernier M**, **Gagnon F**, **Truchon K**, **Bastien M**, **Picard FJ**, **van Belkum A**, **Ouellette M**, **Roy PH**, **Bergeron MG.** 2004. New real-time PCR assay for rapid detection of methicillin-resistant *Staphylococcus aureus* directly from specimens containing a mixture of staphylococci. *J Clin Microbiol* **42:**1875–1884.

How Kit A, **Nielsen HM**, **Tost J.** 2012. DNA methylation based biomarkers: practical considerations and applications. *Biochimie* **94:**2314–2337.

Lao VV, **Grady WM.** 2011. Epigenetics and colorectal cancer. *Nat Rev Gastroenterol Hepatol* **8:**686–700.

Nagamine K, **Hase T**, **Notomi T.** 2002. Accelerated reaction by loop-mediated isothermal amplification using loop primers. *Mol Cell Probes* **16:**223–229.

Nielsen EE, **Cariani A**, **Aoidh EM**, **Maes GE**, **Milano I**, **Ogden R**, **Taylor M**, **Hemmer-Hansen J**, **Babbucci M**, **Bargelloni L**, **Bekkevold D**, **Diopere E**, **Grenfell L**, **Helyar S**, **Limborg MT**, **Martinsohn JT**, **McEwing R**, **Panitz F**, **Patarnello T**, **Tinti F**, **Van Houdt JKJ**, **Volckaert FAM**, **Waples RS**, **Carvalho GR**, **FishPopTrace consortium.** 2012. Gene-associated markers provide tools for tackling illegal fishing and false eco-certification. *Nat Commun* **3:**851.

Nivens DE, **McKnight TE**, **Moser SA**, **Osbourn SJ**, **Simpson ML**, **Sayler GS.** 2004. Bioluminescent bioreporter integrated circuits: potentially small, rugged and inexpensive whole-cell biosensors for remote environmental monitoring. *J Appl Microbiol* **96:**33–46.

Perchiacca JM, **Ladiwala ARA**, **Bhattacharya M**, **Tessier PM.** 2012. Structure-based design of conformation- and sequence-specific antibodies against amyloid β. *Proc Natl Acad Sci USA* **109:**84–89.

Perkins BA, **Caskey CT**, **Brar P**, **Dec E**, **Karow DS**, **Kahn AM**, **Hou Y-CC**, **Shah N**, **Boeldt D**, **Coughlin E**, **Hands G**, **Lavrenko V**, **Yu J**, **Procko A**, **Appis J**, **Dale AM**, **Guo L**, **Jönsson TJ**, **Wittmann BM**, **Bartha I**, **Ramakrishnan S**, **Bernal A**, **Brewer JB**, **Brewerton S**, **Biggs WH**, **Turpaz Y**, **Venter JC.** 2018. Precision medicine screening using whole-genome sequencing and advanced imaging to identify disease risk in adults. *Proc Natl Acad Sci USA* **115:**3686–3691.

Pinzani P, Salvianti F, Cascella R, Massi D, De Giorgi V, Pazzagli M, Orlando C. 2010. Allele specific Taqman-based real-time PCR assay to quantify circulating BRAFV600E mutated DNA in plasma of melanoma patients. *Clin Chim Acta* **411:**1319–1324.

Qian J, Lu Z-X, Mancuso CP, Jhuang H-Y, Del Carmen Barajas-Ornelas R, Boswell SA, Ramírez-Guadiana FH, Jones V, Sonti A, Sedlack K, Artzi L, Jung G, Arammash M, Pettit ME, Melfi M, Lyon L, Owen SV, Baym M, Khalil AS, Silver PA, Rudner DZ, Springer M. 2020. Barcoded microbial system for high-resolution object provenance. *Science* **368:**1135–1140.

Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL, ACMG Laboratory Quality Assurance Committee. 2015. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* **17:**405–424.

Sanseverino J, Gupta RK, Layton AC, Patterson SS, Ripp SA, Saidak L, Simpson ML, Schultz TW, Sayler GS. 2005. Use of *Saccharomyces cerevisiae* BLYES expressing bacterial bioluminescence for rapid, sensitive detection of estrogenic compounds. *Appl Environ Microbiol* **71:**4455–4460.

Shaner NC, Campbell RE, Steinbach PA, Giepmans BNG, Palmer AE, Tsien RY. 2004. Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nat Biotechnol* **22:**1567–1572.

Sibley CD, Peirano G, Church DL. 2012. Molecular methods for pathogen and microbial community detection and characterization: current and potential application in diagnostic microbiology. *Infect Genet Evol* **12:**505–521.

Sokolenko AP, Imyanitov EN. 2018. Molecular diagnostics in clinical oncology. *Front Mol Biosci* **5:**76.

Stenson PD, Mort M, Ball EV, Evans K, Hayden M, Heywood S, Hussain M, Phillips AD, Cooper DN. 2017. The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Hum Genet* **136:**665–677.

Tan CW, Chia WN, Qin X, Liu P, Chen MI-C, Tiu C, Hu Z, Chen VC-W, Young BE, Sia WR, Tan Y-J, Foo R, Yi Y, Lye DC, Anderson DE, Wang L-F. 2020. A SARS-CoV-2 surrogate virus neutralization test based on antibody-mediated blockage of ACE2-spike protein-protein interaction. *Nat Biotechnol* **38:**1073–1078.

Tyagi S, Kramer FR. 1996. Molecular beacons: probes that fluoresce upon hybridization. *Nat Biotechnol* **14:**303–308.

van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R, Friend SH. 2002. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415:**530–536.

Weitz HJ, Ritchie JM, Bailey DA, Horsburgh AM, Killham K, Glover LA. 2001. Construction of a modified mini-Tn*5* *lux*CDABE transposon for the development of bacterial biosensors for ecotoxicity testing. *FEMS Microbiol Lett* **197:**159–165.

Westreich R, Neumann Y, Deutsch O, Krief G, Stiubea-Choen R, Zager D. 2020. Development of saliva-based cTNI point-of-care test: a feasibility study. *Eur Heart J* **41:** ehaa946-1693.

review questions

1. What are the goals of a molecular diagnostic test? What are some of the factors that are considered in the development of an accurate diagnostic test?
2. What is a biomarker?
3. For diagnostic assays, what is meant by sensitivity, specificity, and simplicity?
4. Describe the structural features of an antibody molecule that are important for its function.
5. What are monoclonal antibodies? How are they different from polyclonal antibodies?
6. Briefly, explain how the HAT selection for hybridomas works.
7. Compare and contrast indirect and sandwich ELISAs. Describe an application for each in the diagnosis of a human disease.
8. Describe a method to identify the causative agent of an infection.
9. What is the advantage of using a protein microarray in a diagnostic test compared to an ELISA?
10. Why might an immunological approach be more appropriate than a DNA-based approach to diagnose Alzheimer's disease?
11. What is a hybridization probe? How are hybridization probes employed to identify bacteria present in a patient sample? How are they used to detect disease-associated SNPs?
12. Why is PCR commonly used in molecular diagnostic tests?
13. Describe an allele-specific PCR.
14. How can two different alleles be distinguished using a TaqMan assay?
15. Why is real-time PCR quantitative? Why would this be useful for a diagnostic test?

16. How does droplet digital PCR differ from conventional PCR?
17. What is the major advantage of loop-mediated isothermal amplification compared to conventional PCR amplification? How is this achieved?
18. How can the CRISPR-Cas system be used in a diagnostic test?
19. Describe one application for nucleic acid microarrays in disease diagnosis.
20. Describe an application for an SNP microarray.
21. What is DNA methylation? What role does it play in human disease? How can DNA methylation patterns be detected in clinical samples?
22. When assessing the antibiotic susceptibility of a pathogen, what are the advantages of determining RNA profiles rather than detecting the antibiotic resistance gene directly?
23. What are miRNAs? Why are they useful biomarkers of some diseases?
24. You have been given the task of developing a simple, sensitive, and reproducible diagnostic procedure for a double-stranded DNA virus that is devastating a local cattle population. Because effective treatment of this disease depends on early and accurate diagnosis, you need to be able to detect the very low levels of this virus that are present in infected animals before the onset of disease symptoms. Briefly explain how you would proceed and why you have chosen a particular course of action.
25. How would you develop a microbial biosensor to detect environmental contaminants?

Pharmaceuticals
Human Interferons
Human Growth Hormone
Tumor Necrosis Factor Alpha
Extending Protein Half-Life
Enzymes
DNase I
Alginate Lyase
Phenylalanine Ammonia Lyase
α_1-Antitrypsin
Glycosidases
Masking Nonhuman Epitopes
Toxin-Intein Fusions
Targeting Mitochondria
Bacteria and Therapeutics
Interleukin-10
Leptin
An HIV Inhibitor
Insulin
Parkinson's Disease
Cancer and Bacteria
Recombinant Antibodies
Hybrid Human–Mouse Monoclonal Antibodies
Human Monoclonal Antibodies
Antibody Fragments
Combinatorial Libraries of Antibody Fragments
A Combinatorial Library of Full-Length Antibodies
Shuffling CDR Sequences
Dual-Variable-Domain Antibodies
Bispecific Antibodies against Hemophilia
Anti-HIV Antibodies
Anticancer Antibodies
Antibodies against Various Diseases
Antiobesity Antibodies
Enhanced Antibody Half-Life
Affibody Molecules
SUMMARY
REFERENCES
REVIEW QUESTIONS

5 Protein Therapeutics

Prior to the development of recombinant DNA technology, most human protein pharmaceuticals were available in only limited quantities, they were extremely costly to produce, and, in a number of cases, their biological modes of action were not well characterized. When recombinant DNA technology was first developed, it was heralded as a means of producing a wide range of human therapeutic agents in sufficient quantities for both efficacy testing and eventual human use. This forecast has turned out to be true. Today, the "genes" (more accurately, complementary DNAs [cDNAs]) for several tens of thousands of different proteins that are potential human therapeutic agents have been cloned and expressed. Most of these proteins have been produced in mammalian, fungal, or bacterial host cells, and currently a few thousand of these are undergoing clinical testing with human subjects for the treatment of various diseases. Approximately 750 of these "biotechnology drugs" have been approved for use in the United States and/or the European Union (Table 5.1 provides examples of a few of these proteins). However, it will be several years before some of the other new therapeutic proteins are commercially available, because medical products must first be tested rigorously in animals in preclinical trials and then undergo thorough human clinical trials, which generally last for several years, before being approved for human use. The financial incentive for pharmaceutical companies to develop these new drugs is considerable. It has been estimated that in 2019 the annual global market for human recombinant protein drugs was more than $250 billion. By the end of 2017, 57 monoclonal antibodies were in clinical use. Sales of various monoclonal antibodies in 2020 are estimated to be more than $130 billion,

Molecular Biotechnology: Principles and Applications of Recombinant DNA,
Sixth Edition. Bernard R. Glick and Cheryl L. Patten.
© 2022 American Society for Microbiology.

Table 5.1 Examples of some recombinant proteins that have been approved for human use in the United States or the European Union

Alglucosidase α
Anakinra
Antihemophilic factor
Antithrombin III
α1-Antitrypsin
Botulinum toxin A
C1 esterase inhibitor
Darbepoetin α
Dibotermin
DNase I
Drotrecogin α
Erythropoietin
Factor VIIa
Factor VIII
Factor IX
Fibrinogen
Follicle-stimulating hormone
α-Galactosidase
Galsulfase
Glucagon
β-Glucocerebrosidase analogue
α-Glucosidase
Granulocyte-macrophage colony-stimulating factor
Hirudin
Human bone morphogenic protein 7
Human chorionic gonadotropin
Human growth hormone
Human growth hormone analogue
Human luteinizing hormone
Hyaluronidase
Iduronate-2-sulfatase
Insulin
Insulin analogue
Insulin-like growth factor
IFN-α2a
IFN-α2b
IFN-αN3
IFN-β1a
IFN-β1b
IFN-γ1b
IFN N
IFN analogue
Interleukin-2
Interleukin-2 analogues
Interleukin-11
Interleukin-11 analogue
Kallikrein inhibitor
Keratinocyte growth factor
Laronidase
Novel erythropoiesis-stimulating protein
Osteogenic protein
Platelet-derived growth factor
Stem cell factor
Tissue plasminogen activator
Thyrotropin-α
Truncated tissue plasminogen activator

with 12 "blockbuster" drugs, each with sales of more than $3 billion, constituting nearly half of these sales. These blockbuster drugs include adalimumab (Humira), with sales of nearly $19 billion in 2017; a monoclonal antibody to treat rheumatoid arthritis, rituximab (Rituxan); a monoclonal antibody used to treat individuals with non-Hodgkin's lymphoma; various forms of recombinant human insulin; and trastuzumab (Herceptin), a monoclonal antibody used to treat certain types of breast cancer. Moreover, for the past 10 years, sales of human recombinant protein drugs have been increasing at a rate of around 8 to 10% a year.

Pharmaceuticals

Protein therapeutic agents may be divided into several different categories based on their general mode. A large number of the currently available protein therapeutic agents that are produced by recombinant DNA technology replace a deficient or abnormal protein. Some classic examples of replacing a deficient protein include providing diabetics with insulin or hemophiliacs with specific blood coagulation factors. Other recombinant therapeutics may augment an existing pathway or provide a novel function or activity.

A number of different strategies have been used to isolate the cDNAs for human proteins. More recently, the development of technologies to rapidly and inexpensively determine the genome sequence of human and other organisms (see chapter 2) has dramatically increased our ability to isolate or chemically synthesize disease-associated genes from humans and disease-causing organisms.

Human Interferons

Interferons (IFNs) are animal glycoproteins that are synthesized by cells in response to infection with various pathogens, especially viruses. IFNs "interfere" with viral replication within host cells. IFNs can also activate immune cells, such as natural killer cells and macrophages, increase recognition of infected or tumor cells by upregulating antigen presentation to T lymphocytes, and increase the ability of uninfected host cells to resist a new infection by a virus. There are several different types of IFNs, all with slightly different biological activities. For example, the antiviral activities of human IFN-α2 and human IFN-α1 are approximately the same when measured in a virus-challenged bovine cell line, but IFN-α2 is seven times more effective than IFN-α1 when human cells are treated with virus. IFN-α2 is 30 times less effective than IFN-α1 when mouse cells are used in this assay. IFNs may induce the production of hundreds of other proteins, upregulate major histocompatibility molecules, activate cells that are part of the immune system, promote apoptosis, and inhibit protein synthesis of both viral and host genes. At the present time, 12 different IFNs and IFN derivatives have been approved by the U.S. Food and Drug Administration (FDA) for the treatment of various diseases, including multiple sclerosis, hepatitis B, hepatitis C, various types of cancer, and herpes simplex virus (see Milestone).

Several research groups have engineered IFNs with combined properties of different members of the IFN-α gene family (very similar proteins encoded by different genes) that vary in the extents and specificities of their antiviral activities. This was achieved by splicing a portion of one IFN-α

Synthesis in *E. coli* of a Polypeptide with Human Leukocyte Interferon Activity

milestone

S. Nagata, H. Taira, A. Hall, L. Johnsrud, M. Streuli, J. Ecsödi, W. Boll, K. Cantell, and C. Weissmann

Nature **284**:316–320, 1980

The late 1970s and early 1980s were a time of tremendous excitement in molecular biotechnology. The promise of this new technology was being touted to both the public and large institutional investors. One of the products of the new biotechnology that captured the imagination of a large number of people was IFN, which at the time was seen by many as a possible miracle cure for a wide range of diseases caused by viruses and cancer. Thus, the isolation of a human IFN cDNA and its subsequent expression in *E. coli* were reported in newspapers and magazines around the world.

Several features of IFN made it particularly difficult to synthesize and isolate a cDNA encoding the polypeptide. First, although IFN had been purified more than 80,000-fold, only minuscule amounts were available, so researchers did not even have an accurate estimate of its molecular mass. Second, unlike many other proteins, IFN did not have a chemical or biological activity that was easy to monitor. At the time, its activity was measured by the reduction in the cytopathic effect of an animal virus on cells in culture, which was an extremely complex and time-consuming process. Third, unlike insulin, researchers had no idea if there was one particular human cell that produced high levels of IFN and therefore could serve as a source of mRNA that was enriched for IFN mRNA. These problems notwithstanding, a cDNA encoding IFN was eventually isolated and characterized. Since that time, researchers have discovered that there are several different types of IFNs. Unfortunately, IFN is not the panacea that was dreamed of by both investors and the press. However, the genes for several IFNs have been isolated, and clinical trials have shown that they are effective treatments for a variety of viral diseases.

after translation, a hybrid protein that exhibits novel properties, i.e., properties different from either of the contributing genes.

In one study, portions of the genes encoding IFN-α2 and IFN-α3 were combined to create a protein with novel IFN activities. Comparison of the sequences of the two IFN-α cDNAs indicated that they had common restriction sites at corresponding positions in the nucleotide sequences. Digestion of both cDNAs at these sites and ligation of the DNA fragments yielded a number of hybrid derivatives of the original genes (Fig. 5.1A). These hybrids were expressed in *Escherichia coli*, and the resultant proteins were purified and examined for various biological functions. When tested for the extent of protection of mammalian cells in culture against viral infection, some of the hybrid IFNs were found to have greater activity than either of the parental molecules. In addition, many of the hybrid IFNs induced test cells to synthesize (2′-5′)-oligoisoadenylate synthetase. This enzyme generates (2′-5′)-linked oligonucleotides, which in turn activate a latent cellular endoribonuclease that cleaves viral mRNA. Other hybrid IFNs had an antiproliferative activity against various human cancers that was greater than that of either of the parental molecules. Subsequently, another group of scientists generated other hybrid IFN molecules by a variation of the above-mentioned procedure. In this case, the entire IFN-α cDNA family was PCR amplified and then partially digested with the enzyme DNase into small DNA fragments (~50 to 60 nucleotides long) before the fragments were shuffled and amplified by PCR (Fig. 5.1B). This procedure generated new IFN cDNAs because the PCR mixture contains many overlapping single-stranded DNAs that can act as PCR primers (see chapter 2). Following testing of the many shuffled IFN cDNAs, it is possible to select hybrid IFNs with vastly improved antiviral or antiproliferative activities. In fact, some hybrid IFNs have undergone successful clinical trials (Box 5.1) and have been approved for use as human therapeutic

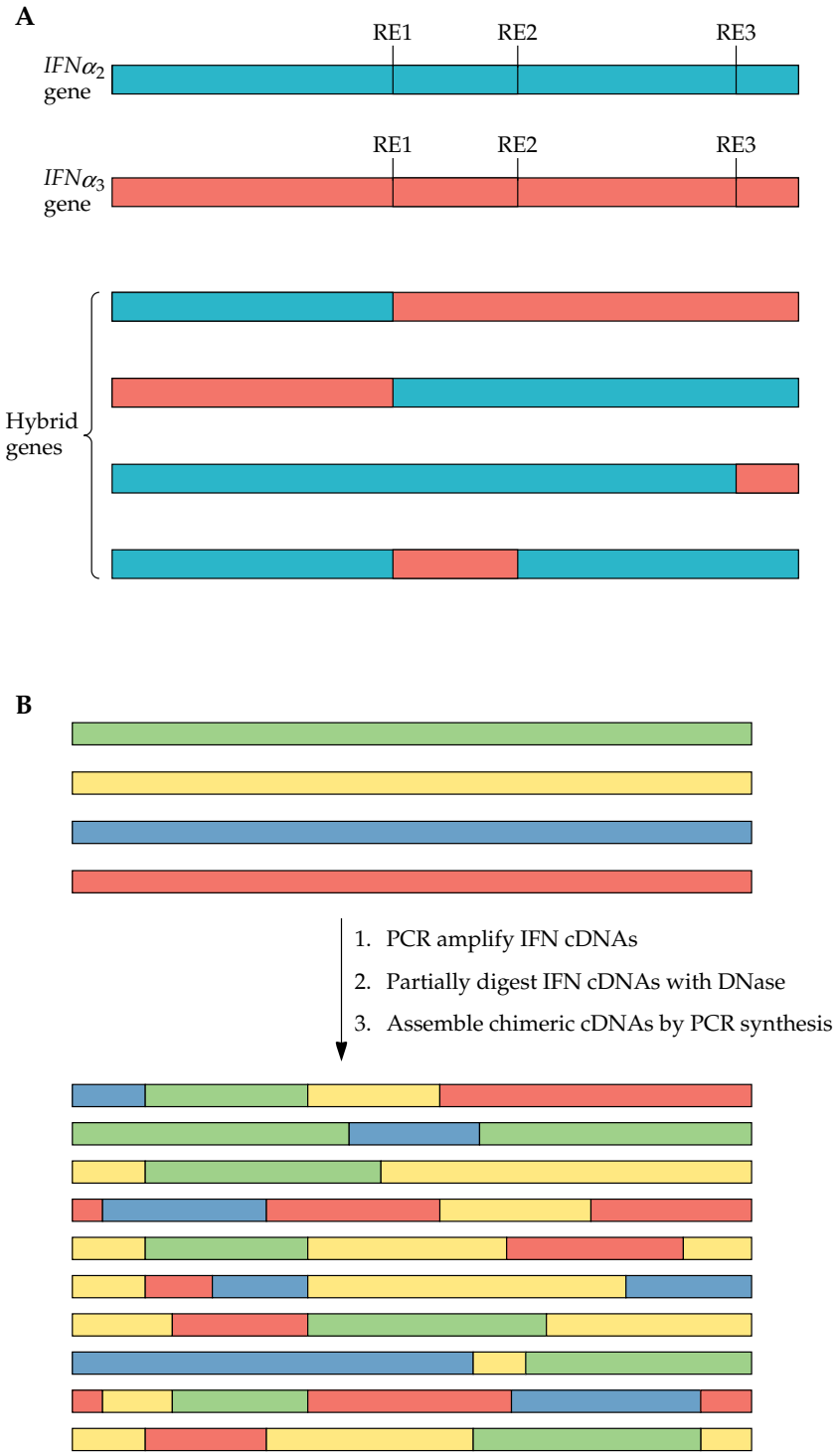


Figure 5.1 Two strategies for constructing hybrid IFN genes. **(A)** Structures of the IFN-α2 and IFN-α3 genes and four hybrid genes. Comparison of the sequences of the IFN-α2 and IFN-α3 genes shows shared restriction enzyme sites (RE1, RE2, and RE3). Digestion of the genes at the indicated restriction sites and ligation of the resulting fragments generate a number of different hybrid IFN genes, of which four possibilities are shown. **(B)** Construction of hybrid IFN-α genes following partial DNase digestion and PCR synthesis of several different IFN-α genes. The resultant IFN-α gene-shuffled libraries are tested for antiproliferative and antiviral activities.

Clinical Trials

box 5.1

After the discovery of a new drug or course of treatment, and before it is made available to the public, it is essential that an extensive analysis of its safety and efficacy be conducted and then reviewed by an impartial agency. Although many countries have developed their own approaches to test new therapeutics, the "gold standard" for clinical trials is the set of requirements established by the FDA. This process is briefly described here.

The preclinical phase of therapeutic drug development (i.e., the initial stage of the process of bringing a new therapeutic agent to market) entails thorough and extensive laboratory research on the mode of action, structure, and other biochemical and physical properties of a potential new drug. Scientists working at universities, research institutes, and drug and biotechnology companies are continually discovering and testing new molecules, as well as new uses for known compounds. However, it is impossible to know with any certainty which avenues of research will eventually bear fruit. Once a promising result has been obtained in the laboratory, and it has been shown to be reproducible, sufficient quantities of a highly purified version of the potential therapeutic compound must be produced so that it can be tested on small animals, such as mice. If the animal tests are positive and there is no evidence of any serious side effects, the organization seeking to commercialize the research files an "investigational new drug" application with the FDA. This is an application to begin the process of clinical trials. Based on the preclinical research data that are provided, about 85% of these applications are approved.

Clinical trials are conducted in three distinct phases (described below), generally requiring a total of about 7 to 9 years at a cost of approximately $100 million to $400 million to complete. At each stage, various compounds are dropped from consideration based on the results obtained. Eventually, approximately 20% of the compounds that looked promising based on preclinical results will, after a careful review of all the data, finally be approved. This slow and expensive process is claimed to be "the most effective method ever devised to assess the efficacy of a treatment" (Zivin, 2000). The three phases of the FDA review process are as follows.

Phase I: With between 10 and 100 healthy people, the safety of the drug and, starting with very low doses, the highest dosages that can be administered are assessed. When there is a chance that serious side effects may result, individuals affected with the disorder that the drug is designed to alleviate may be used.

Phase II: With 50 to 500 affected patients, the optimal dosing regimen is determined. A control group is used so that it is possible to clearly distinguish between the effects of the drug and the natural remission of the disease. The use of a control group also helps to delineate real from apparent side effects of the treatment.

Phase III: Depending upon the disease, approximately 300 to 30,000 patients who have the disease are tested. After it is established that the drug is not harmful (phase I) and the optimal dosing regimen has been determined (phase II), the effectiveness of the treatment needs to be proven.

The requirement for careful and thorough clinical trials ensures both the safety and efficacy of approved drugs. However, since the costs of both the preclinical research and the clinical trials are borne by pharmaceutical companies, this system makes it difficult for small companies that discover a new product to eventually bring that product to market without the involvement of a large corporation with significant financial resources. Furthermore, the high cost of clinical trials and the low probability of a new drug being approved mean that it is unlikely that therapeutic agents will even be considered for clinical trials unless there is a strong possibility that there will be significant financial gains from the sale of that agent. Representatives from the pharmaceutical industry estimate that in 2018 it cost more than $2.6 billion to develop a new FDA-approved drug, including both laboratory research and clinical trials. However, a very large fraction of this estimated cost of bringing drugs to market is the so-called "lost opportunity cost." That is, many pharmaceutical companies include in their calculations what their money would have earned if it had been invested in the stock market instead of in new drug research and development. Nevertheless, regardless of how one calculates the actual cost of bringing a new pharmaceutical to market, it is quite expensive and clearly beyond the means of most small- to medium-size companies. This financial disincentive may discourage research on therapeutic agents for diseases that either affect only a relatively small number of people or affect only populations in poor, underdeveloped countries.

agents. The strategies for creating hybrid IFNs can also be applied to other human gene families whose products have therapeutic potential.

Hepatitis C virus infection is one of the most common causes of liver disease, which affects approximately 200 million people worldwide. Many of these individuals eventually develop either cirrhosis of the liver

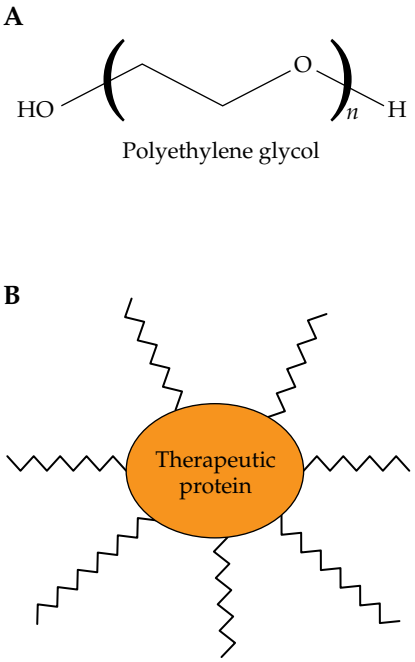


Figure 5.2 (A) Structure of polyethylene glycol. **(B)** Schematic representation of a pegylated therapeutic protein. The polyethylene glycol residues attached to the protein (shown as zigzag lines) shield the protein from proteases and antibodies and increase the protein's effective size and solubility, thereby decreasing the rate at which it is cleared from the kidneys.

or hepatocellular carcinoma (liver cancer). Therapeutic agents that maximize an early antiviral response and maintain viral suppression throughout the course of therapy give infected individuals the best chance of achieving lasting eradication of the virus from an infected individual. One effective treatment for hepatitis C includes the combined use of the antiviral chemical compound ribavirin (a guanosine analog) with IFN-α. However, for this sort of application, longer-acting IFNs are needed to lower the required dosage, minimize the side effects from IFN treatment, and decrease the required frequency of the treatments. One approach to creating long-acting IFNs includes pegylation of the therapeutic agent. The process of pegylation entails covalently attaching the chemical polyethylene glycol (PEG) to proteins (Fig. 5.2). This binding is typically achieved by incubation of a reactive derivative of polyethylene glycol with the target protein molecule. As a consequence of pegylation, the effective size of modified IFNs in solution increases and the protein is protected from protease digestion, thereby prolonging the circulatory time of the therapeutic protein by reducing its renal (kidney) clearance. Some pegylated IFNs, following testing for safety and efficacy, have been approved for clinical use. For example, pegylated IFN-β1 has been approved to treat individuals with multiple sclerosis (MS). This form of IFN allows patients to receive less frequent IFN treatments, i.e., subcutaneous injections once every 2 weeks. Moreover, pegylation of IFN-α increased its half-life 10-fold, allowing patients to be treated once per week rather than three times per week. Pegylated IFN-α-2a and -α-2b have been approved by the U.S. FDA for treatment of human patients.

Human Growth Hormone

Human growth hormone (somatotropin) is a 191-amino-acid pituitary protein that stimulates the production of insulin-like growth factor 1. Insulin-like growth factor 1 is an essential component of the promotion of growth in children and in adults. Infants and children who lack sufficient endogenous levels of human growth hormone, patients with chronic renal insufficiency (defective kidneys), and individuals with Turner syndrome (an inherited disease resulting from a chromosomal abnormality) respond to treatment with growth hormone, which stimulates tissue and bone growth, increases protein synthesis and mineral retention, and decreases body fat storage.

Human growth hormone was one of the first therapeutic proteins in the world to be approved for human use. Prior to its production by recombinant DNA technology, human growth hormone used to treat various deficiencies was extracted from the pituitary glands of human cadavers; growth hormone from other primates was found to be inactive in humans. In1985, cases of Creutzfeldt-Jakob disease were found in some individuals that had previously received cadaver-derived human growth hormone. Based on the assumption that infectious prions causing the disease were transferred along with the cadaver-derived growth hormone, this product was removed from the market. The first recombinant growth hormone was called somatrem (Protropin); it was produced and marketed by Genentech beginning in 1985. It had an amino acid sequence that was identical to that of human growth hormone, except that there was an extra methionine residue at the N-terminal end of the polypeptide chain (which

was thought to increase its half-life). It was discontinued in the late 1990s. At the present time, six different major pharmaceutical companies manufacture and sell recombinant human growth hormone. The recombinant form of the protein is produced in *E. coli* and is identical to native pituitary-derived human growth hormone.

Treatment of children with human growth hormone deficiency with the recombinant protein typically entails daily subcutaneous injections during the years when the child is growing. The cost of the treatment varies depending on the country and the size of the child but is generally approximately $40,000 to $50,000 U.S. per year. In addition to children with growth hormone deficiency, in 2004, the U.S. FDA approved the use of recombinant human growth hormone for individuals whose short stature was caused by a variety of medical conditions other than human growth hormone deficiency.

Since the current treatment regimen is both inconvenient and expensive, it would be advantageous to have a long-lasting form of human growth hormone. To this end, DNA encoding the extracellular domain of the human growth hormone receptor was fused to human growth hormone cDNA with a linker DNA sequence between them encoding a 20-amino-acid-long linker peptide consisting of four repeats of the amino acids Gly_4Ser (Fig. 5.3). This construct has a very strong tendency to dimerize as the growth hormone moiety from one molecule binds with the receptor portion of another molecule. When this growth hormone construct was tested in rats, a single injection promoted growth for 5 days (120 hours), compared to the usual requirement in rats for (at least) daily injections (Fig. 5.4). The data that were obtained from subsequent experiments with male *Macaca fascicularis*, crab-eating macaques (a monkey species), is very similar to the data obtained with rats. However, the rate of clearance was even more delayed in monkeys than in rats, with the modified growth hormone levels remaining elevated for 7 to 10 days, compared to native growth hormone, where the level dropped to an undetectable level after approximately 6 hours. It is thought that the dimerization of the growth hormone construct stabilizes human growth hormone *in vivo* so that it is

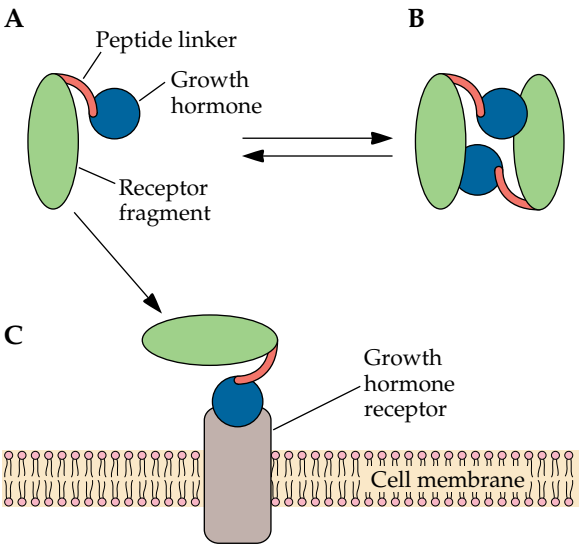


Figure 5.3 Derivatization of growth hormone by coupling it to a portion of the growth hormone receptor using a specific 20-amino-acid peptide. **(A)** Monomeric derivative; **(B)** dimeric derivative; **(C)** monomeric derivative bound to a growth hormone receptor.

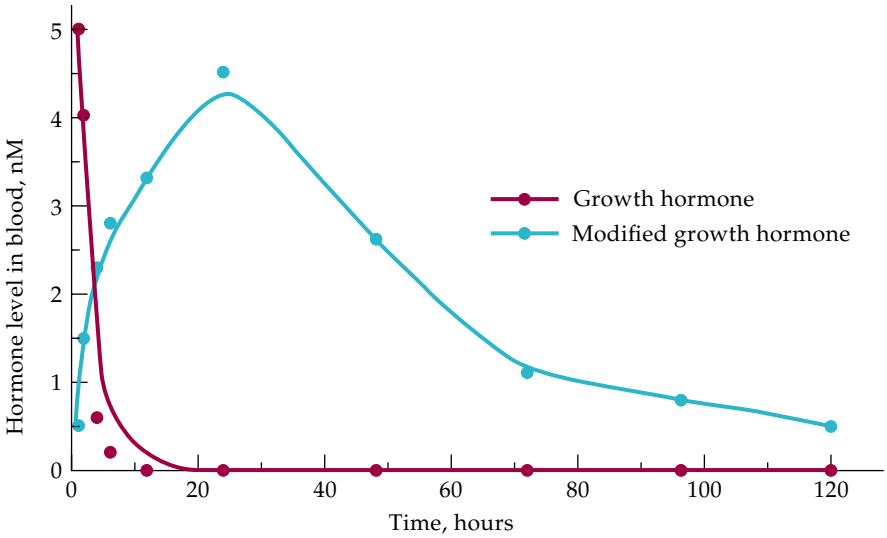


Figure 5.4 Growth hormone and modified growth hormone levels (in nM) in rat blood following subcutaneous injection.

cleared from plasma approximately 300 times more slowly than free monomeric human growth hormone. Under these conditions, the active monomeric form (Fig. 5.3A) is slowly released from the inactive dimeric growth hormone (Fig. 5.3B), allowing it to bind to the cellular growth hormone receptor (Fig. 5.3C). More recently, when a fusion protein that did not include the peptide linker was constructed and tested, this protein yielded growth hormone levels that were 1.7 times higher than when the linker was a part of the construct. Therefore, all subsequent studies, including human trials, have involved the fusion protein without a peptide linker. Since the fusion protein is a single polypeptide that does not require any additional manipulation, the costs of its preparation compared to pegylated growth hormone (which is currently in use in China for the treatment of children with growth hormone deficiency) are significantly reduced. This unique modified slow-release fusion growth hormone protein showed promising results in phase II clinical trials. In addition, in phase III clinical trials, using twice monthly injections of the fusion protein was as effective as daily injections of recombinant growth hormone.

In an attempt to boost muscle mass and athletic performance, human growth hormone has been abused by competitors in various sports since its widespread availability in the 1970s, and its use has been banned by the International Olympic Committee. However, because urine analysis could not detect the presence of synthetic (i.e., recombinant) human growth hormone, the ban was unenforceable until the early 2000s, when blood tests that could distinguish between natural and artificial human growth hormone (by small differences in the posttranslational modification of the protein which do not affect its activity) were developed. In February 2010, a British rugby player had the dubious distinction of being the first athlete to be caught illegally using recombinant human growth hormone. Subsequently, in their agreement with Major League Baseball before the 2012 season, the Baseball Player's Union agreed to the random testing of professional baseball players for the presence of excessive amounts of human growth hormone in their systems.

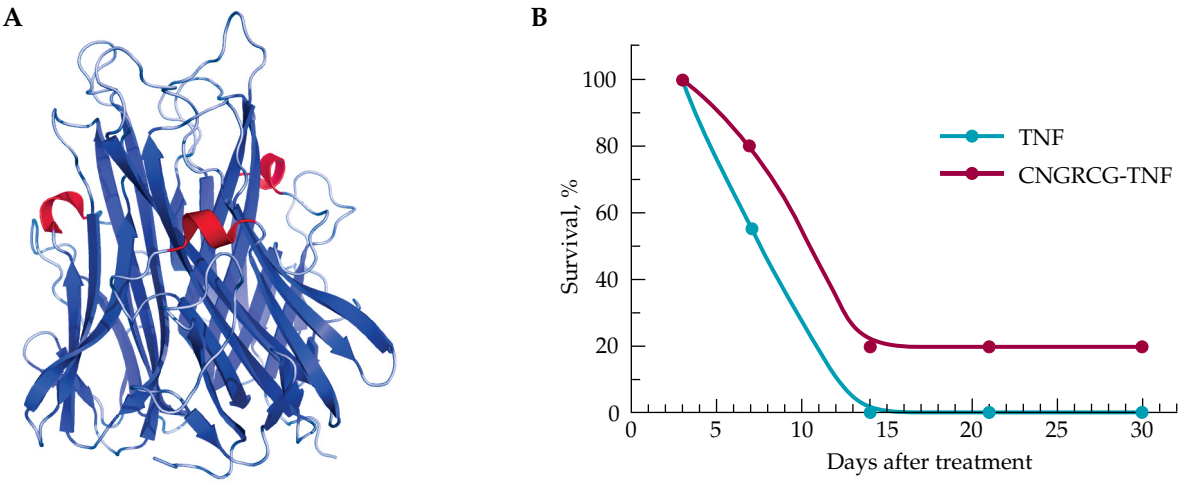
Tumor Necrosis Factor Alpha

Tumor necrosis factor (TNF) is a cytokine (a cell signaling molecule) whose primary role is regulation of the immune system. TNF is produced as a 212-amino-acid-long transmembrane protein arranged in stable homotrimers that are subsequently cleaved by the metalloprotease TNF-converting enzyme to form a soluble homotrimeric cytokine TNF (Fig. 5.5A).

While a number of studies have clearly shown that tumor necrosis factor alpha (TNF-α) is a potent antitumor agent, it has not been widely used as an antitumor treatment in this capacity because of its severe toxicity. If TNF-α could be delivered directly to its intended site of action, i.e., the tumor, then lower doses could be used, and the unwanted side effects would be reduced. To develop a version of TNF-α with tumor specificity, DNA encoding the peptide Cys-Asn-Gly-Arg-Cys-Gly (i.e., cysteine-asparagine-glycine-arginine-cysteine-glycine), which targets a particular tumor cell surface protein, was fused to TNF-α DNA. When the fusion protein was expressed, it contained a 6-amino-acid extension at its N-terminal end. In mice, the cytotoxic activities of Cys-Asn-Gly-Arg-Cys-Gly–TNF-α and TNF-α were identical, indicating that the additional amino acids did not prevent protein folding, trimer formation, or binding to receptors. However, the modified version of TNF-α was 12 to 15 times more effective at inhibiting tumor growth than the unmodified form, and a higher percentage of mice with lymphoma survived after treatment with the modified factor (Fig. 5.5B). In addition, all the mice that were treated with the modified factor and survived for 30 days also survived a second and third challenge with mouse lymphoma (tumor-causing) cells. These data indicate that there is a significant benefit, at least in mice, to fusing TNF-α with a short targeting peptide.

In addition to the above-mentioned hexapeptide-fused TNF-α, scientists have tested the efficacy of pegylated TNF-α. In preclinical mouse trials, pegylated TNF-α had been found to have a decreased level of toxicity and increased efficacy in comparison to the native TNF-α.

Figure 5.5 **(A)** Three-dimensional representation of TNF; **(B)** survival of lymphoma-bearing mice following treatment with 3 μg of either TNF or Cys-Asn-Gly-Arg-Cys-Gly–TNF (CNGRCG-TNF) as a function of the number of days after treatment.



Extending Protein Half-Life

As discussed above for IFN, a protein therapeutic agent with an extended half-life is longer-acting, requires lower doses, and generally has fewer side effects, and the treatments are not required as frequently. The most common strategies for extending the *in vivo* half-life of therapeutic proteins are pegylation and the fusion of the therapeutic protein with human serum albumin. However, each of these approaches has its limitations. For example, while pegylation may both reduce the immunogenicity and increase the stability of therapeutic proteins, it may also significantly increase the costs of the therapeutic protein. In some cases, pegylated proteins or their metabolites may accumulate in the kidney, thereby interfering with normal kidney functioning, so pegylation is not necessarily the best long-term strategy for stabilizing therapeutic proteins. Similarly, fusion proteins that include human serum albumin do not always behave as expected. Thus, scientists have sought to determine whether an unstructured synthetic protein that is covalently attached to the target therapeutic protein might provide some of the benefits of pegylation and fusion to human serum albumin without any of the drawbacks. In designing such a protein, it was decided to exclude all hydrophobic amino acids (which often lead to compact structures or protein aggregation). Amide-containing amino acids that may become unstable upon long-term storage, positively charged side chains that may cause binding to membranes, and cysteine residues that become either oxidized or cross-linked were also excluded. Ultimately, only six amino acids were used in the design of the unstructured protein, i.e., alanine, glutamic acid, glycine, proline, serine, and threonine. A DNA library encoding randomized 36-amino-acid peptides was constructed and screened for expression level. Highly expressed sequences were ligated to form multimers of those sequences and then rescreened for maximal activity. The five most highly expressed proteins were tested for (i) stability of the genetic construct, (ii) solubility, (iii) heat stability, (iv) aggregation resistance, and (v) the presence of contaminating host cell proteins and lipopolysaccharides, eventually resulting in the selection of a single 864-amino-acid protein called XTEN. XTEN fusion proteins are generally highly soluble, properly folded in *E. coli*, nonimmunogenic, and biodegradable. The DNA encoding the XTEN protein was fused at a genetic level to the 3′ end of a target therapeutic protein, while a cellulose-binding domain was fused to the therapeutic protein gene at the 5′ end (Fig. 5.6). Following the translation of the fusion construct into protein in *E. coli*, the cell lysate was purified by passage over a cellulose column, with the fusion protein binding strongly to the column. The cellulose-binding domain was subsequently removed from the fusion protein by digestion with a specific protease (the amino acid recognition site for that protease had previously been engineered into the region between the cellulose-binding domain and the therapeutic protein). The proteins (and peptides) exenatide (similar to glucagon-like peptide 1), glucagon, factor VII, and human growth hormone have all been fused to XTEN and purified as described above. In mice, rats, and monkeys, the plasma half-life of the fusion protein was at least 3 times, and as much as 125 times, greater than the unfused protein. It is argued that this approach is less expensive and more effective than pegylation or fusion to known proteins such as albumin. However, human testing has just begun. In this regard, phase III clinical trials of XTENylated human growth hormone have been initiated. Nevertheless, there are still significant

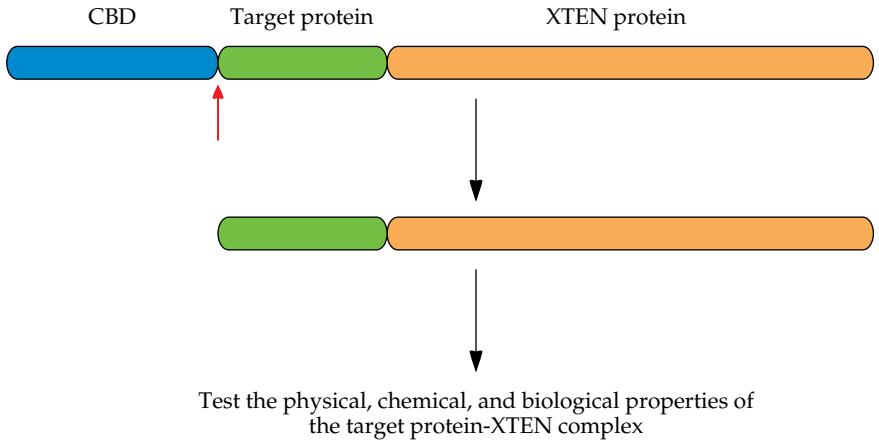


Figure 5.6 Schematic representation of a tripartite fusion protein including a cellulose-binding domain (CBD), a target therapeutic protein, and the 864-amino-acid unstructured XTEN protein. The red arrow indicates the site where, following digestion with a specific protease, the CBD is removed.

technical hurdles that need to be overcome before this technology can be used on a large scale.

Enzymes

DNase I

Cystic fibrosis is one of the most common fatal hereditary diseases among Europeans and their descendants, with approximately 30,000 diagnosed cases in the United States and another 23,000 cases in Canada and Europe. It is estimated that a mutant cystic fibrosis gene is carried by 1 in 29 Europeans, 1 in 65 African Americans, and 1 in 150 Asians. Individuals with cystic fibrosis are highly susceptible to bacterial infections in their lungs. Antibiotic treatment of patients who have these recurring infections eventually leads to the selection of antibiotic-resistant bacteria. The presence of bacteria, some alive and some lysed, contributes to the accumulation of a thick mucus in the lungs of these patients, making breathing very difficult and acting as a source for further infection. The thick mucus in the lungs is the result of the combination of the alginate that is secreted by the living bacteria, the DNA that is released from lysed bacterial cells, and degenerating leukocytes that accumulate in response to the infection, as well as filamentous actin derived from the cytoskeletons of damaged epithelial cells (Fig. 5.7). To address this problem, scientists at the U.S. biotechnology company Genentech isolated the cDNA for the human enzyme deoxyribonuclease I (DNase I) and subsequently expressed this cDNA in Chinese hamster ovary (CHO) cells in culture. DNase I can hydrolyze long polymeric DNA chains into much shorter DNA fragments. The purified enzyme is delivered in an aerosol mist to the lungs of patients with cystic fibrosis. The DNase I decreases the viscosity and adhesivity of the mucus in the lungs and makes it easier for these patients to breathe. While this treatment, which is commercially available in many countries around the world, is not a cure for cystic fibrosis, it nevertheless relieves the most severe symptom of the disease in most patients.

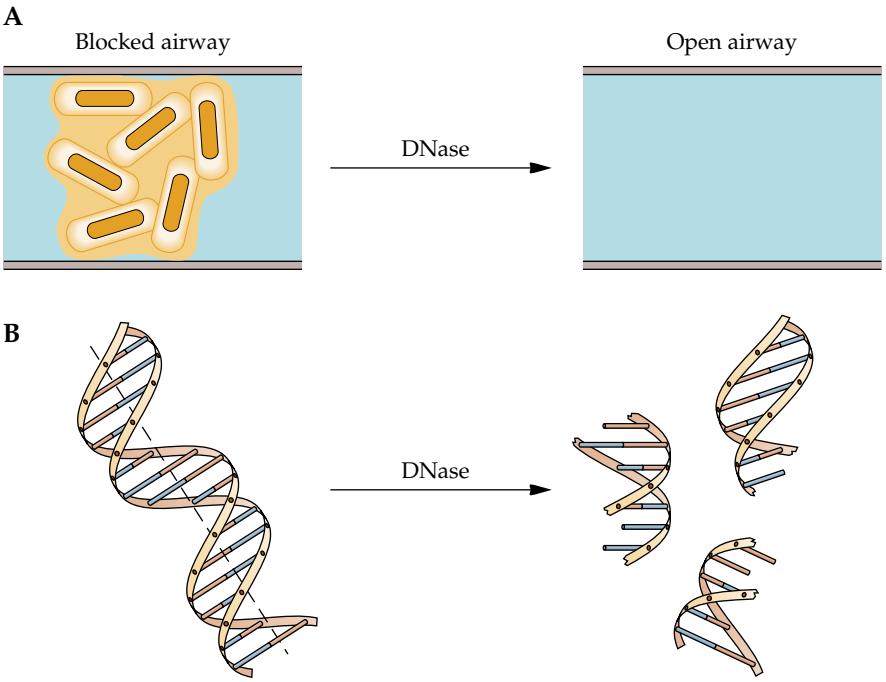
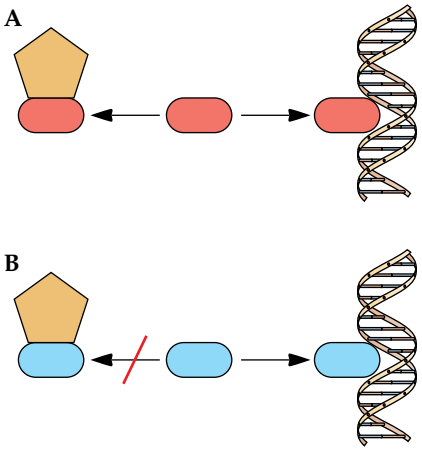


Figure 5.7 **(A)** Schematic representation of a portion of a human lung occluded by a combination of live alginate-secreting bacterial cells, lysed bacterial cells, and leukocytes and their released DNA being cleared by digestion by aerosol-delivered DNase I. **(B)** Digestion of DNA by DNase I.

The monomeric form of actin, which is abundant in most cells, binds very tightly to DNase I (inhibitor constant [*Ki*] = ~1 nM), which significantly inhibits its ability to cleave DNA (Fig. 5.8). This interaction can limit the effectiveness of DNase I as a therapeutic agent. On the basis of X-ray crystallographic studies, it was possible for scientists at Genentech to predict which amino acid residues of DNase I interacted with actin and were therefore possible targets for change by directed mutagenesis (see chapter 2). For example, changing amino acid 144 from alanine to arginine or amino acid 65 from tyrosine to arginine decreased the binding of DNase I to actin up to 10,000-fold. The actin-resistant proteins had 10- to 50-fold more DNase I activity than the native enzyme. In addition to its use with cystic fibrosis patients, the modified DNase I may be used to treat chronic bronchitis, pneumonia, bronchiectasis, emphysema, asthma, or systemic lupus erythematosus. In 1995, scientists working for Genentech were awarded a patent describing the development and use of this modified form of DNase I; this patent has now expired. The enzyme was approved for use by the FDA in 1994; it had worldwide sales of approximately $750 million in 2018.

Figure 5.8 **(A)** Competition between DNA and actin (shown in orange) for the binding of DNase I, shown in red. **(B)** Modified DNase I (blue) continues to bind tightly to DNA but has a significantly decreased affinity for actin.



Alginate Lyase

Alginate is a polysaccharide polymer that is produced by a wide range of seaweeds and both soil and marine bacteria. Alginate is composed of polymers of the sugars β-D-mannuronate and α-L-guluronate (Fig. 5.9). The properties of a particular alginate depend on the relative amounts

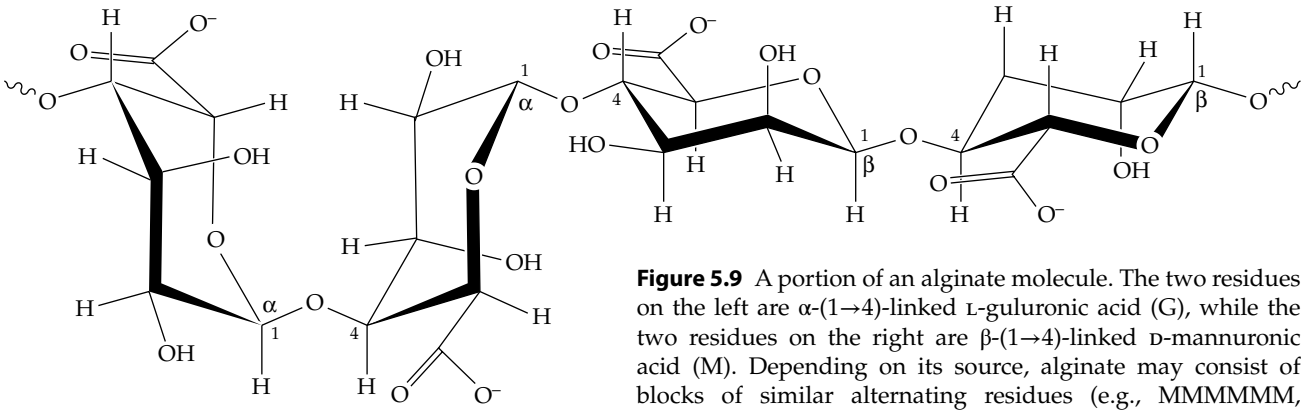


Figure 5.9 A portion of an alginate molecule. The two residues on the left are α-(1→4)-linked L-guluronic acid (G), while the two residues on the right are β-(1→4)-linked D-mannuronic acid (M). Depending on its source, alginate may consist of blocks of similar alternating residues (e.g., MMMMMM, GGGGGG, and GMGMGMGM), all of which have different conformations and behaviors.

and distribution of these two saccharides. For example, stretches of α-L-guluronate residues form both interchain and intrachain cross-links by binding calcium ions, and the β-D-mannuronate residues bind other metal ions. The cross-linked alginate polymer forms an elastic gel. In general, the viscosity of an alginate polymer is related to its structure and molecular size.

The excretion of alginate by mucoid strains of the bacterium *Pseudomonas aeruginosa* that infect the lungs of patients with cystic fibrosis significantly contributes to the viscosity of the mucus in the airways. Once mucoid strains of *P. aeruginosa* have become established in the lungs of cystic fibrosis patients, it is almost impossible to eliminate them by antibiotic treatment. This is because the bacteria form biofilms, surface-associated microcolonies surrounded by an exopolysaccharide matrix (Fig. 5.10A) in which the alginate prevents the antibiotics from coming into contact with the bacterial cells. In one experiment, it was shown that the addition of the enzyme alginate lyase, which can liquefy bacterial alginate, together

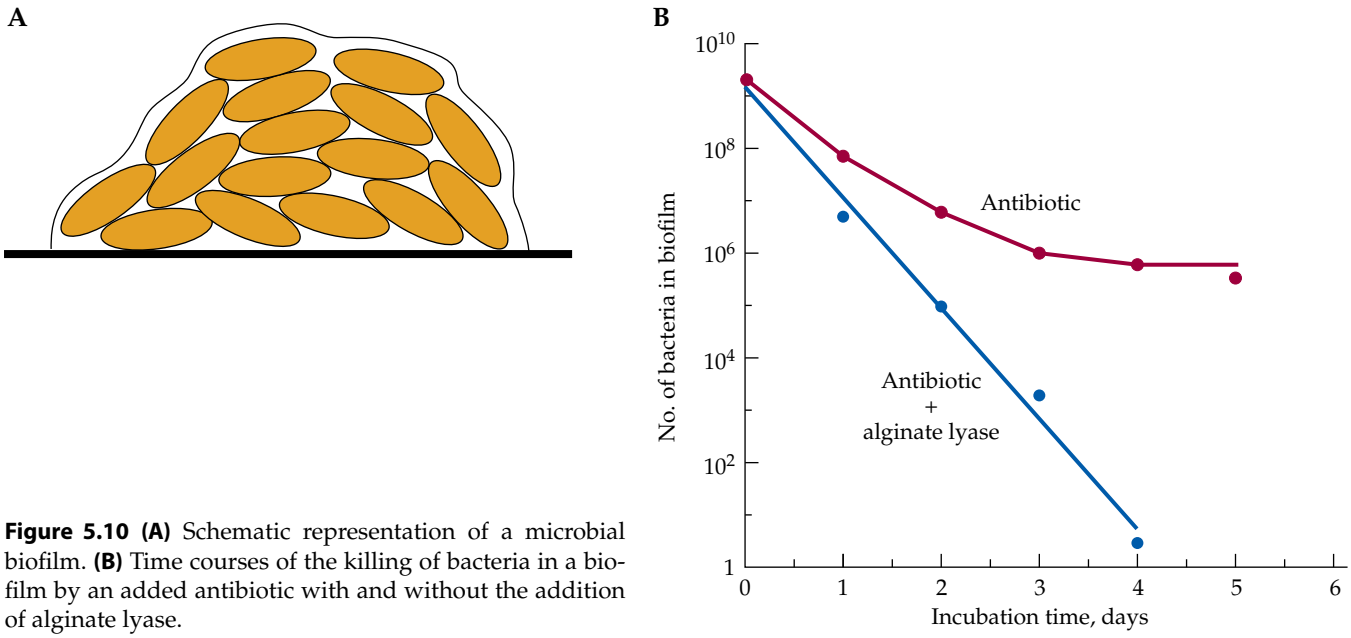


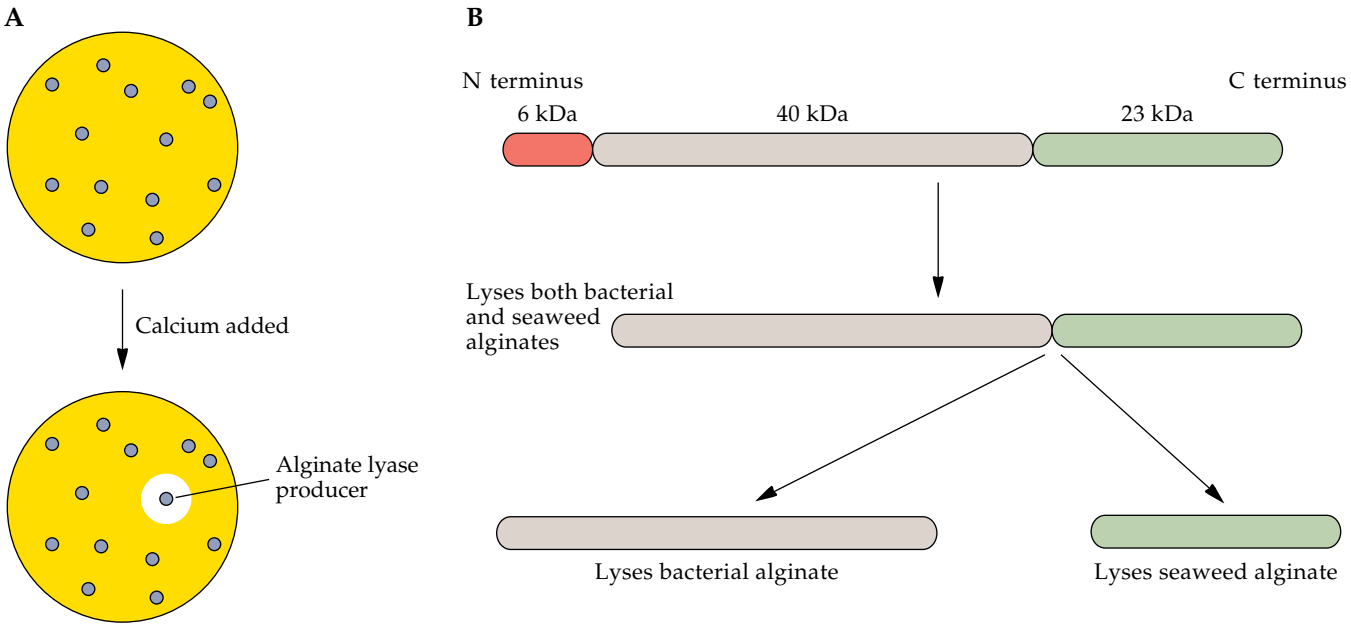
Figure 5.10 (A) Schematic representation of a microbial biofilm. **(B)** Time courses of the killing of bacteria in a biofilm by an added antibiotic with and without the addition of alginate lyase.

with or prior to antibiotic treatment, significantly decreased the number of bacteria found in biofilms (Fig. 5.10B). Thus, alginate lyase treatment not only decreases the viscosity of the mucus but also facilitates the ability of added antibiotics to kill the infecting bacterial cells. This result suggests that in addition to the DNase I treatment described above, depolymerization of the alginate might help clear blocked airways of individuals with cystic fibrosis.

An alginate lyase gene has been isolated from a *Flavobacterium* species, a Gram-negative soil bacterium that is a strong producer of this enzyme. A *Flavobacterium* genomic DNA library was constructed in *E. coli* and screened for alginate lyase-producing clones by plating the entire library onto solid medium containing alginate. Following growth, colonies that produced alginate lyase formed a halo around the colony when a solution of calcium was added to the plate (Fig. 5.11A). In the presence of calcium, all of the alginate in the medium, except in the immediate vicinity of an alginate lyase-positive bacterial clone, becomes cross-linked and opaque. Since hydrolyzed alginate chains do not form cross-links, the medium surrounding an alginate lyase-positive clone is transparent.

Analysis of a cloned DNA fragment from one of the positive colonies revealed an open reading frame encoding a polypeptide with a molecular

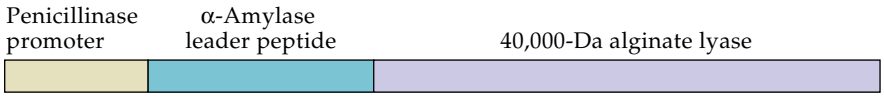
Figure 5.11 (A) Schematic representation of the detection of an alginate lyase-producing clone from a clone bank of a *Flavobacterium* sp. in *E. coli*. The alginate that is present in the growth medium is digested by alginate lyase secreted by an *E. coli* clone. The alginate in the vicinity of such a colony is not cross-linked when a calcium solution is added to the solid medium and instead produces a clear zone (halo) surrounding the colony. **(B)** Processing of the recombinant *Flavobacterium* alginate lyase protein precursor in *E. coli*. A 6-kDa peptide is removed from the N terminus of the 69-kDa precursor to yield a 63-kDa protein that can depolymerize alginate from both seaweed and bacteria. A second proteolytic cleavage event converts the 63-kDa protein into a 23-kDa protein that is active against seaweed alginate and a 40-kDa protein that hydrolyzes bacterial alginate.



mass of approximately 69,000 Da. Detailed biochemical and genetic studies indicated that this polypeptide is a precursor of the three different alginate lyases produced by the *Flavobacterium* sp. (Fig. 5.11B). After the 69,000-Da precursor is produced, a proteolytic enzyme cleaves off an N-terminal peptide of about 6,000 Da. The resultant 63,000-Da protein can lyse both bacterial and seaweed alginates. Cleavage of the 63,000-Da protein yields a 23,000-Da enzyme that depolymerizes seaweed alginate and a 40,000-Da enzyme that is effective against bacterial alginate. To produce large amounts of the bacterial alginate-specific enzyme, the DNA corresponding to the 40,000-Da enzyme was amplified by PCR and then inserted into a *Bacillus subtilis* plasmid vector fused to the coding sequence for a *B. subtilis* α-amylase leader peptide to direct the secretion of the protein and placed under the transcriptional control of a *Bacillus* penicillinase gene promoter (Fig. 5.12). Transformation of *B. subtilis* cells with this construct yielded colonies that produced large halos on solid medium containing alginate after calcium was added. When these transformants were grown in liquid medium, the recombinant alginate lyase was secreted into the culture broth. Further tests showed that the enzyme efficiently liquefied alginates that were produced by mucoid strains of *P. aeruginosa* that had been isolated from the lungs of patients with cystic fibrosis.

Additional studies are necessary to determine whether a recombinant alginate lyase derived from a nonhuman source is an effective therapeutic agent. Unfortunately, since most sources of alginate lyase are nonhuman, these proteins are not necessarily a good choice for use as a therapeutic agent, as they are likely to elicit an immune response in the patient, thereby causing a range of complications. To significantly decrease the antigenicity of the alginate lyase, the enzyme may be pegylated. However, pegylation of alginate lyase results in a large decrease in the activity of the enzyme. Nevertheless, it is possible to avoid the inactivation of the enzyme while still decreasing its antigenicity by adding polyethylene glycol only to specific amino acids. To do this, several specific mutants of alginate lyase were created by directed mutagenesis in which one amino acid residue at a time was changed to cysteine (Fig. 5.13). Then, the modified alginate lyases were each reacted with a modified form of polyethylene glycol that formed a covalent bond with the (newly introduced and only) cysteine residue. The pegylated enzymes were then tested for alginate lyase activity, antigenicity, and the ability to disrupt mucoid biofilms produced by the bacterium *P. aeruginosa*. One of the modified enzymes (in which an alanine residue at position 53 was changed to a cysteine residue) displayed behavior that was superior to those of all of the other modified enzymes and the native enzyme. Specifically, the selected modified enzyme degraded bacterial alginate 80% more rapidly than the native enzyme, disrupted 94% of an established *P. aeruginosa* biofilm (compared to 75% disruption of the

Figure 5.12 DNA construct encoding the 40,000-Da alginate lyase protein. After translation, the leader peptide from a *B. subtilis* α-amylase gene is fused to the N terminus of the alginate lyase coding sequence. The DNA construct is under the transcriptional control of a *B. subtilis* penicillinase gene expression system.



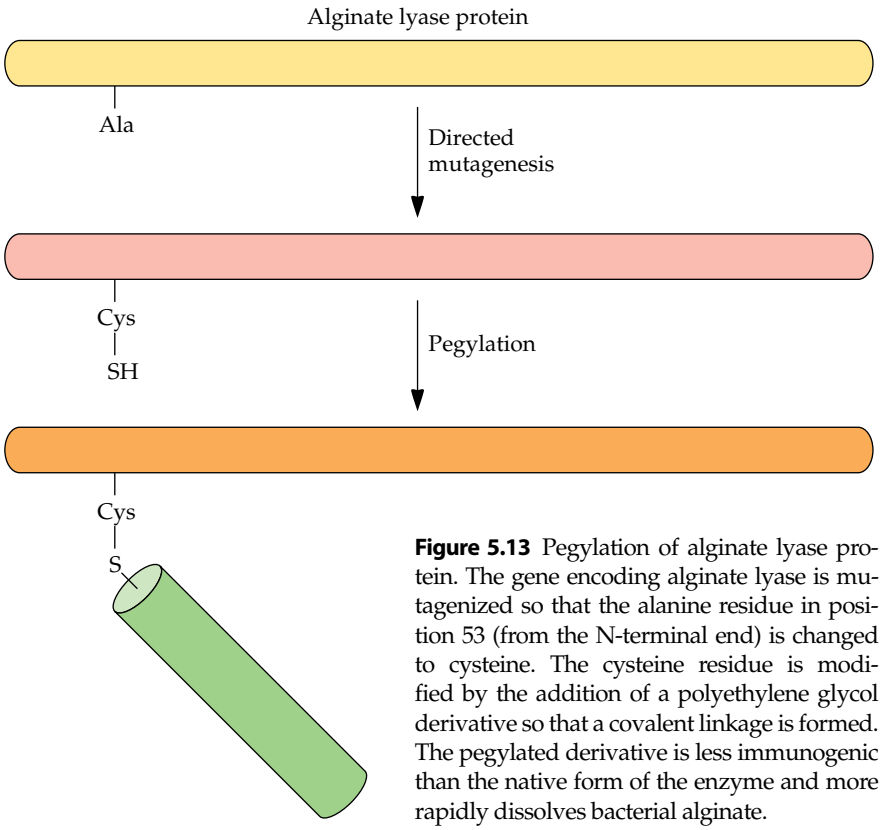


Figure 5.13 Pegylation of alginate lyase protein. The gene encoding alginate lyase is mutagenized so that the alanine residue in position 53 (from the N-terminal end) is changed to cysteine. The cysteine residue is modified by the addition of a polyethylene glycol derivative so that a covalent linkage is formed. The pegylated derivative is less immunogenic than the native form of the enzyme and more rapidly dissolves bacterial alginate.

biofilm by the native enzyme), and had reduced immunogenicity. This result, although preliminary, is very encouraging. The next step will be to test this modified enzyme in animal models of cystic fibrosis.

Phenylalanine Ammonia Lyase

The human genetic disease phenylketonuria is an autosomal recessive trait; i.e., both parents must pass on the defective gene, resulting in the production of an impaired phenylalanine hydroxylase enzyme. In the United States, approximately 1 of every ~13,000 to 21,000 newborns has phenylketonuria, with different populations having somewhat different frequencies of this genetic disease. Approximately 500 different mutations in the gene encoding phenylalanine hydroxylase can lead to an enzyme with decreased activity, preventing the disposal of excess amounts of phenylalanine. When phenylalanine hydroxylase, which oxidizes phenylalanine to tyrosine, is deficient, the normal cognitive development of an individual is impaired, and mental retardation ensues due to a buildup of phenylalanine (Fig. 5.14A). Following diagnosis of phenylketonuria, either prenatally or shortly after birth, the treatment entails a controlled semisynthetic diet, first developed in the mid-1950s, with low levels of phenylalanine through infancy and possibly for life. People with phenylketonuria need to consume diets that avoid high-protein foods such as milk, cheese, eggs, nuts, soybeans, beans, chicken, beef, pork, fish, peas, and beer, as well as the synthetic sweetener aspartame (a methyl ester of aspartate-phenylalanine dipeptide). A possible treatment for this disease

would be the administration of the enzyme phenylalanine hydroxylase. Unfortunately, phenylalanine hydroxylase is a multienzyme complex that is not very stable and requires a cofactor (tetrahydrobiopterin) for activity. Moreover, depending on the source of the enzyme (i.e., enzyme replacement therapy), phenylalanine hydroxylase may be sufficiently immunogenic to cause severe problems if this enzyme is used to treat people with phenylketonuria. Some scientists are trying to overcome the immunogenicity of administered phenylalanine hydroxylase by developing pegylated versions of this enzyme, and this approach has been successful with some patients.

On the other hand, phenylalanine ammonia lyase, which converts phenylalanine to ammonia and *trans*-cinnamic acid (Fig. 5.14B), is a stable enzyme that does not require a cofactor and could potentially prevent the accumulation of phenylalanine in phenylketonuria patients. The cinnamic acid is largely converted to hippuric acid and then excreted in the urine. To test this concept, the gene for phenylalanine ammonia lyase from the yeast *Rhodosporidium toruloides* was cloned and overexpressed in *E. coli*. Preclinical studies were conducted with mice that were defective in producing phenylalanine ammonia lyase and therefore accumulated phenylalanine. With these mice, plasma phenylalanine levels were lowered when phenylalanine ammonia lyase was injected intravenously or encapsulated enzyme was administered orally. Thus, in short-term experiments in mice, phenylalanine ammonia lyase is an effective substitute for phenylalanine hydroxylase, and the orally delivered enzyme is sufficiently stable to survive the mouse gastrointestinal tract and still function. However, with repeated administration of nonhuman phenylalanine ammonia lyase, this enzyme is immunogenic, and its continued use becomes problematic. Thus, this enzyme was pegylated to significantly reduce its immunogenicity, while largely retaining its therapeutic catabolic activity, and was approved for human therapeutic use in 2017. More recently, researchers identified a double mutant of phenylalanine ammonia lyase derived from the cyanobacterium *Anabaena variabilis* that displayed a lower level of immunogenicity. The gene for this enzyme was expressed in *E. coli*, purified, and then pegylated. The pegylated version of the *A. variabilis* enzyme is currently undergoing clinical trials.

It is envisioned that a combination of oral enzyme therapy with pegylated phenylalanine ammonia lyase and a less stringent low-phenylalanine diet might serve to improve the quality of life of individuals affected with this disease. However, it is important to note that the pegylated enzyme is administered subcutaneously (under the skin) rather than orally. In an attempt to develop a more efficacious means of delivering this therapeutic agent, scientists are currently experimenting with the use of live bacteria, including both specific strains of *E. coli* (Nissle) and *Lactobacillus reuteri*, as delivery vehicles.

α_1-Antitrypsin

A number of different bacterial and viral proteins that are involved in the pathogenicity of the organism are synthesized as inactive precursor proteins. The cleavage and activation of many of these precursor proteins is performed, following human infection, by human proteases when the protease recognizes the amino acid sequence –Arg-X-Lys/Arg-Arg↓–,

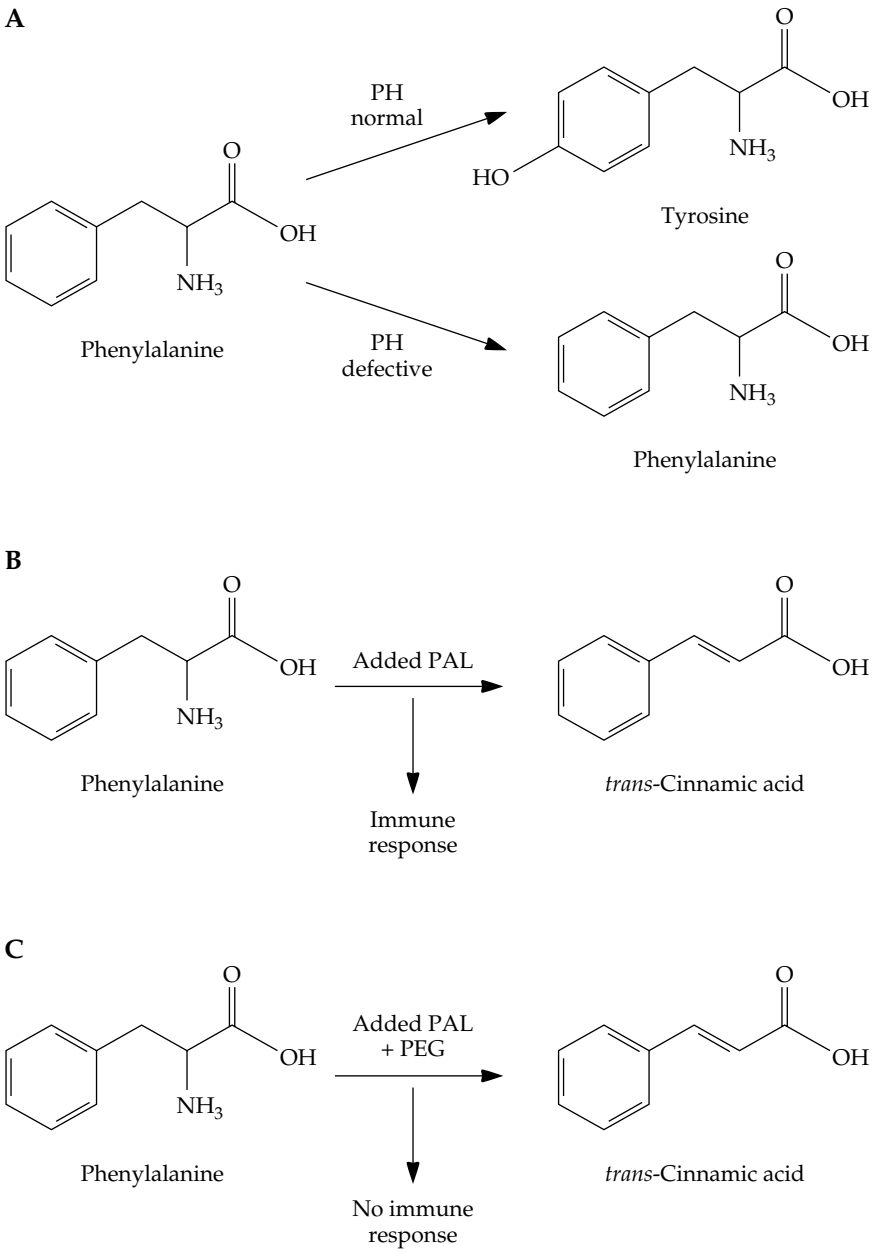


Figure 5.14 Fate of dietary phenylalanine in different individuals. **(A)** Individuals with normal levels of phenylalanine hydroxylase (PH) activity convert phenylalanine to tyrosine, while those with a defective version of this enzyme activity do not. **(B)** Individuals who are given phenylalanine ammonia lyase (PAL) can convert phenylalanine to ammonia and *trans*-cinnamic acid; however, they may mount an immune response directed against PAL. **(C)** Immunogenicity is abolished when the PAL is pegylated.

with peptide bond cleavage occurring on the C-terminal side of the C-terminal Arg (as indicated by the arrow), where X is any of the 20 common amino acids. Since a number of different pathogen proteins contain this amino acid sequence and are therefore activated by this processing step, a therapeutic agent that targets the processing protease and blocks

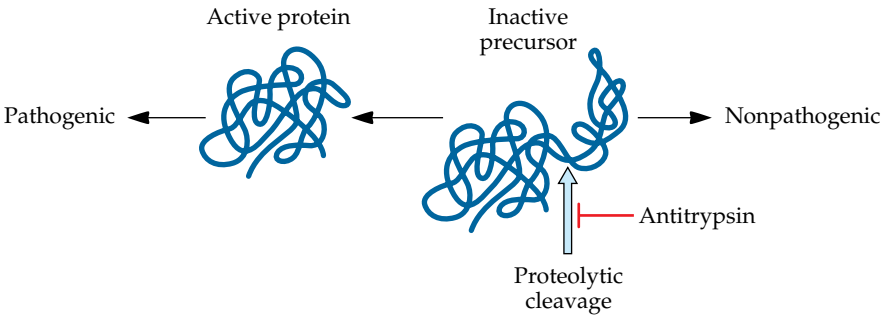


Figure 5.15 Schematic representation of α_1-antitrypsin inhibiting the proteolytic cleavage (activation) of pathogenic precursor proteins by human proteases.

its activity might act as a broad-spectrum antibacterial and antiviral agent (Fig. 5.15). When a variant of human protease inhibitor α_1-antitrypsin (which normally protects tissues from enzymes of inflammatory cells) was genetically engineered to slightly alter its specificity and tested in tissue culture experiments, the protein blocked the processing of human immunodeficiency virus (HIV) type 1 glycoprotein gp160, as well as measles virus protein F_0, and consequently, in both cases, the production of infectious viruses. When the α_1-antitrypsin variant was added to cell cultures, it blocked the production of human cytomegalovirus, a major cause of illness and death in organ transplant recipients and AIDS patients. The α_1-antitrypsin variant is both potent and selective. Against human cytomegalovirus, it is at least 10-fold more effective than any currently used viral inhibitory agent. Its efficacy has been demonstrated in cell culture; however, it remains to be determined if the strategy is effective in animals.

Glycosidases

The ABO blood group system is based upon the presence or absence of specific carbohydrate residues on the surfaces of erythrocytes, endothelial cells, and some epithelial cells. The monosaccharide that determines blood group A is a terminal α-1,3-linked *N*-acetylgalactosamine, while the corresponding monosaccharide of blood group B is α-1,3-linked galactose (Fig. 5.16). Group O cells lack both of these monosaccharides at the ends of their oligosaccharide chains and instead contain α-1,2-linked fucose, which is designated the H antigen. Plasma from blood group A individuals contains antibodies against the B antigen, blood group B individuals have antibodies against the A antigen, and blood group O individuals have antibodies against both the A and B antigens. In practice, this means that individuals with either anti-A or anti-B antibodies cannot safely receive a blood transfusion containing the incompatible antigen, since this is likely to cause a severe immune response (Table 5.2). As a consequence, blood group AB individuals are said to be universal recipients while those from blood group O are universal donors. Thus, when a blood transfusion is required, it is advantageous to have a large supply of plasma that is from blood group O (e.g., in an emergency situation, there may not be sufficient time to check a patient's blood group). On a typical day, approximately 16,500 liters of blood are utilized in the United States. Fortunately, digestion of blood cells from either type A or B with specific glycosidases can

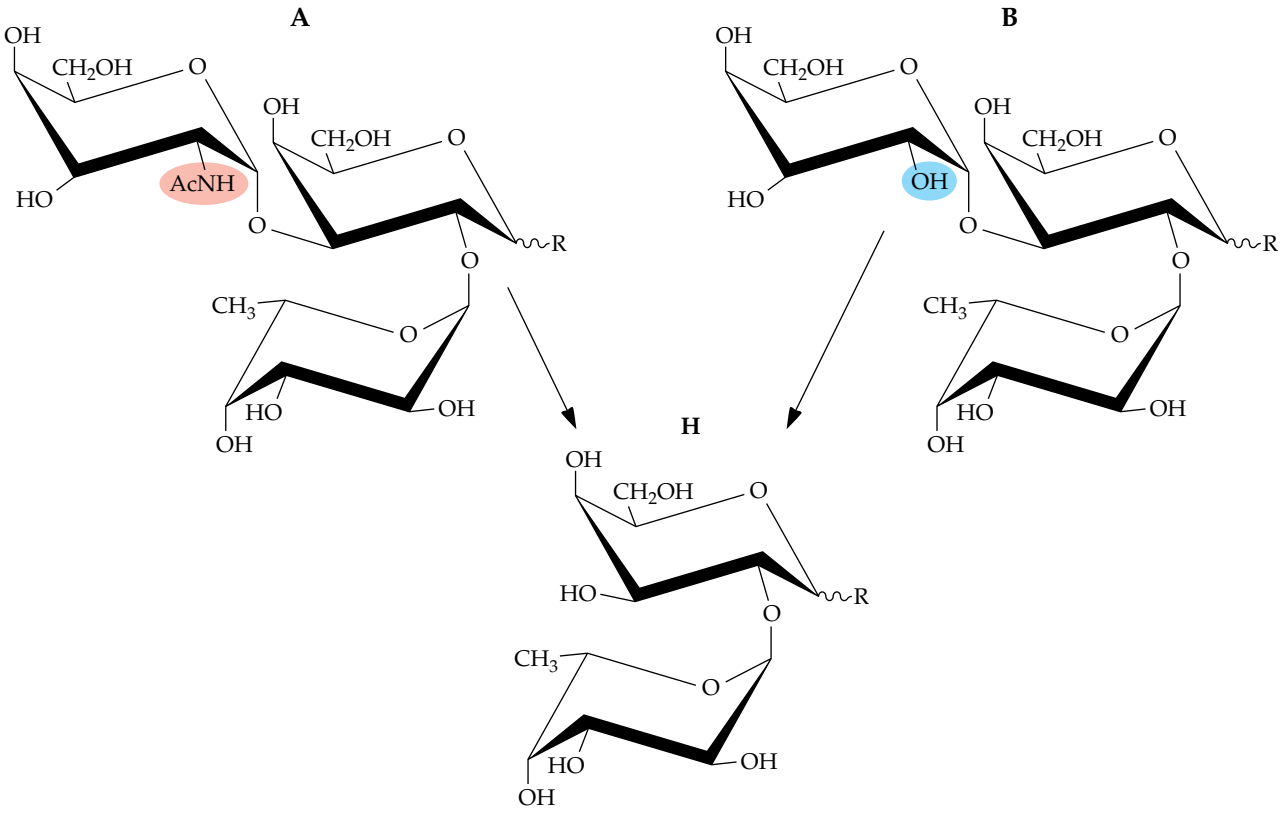


Figure 5.16 Digestion of the monosaccharides that determine blood groups A and B to obtain the H antigen (i.e., blood group O) with specific glucosidases. The major chemical differences between the monosaccharides that determine blood groups A and B are highlighted. AcNH stands for an acetyl moiety covalently bound to a nitrogen atom.

cause types A, B, and AB to be converted into type O (Fig. 5.16). These enzymes were found following an extensive screening process of 2,500 fungal and bacterial isolates. Eventually, an active α-*N*-acetylgalactosamidase, which converts group A blood cells to group O, was found in the Gram-negative bacterium *Elizabethkingia meningoseptica*, and an α-galactosidase A, which converts group B blood cells to group O, was found in *Bacteroides fragilis* (also a Gram-negative bacterium). The genes that encoded these enzymes were isolated, and the proteins were characterized. Both of the

Table 5.2 Compatible and incompatible blood groups

	Recipient blood type			
Donor blood type	A	B	AB	O
A	Compatible	Incompatible	Compatible	Incompatible
B	Incompatible	Compatible	Compatible	Incompatible
AB	Incompatible	Incompatible	Compatible	Incompatible
O	Compatible	Compatible	Compatible	Compatible

Individuals from one blood group may safely receive a blood transfusion from individuals from a compatible blood group but not from someone from an incompatible blood group.

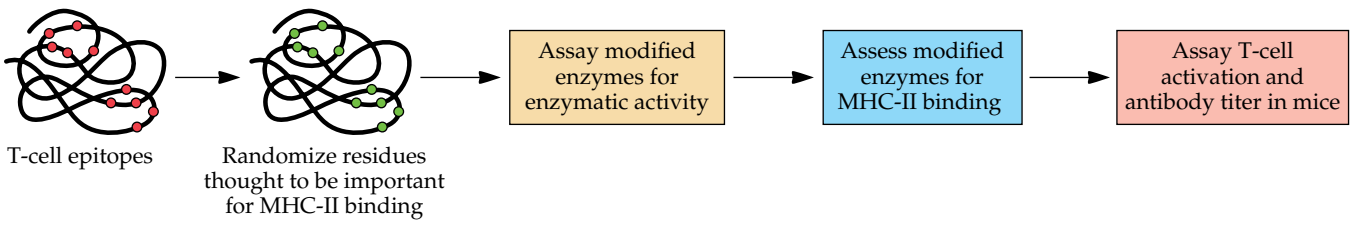
enzymes have a high specificity for cleaving the appropriate monosaccharide under conditions that maintain the integrity and functioning of the treated red blood cells. Moreover, each enzyme could readily be removed from the treated red blood cells following treatment. Unfortunately, this approach has only met with limited success, as the enzymes are not efficient enough. However, more recently, scientists have discovered more efficient enzymes from the human gut (anaerobic Gram-positive) bacterium *Flavonifractor plautii*. These enzymes convert human type A blood (nearly one-third of the human population is type A) to universal type O blood. It is hoped that, following extensive testing, this work can become commercialized within the next few years, thereby alleviating some of the shortage in the blood supply.

Masking Nonhuman Epitopes

For a number of human diseases, an adequate level of a particular enzyme is lacking, such as phenylalanine hydrolase as described above. Some of these diseases can be treated by the administration of human-derived enzymes, typically produced from cloned human cDNAs. However, there are a number of diseases where a needed human enzyme is not known or not readily available. In these instances, enzymes with the required enzymatic activity can often be found in bacteria or fungi. Unfortunately, the therapeutic use of heterologous enzymes can be complicated by the immunogenicity of these enzymes in humans, resulting in the production of antienzyme antibodies that can in turn mediate adverse effects such as hypersensitivity reactions, anaphylactic shock, and the inactivation of the heterologous enzyme. This problem can sometimes be avoided by pegylation of the nonhuman enzyme, as described above. An alternative approach to this problem includes the genetic modification of the gene or cDNA that encodes the nonhuman enzyme so that many or all of the epitopes (antigenic determinants) that might otherwise induce an immune response in humans to this protein become masked.

One group of researchers has devised a scheme to mask the epitopes on a potentially therapeutic enzyme from a nonhuman organism (Fig. 5.17). First, a computer program was used to predict possible (human) T-cell epitopes (i.e., those few amino acids that are sufficient to elicit an immune response) on the therapeutic enzyme. T-cell epitopes are normally bound by the human major histocompatibility complex

Figure 5.17 Schematic representation of modification of a nonhuman enzyme so that it is no longer immunogenic when introduced into patients that lack a sufficient level of the human version of this enzyme. The abbreviation MHC-II stands for the human major histocompatibility complex.



(MHC-II) and elicit T cell-dependent immune responses in humans. The antigenicity of a protein (e.g., a potentially therapeutic enzyme) can be decreased by altering some of the amino acids in these epitopes. Second, the nucleotides in the region of the DNA encoding the therapeutic enzyme that contains the identified T-cell epitopes (typically around nine nucleotides long, encoding three amino acids) may be substituted combinatorially (i.e., with all possible combinations) in an effort to change the amino acids within the identified epitopes without changing the activity of the enzyme. Third, all of the modified enzymes are tested for enzymatic activity, and any modified enzymes that have significantly decreased enzyme activity are discarded. Fourth, the active modified enzymes are tested for the ability to bind to the MHC-II complex, and those that exhibit reduced binding compared to the wild-type enzyme are selected. Finally, the modified therapeutic enzymes with reduced MHC-II binding are assayed for their ability to induce antibodies and a T-cell response in mice. When this seemingly complex procedure was carried out, it yielded candidate therapeutic enzymes in mice, opening up the possibility that many different enzymes that are derived from various nonhuman sources may eventually become human therapeutic enzymes.

Toxin-Intein Fusions

With the widespread dissemination of antibiotics (initially penicillin and streptomycin) following World War II, a large number of infectious bacterial diseases are no longer a significant threat to people. Unfortunately, as a consequence of the overuse of antibiotics, there has been a worldwide emergence of antibiotic-resistant bacteria. Some scientists have suggested that unless new antibacterial agents are developed, in the next 15 to 20 years antibiotic-resistant bacteria could again become a major worldwide health problem. To deal with this impending problem, scientists have sought to develop a range of antibacterial agents, including new antibiotics (see chapter 8), antibacterial vaccines (see chapter 7), and some unique therapeutic agents.

In the past 20 years, the bacterium *Vibrio cholerae* typically has caused 21,000 to 143,000 deaths per year. Part of the reason for this large number of deaths is related to the fact that many of the disease-causing bacterial strains are antibiotic-resistant. One unique therapeutic strategy directed against some human-pathogenic bacteria such as *V. cholerae* has recently been reported (Fig. 5.18). In this case, a nonpathogenic *E. coli* cell engineered with a plasmid that encodes a bacterial toxin is used as a delivery vehicle to transfer (via conjugation) the toxin-synthesizing potential to the target *V. cholerae* cells. Once the genes responsible for toxin production have been transferred to the pathogen, they may be activated and, as a result, kill the pathogen. In laboratory tests, 100% of the antibiotic-resistant *V. cholerae* cells that were present received the plasmids and were killed. When zebrafish larvae or brine shrimp (*Artemia salina*) infected with *V. cholorae* were subsequently infected with the toxin-encoding *E. coli*, the pathogen was completely killed. While this approach admittedly will require a considerable amount of fine-tuning (including working out how to turn on the toxin genes following administration to a live animal) and subsequent testing, this type of

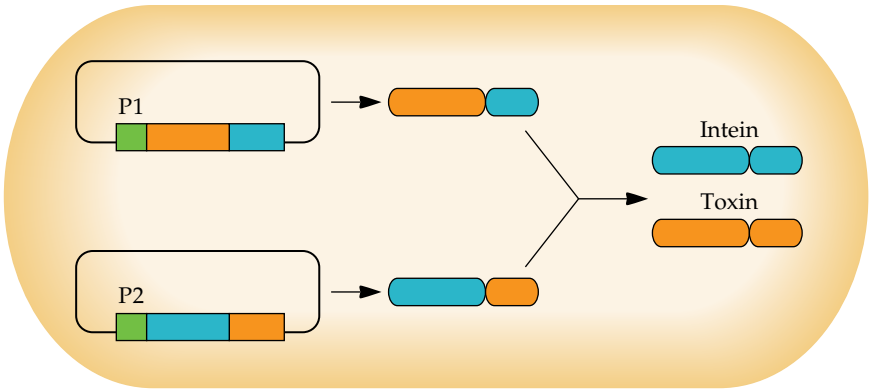


Figure 5.18 Schematic representation of the construction and employment of a toxin-intein fusion protein as a means of killing antibiotic-resistant *V. cholerae* bacteria. An *E. coli* cell is transformed with two different plasmids, one encoding half of the toxin protein fused to the N-terminal portion of the intein with both under the transcriptional control of the P_{BAD} promoter (P1, activated by the addition of arabinose), while the other plasmid encodes the other half of the toxin protein fused to the C-terminal protein of the intein with this construct under the transcriptional control of the P_{LAC} promoter (P2, activated by the addition of lactose or IPTG). After the two promoters are activated, two short-lived toxin-intein fusion proteins are produced. The proteins are recombined to form separate intein and toxin molecules (inteins act as protein introns that can excise and remove themselves from adjacent proteins and join the remaining portions of the nonintein protein). The toxin molecule then kills the cell. To kill a target pathogen, the engineered *E. coli* strain is conjugated with the pathogenic bacterium (e.g., *V. cholerae*) prior to inducing the expression of the genes on the two plasmids.

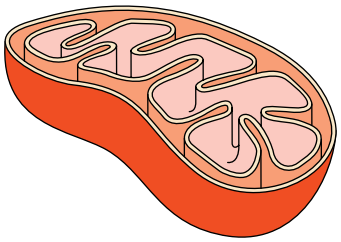
procedure allows for the complete killing of pathogenic bacteria without the use of antibiotics.

Targeting Mitochondria

A mitochondrion is a membrane-enclosed organelle, about 0.5 to 1.0 μm in diameter (Fig. 5.19A), found in most eukaryotic cells that plays a critical role in the production of ATP and in cellular energy metabolism. The human mitochondrial genome consists of circular DNA 16,569 bp

Figure 5.19 **(A)** Targeting proteins to the mitochondria. Shown is a schematic view of a typical mitochondrion. **(B)** Amino acid sequence of the TAT peptide. **(C)** Domains of a therapeutic protein directed to the mitochondria, including the TAT peptide, MTS (a mitochondrial transit sequence), and LAD (the enzyme lipoamide dehydrogenase).

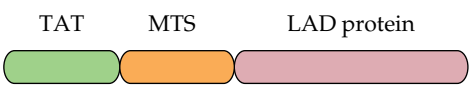
A



B

Tyr-Gly-Arg-Lys-Lys-Arg-Arg-Gln-Arg-Arg-Arg

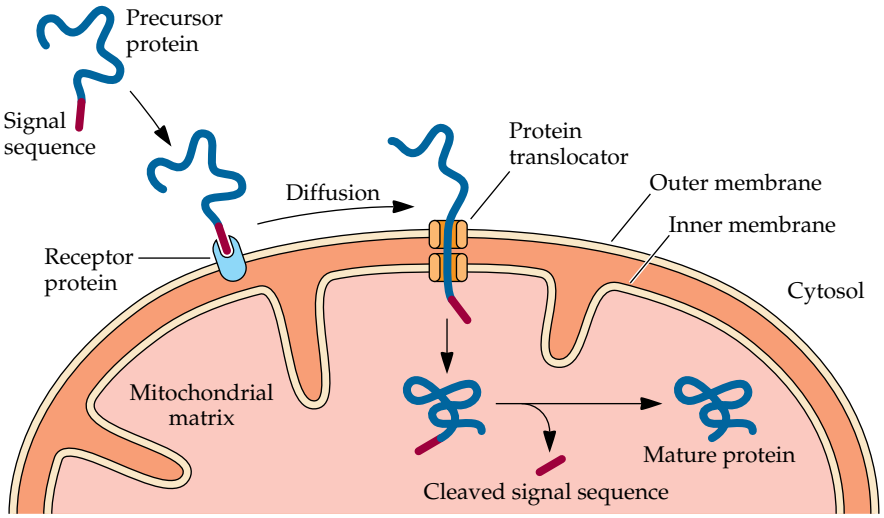
C



in size, with 37 genes encoding 37 separate proteins. Moreover, more than 1,000 additional proteins that function within the mitochondrion are encoded in the nuclear DNA, synthesized in the cytosol, and subsequently imported into the mitochondria across the characteristic double membrane. The nucleus-encoded proteins are synthesized with an N-terminal mitochondrial targeting sequence (15 to 70 amino acids long), which allows them to be imported into the mitochondria (Fig. 5.20). Once a protein is inside, the mitochondrial targeting sequence is recognized by a specific peptidase and cleaved off, allowing the protein to function without any impediment.

Mitochondrial dysfunction can lead to a wide range of disorders. It has been estimated that mitochondrial disease affects approximately 1 in every 8,000 people. Most mitochondrial disorders are caused by defects in nuclear genes. They typically appear as neurological disorders, but they can also manifest as myopathy (muscle disease), diabetes, multiple endocrinopathy (endocrine disease and hormone imbalance), or any one of a variety of other systemic disorders. On the other hand, diseases that are caused by mutation of the mitochondrial genes include Kearns–Sayre syndrome (a neuromuscular disorder), MELAS syndrome (mitochondrial myopathy, encephalopathy, lactic acidosis, and stroke; this syndrome is a form of dementia), and Leber hereditary optic neuropathy (a form of vision loss). The vast majority of the diseases that are caused by mutations in mitochondrial DNA are transmitted by a mother to her children, as the zygote derives its mitochondria from the ovum—i.e., mitochondrial DNA is maternally inherited. This is largely because enzymes that might be administered to replace a deficiency due to a mutation must cross both the cell (plasma) membrane and the mitochondrial double membrane. Inefficient entry reduces the bioavailability of the therapeutic protein so that, at present, there are no effective therapies for mitochondrial disorders. However, symptomatic therapy (i.e., reducing the manifestations of

Figure 5.20 Schematic representation of the importation into the mitochondrial matrix of a protein encoded within the nuclear DNA and then synthesized in the cytosol.



the disease) is typically utilized to deal with a variety of mitochondrial disorders, e.g., diabetes.

Currently, enzyme replacement therapy is used to treat several lysosomal storage diseases such as Gaucher disease, Fabry disease, Hunter syndrome, and Pompe disease. Enzyme replacement therapy might also be an effective strategy for mitochondrial disorders using a simple and effective means of directing enzymes to the mitochondrion. One solution is to synthesize the target enzyme as a fusion protein with a protein transduction domain (also known as a cell-penetrating peptide), which is a short (8 to 30 amino acids) cationic peptide that mediates passage across cell membranes. The most widely investigated and tested protein transduction domain is an 11-amino-acid-long arginine- and lysine-rich peptide (Fig. 5.19B) that is equivalent to amino acid residues 47 to 57 from the transcriptional activator of transcription (TAT) protein encoded by HIV. TAT fusion proteins (including the TAT peptide fused to a target therapeutic protein) may be rapidly and efficiently introduced into cultured cells, intact tissues, and live tissues (when injected into mice) while retaining the biological activity of the target protein.

The protein lipoamide dehydrogenase (LAD), the third catalytic subunit of three multicomponent enzyme complexes in the mitochondrial matrix, including pyruvate dehydrogenase, is responsible for the metabolism of carbohydrates and amino acids. A defect in LAD results in extensive metabolic disturbances and biochemical abnormalities in an affected individual. The clinical course of LAD deficiency is variable, presenting in infancy as a neurological disease or later in life with life-threatening episodes of liver failure or myoglobinuria (i.e., the presence of myoglobin in the urine). In one series of experiments, the ability of the TAT peptide to help to deliver human LAD enzyme into human cells in culture was examined. In this case, the genetic construct that was used encoded the TAT peptide, a mitochondrial targeting sequence peptide (MTS), and the LAD enzyme (Fig. 5.19C). When it was attached to the TAT peptide, the enzyme LAD was efficiently delivered into cells in culture and to their mitochondria, restoring LAD activity to normal values. In a subsequent study using LAD-deficient mice, it was shown that TAT was able to mediate enzyme replacement therapy *in vivo*. That is, it was possible to demonstrate the delivery of the TAT–MTS-LAD protein into deficient mouse tissues, and a single administration of the TAT–MTS-LAD protein resulted in a significant increase in the enzymatic activity of the mitochondrial multienzyme complex pyruvate dehydrogenase within the liver, heart, and brain in the treated mice. Following the demonstration that this tactic works well with human cells in tissue culture, the researchers involved in this study patented their method. If this approach turns out to be as effective *in vivo* in humans as in mice, it has the potential to provide a general and highly effective means of treating individuals who have a wide range of different mitochondrial disorders.

Bacteria and Therapeutics

It has been estimated that the human body contains more bacterial cells than human cells. Some of the bacteria that colonize human tissues are commensal (they coexist without having any appreciable effect on humans). Other bacteria have a mutualistic relationship with various human

tissues and, as a result, facilitate human health. Some bacteria are harmful to human health and are either pathogenic, produce metabolites that impair human health, or can modify or metabolize various drugs, thereby rendering certain treatments ineffective. In this section, the discussion primarily focuses on the use of mutualistic bacteria to deliver therapeutic agents. In addition, there is some discussion of how some bacteria can negatively impact therapeutic treatment.

Lactic acid bacteria are widely used in the production and preservation of fermented foods, and many have been given the designation "generally regarded as safe" according to the U.S. FDA. Included in this group of organisms are various species of *Lactobacillus*, *Leuconostoc*, *Lactococcus*, *Pediococcus*, and *Streptococcus*. Many of these organisms are members of the indigenous microflora of the human gut and have been recognized for their health-promoting properties. Some strains of lactic acid bacteria, notably lactobacilli, are used in probiotic products that are often sold in health food stores. A probiotic is a live microorganism that confers a health benefit by altering the indigenous microflora of the intestinal tract. Lactic acid bacteria have also been used to treat several gastrointestinal disorders, including lactose intolerance, traveler's diarrhea, antibiotic-associated diarrhea, infections caused by various bacterial and viral pathogens, and immunopathological disorders, such as Crohn's disease and ulcerative colitis. Not only are lactic acid bacteria easy to deliver orally, but they also have the potential to elicit both mucosal and systemic immune responses. Mucosal delivery of therapeutics is thought to increase their efficacy and specificity while simultaneously minimizing the side effects of systemic administration.

Quite recently, scientists have discovered that some of the trillions of bacteria that form the gut microbiota have the ability to metabolize and therefore modify a number of drugs and toxins that are ingested by humans. Thus, gut bacteria can sometimes modify the efficacy of various drugs that are taken orally and used to treat a number of different diseases. Notwithstanding the currently limited number of examples of this phenomenon, it is likely that, given the very large number of different bacteria within the human gut, this is much more common than is believed at the present time.

In the past several years, lactic acid bacteria have been used as a host system to express various foreign genes (Table 5.3), with the idea that these bacteria facilitate the delivery of the encoded proteins to the human gut. In particular, *Lactococcus lactis* has been developed as a host for this purpose. *L. lactis* is a nonpathogenic, noninvasive, noncolonizing Gram-positive bacterium that is often used in the production of fermented foods and has been used for many years as a human probiotic. Most *Lactococcus* strains are able to survive passage through the human gastrointestinal tract; however, they are generally unable to colonize it. On the other hand, *Lactobacillus* strains typically survive passage through the human gastrointestinal tract and can subsequently colonize it. Moreover, unlike *E. coli*, eukaryotic proteins produced in *L. lactis* generally do not form insoluble inclusion bodies. In addition, *L. lactis* can be genetically engineered to secrete the target protein facilitating its delivery directly to the target tissue; therefore, it is not necessary to purify the target protein. Finally, it is worth noting that there is a distinction between the use of *L. lactis* to

Table 5.3 Some therapeutic proteins that have been expressed in *Lactococcus lactis*

Therapeutic protein	Target pathogen or disease
Tetanus toxin fragment C	Tetanus
β-toxoid	*Clostridium perfringens*
GroEL heat shock protein	*Brucella abortus*
LcrV antigen	*Yersinia pseudotuberculosis*
MrpA structural fimbrial protein	*Proteus mirabilis*
Outer membrane protein Cag12	*Helicobacter pylori*
Urease subunit B	*Helicobacter pylori*
PspA, PsaA, or PppA antigen	*Streptococcus pneumoniae*
Pili proteins	Group B *Streptococcus*
C-repeat region of M protein	*Streptococcus pyogenes*
Listeriolysin O	Listeriosis
E7 of human papillomavirus 16	Cervical cancer
L1 capsid protein of human papillomavirus 16	Cervical cancer
NSP4 or VP7 antigen	Rotavirus
VP2 and VP3 antigens	Infectious bursal disease virus
Envelope protein	HIV
Nucleocapsid protein	SARS virus
Envelope protein E domain III	Dengue virus
MSP-1	*Plasmodium yoelii* (malaria)
Cyst wall protein 2	*Giardia lamblia*
Murine IL-10	Inflammatory bowel disease
Human IL-10	Crohn's disease
Murine trefoil factors	Inflammatory bowel disease
Mig and IP-10 cytokines	Cancer immunotherapy
Human leptin	Weight control
Human IFN-β	Ulcerative colitis
Ovine IFN-ω	Enteric viral infections
Catalase	Gut cancer and inflammatory diseases
Rat heme oxygenase-1	Hemorrhagic shock or endotoxemia
Major birch pollen allergen	Allergy
Murine IL-12	Cancer immunotherapy, allergy, asthma
LerV protein	Inflammatory bowel disease
Cyanovirin	HIV
Epidermal growth factor	Intestinal development in newly weaned infants
Human fibroblast growth factor 2, IL-4, and colony stimulating growth factor 1	Wound healing in diabetes
Lysostaphin	*Staphylococcus aureus*

IL-10, interleukin-10; IL-4, interleukin-4; SARS, severe acute respiratory syndrome.

deliver human or animal therapeutic agents and its use as a live vaccine. In the former instance, the therapeutic agent typically has biological activity of its own, while in the latter case, the bacterium delivers a protein that elicits an immune response.

In the past 10 to 15 years, there has been a significant increase in the number of therapeutic agents (and vaccines) that have been delivered to mucosal surfaces using lactic acid bacteria. While most of these studies have been performed with mice, positive results have also been obtained with pigs, chickens, ferrets, and humans. By the end of 2019, there were

Table 5.4 Clinical trials using lactic acid bacteria to deliver recombinant proteins

Recombinant protein	Bacterial host	Use	Clinical trial phase
Influenza vaccine antigen	*Lactococcus lactis*	Influenza vaccine	I
Respiratory syncytial virus-F antigen	*Lactococcus lactis*	RSV vaccine	I
Human IL-10	*Lactococcus lactis*	Ulcerative colitis	II
Human trefoil factor-1	*Lactococcus lactis*	Oral mucositis	II
Anti-TNF α antibody fragment	*Lactococcus lactis*	Inflammatory bowel disease	I
Human IL-10 and proinsulin	*Lactococcus lactis*	Type I diabetes	I
Human papillomavirus 16 E7	*Lactobacillus casei*	Human papillomavirus vaccine	II
Anti-human immunodeficiency virus antibody fragment	*Lactobacillus jensenii*	Human immunodeficiency virus prevention	I (planned)

seven ongoing and one planned clinical study employing lactic acid bacteria as therapeutic delivery systems (Table 5.4). A few examples of these studies are given below.

Interleukin-10

Ulcerative colitis and Crohn's disease, both inflammatory diseases of the intestinal tract, affect approximately 1 in every 500 to 1,000 people in the developed countries of the world. Ulcerative colitis is associated with an excess of type 2 T helper cell cytokines, including interleukin-4 and interleukin-5, whereas in Crohn's disease, type 1 T helper cell cytokines, including TNF-α, IFN-α, and interleukin-2, are overproduced. The treatment for Crohn's disease often includes trying to lower the levels of cytokines, especially TNF-α. Another approach to control Crohn's disease is to increase the level of anti-inflammatory cytokine interleukin-10 because it modulates the regulatory T cells that control inflammatory responses to intestinal antigens. However, interleukin-10 currently must be administered by either frequent injections or rectal enemas and therefore is not a convenient treatment. To overcome this problem, the bacterium *L. lactis* was engineered to synthesize and secrete interleukin-10.

Experiments were performed with mice to test whether interleukin-10-secreting *L. lactis* could be used to treat inflammatory bowel disease, where inflammatory bowel disease is a general term that includes both Crohn's disease and ulcerative colitis (Fig. 5.21A). In two mouse models of ulcerative colitis, mice were fed interleukin-10-secreting *L. lactis*: one in which inflammation of the colon was induced by the addition of 5% dextran sulfate to the drinking water and one that is genetically incapable of synthesizing interleukin-10 and provided an animal model for ulcerative colitis. In both of these cases, the engineered *L. lactis* significantly alleviated the symptoms of the disease, establishing that this approach works in principle. However, these mouse models for inflammatory bowel disease are not identical to the disease in humans, and many questions remained before the treatment could be used with humans.

One concern about the use of an interleukin-10-secreting *L. lactis* strain as a therapeutic approach is the possibility that the genetically modified bacterium will be, unintentionally, released into the environment. If this were to happen, the plasmid carrying the interleukin-10 gene and any plasmid-borne antibiotic resistance marker genes could be spread to other

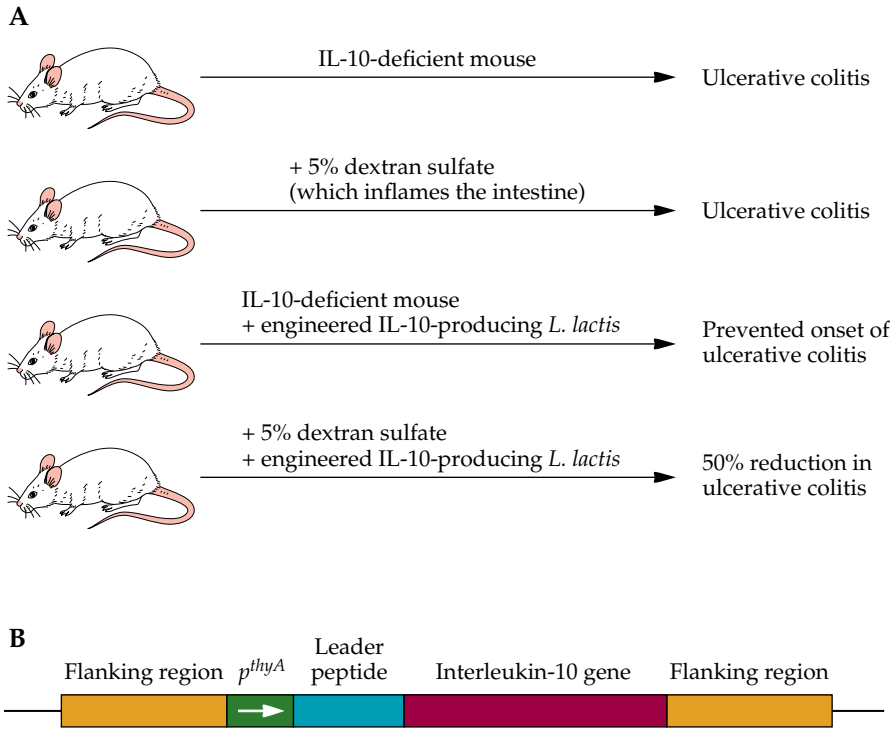


Figure 5.21 **(A)** Schematic representation of the effects of intestinal interleukin-10 (IL-10)-secreting *L. lactis* bacteria on inflammatory bowel disease in mice. **(B)** Genetic construct integrated into the chromosomal DNA of *L. lactis* in place of its thymidylate synthase gene. The promoter (p^{thyA}) is from the *L. lactis* thymidylate synthase gene. The IL-10 gene was chemically synthesized so that its codon usage was optimized for expression in *L. lactis*, thereby ensuring a high level of protein expression.

bacteria in the environment. To prevent this from occurring, a synthetic human interleukin-10 gene was inserted, by homologous recombination, into the thymidylate synthase gene, *thyA*, in the bacterial chromosome of *L. lactis* (Fig. 5.21B). Insertion disrupted the function of the *thyA* gene, which is essential for the production of thymine and thymidine and therefore for the growth of the bacterium. This strain produced interleukin-10 and grew well in the laboratory when either thymidine or thymine was added to the culture medium. However, when it was deprived of thymidine and thymine, the viability of the bacterium declined by several orders of magnitude. When this modified bacterium was tested in pigs, whose digestive tract is similar to that of humans, they thrived and actively produced interleukin-10. In addition, laboratory experiments demonstrated that the modified *L. lactis* was extremely unlikely to acquire a thymidylate synthase gene from other bacteria in the environment, confirming both the safety and efficacy of this approach.

Phase II clinical trials with this interleukin-10-secreting *L. lactis* strain have been initiated. To date, with individuals receiving this treatment, a significant decrease in Crohn's disease symptoms has been observed, with only minor adverse events. Moreover, engineered *L. lactis* bacteria isolated from the patients' feces were not able to grow without the addition of thymidine, indicating that the engineered *L. lactis* did not acquire a

thymidylate synthase gene and the containment strategy was effective. Thus, initial indications are that this strategy appears to be working as well in humans as it did in small animals.

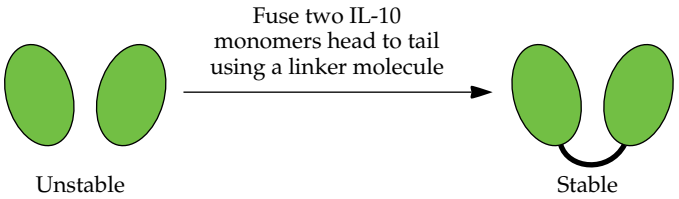
One of the problems with human interleukin-10 is that this protein has a very short half-life so that frequent treatments are required. Fortunately, the half-life of this protein is extended somewhat by either pegylation or by producing it in *L. lactis*. To create a more stable, and hopefully more efficacious, version of interleukin-10, researchers have developed a modified version of this molecule. In its native form, interleukin-10 is a noncovalent homodimer of two 160-amino-acid-long subunits, whereas the engineered version is expressed as a single polypeptide comprising the two subunits joined by a 15-amino-acid-long glycine and serine spacer peptide (Fig. 5.22). The size of the spacer was modeled by a computer algorithm to allow for the proper folding of the fusion protein. When this protein was tested *in vitro*, it showed a significantly increased half-life and (in limited *in vitro* tests) appeared to have a similar range of biological activity compared to the native protein. However, it remains to be seen whether this molecule will be as effective as hoped when it is expressed in lactic acid bacteria and then used to treat mice and other animals.

In a more recent study, instead of using interleukin-10, scientists employed *L. lactis* strains to deliver the serine protease inhibitors elafin and secretory leukocyte protease inhibitor to treat inflammatory bowel disease in mice. Initial *in vivo* results indicate that the treatment of mice with these protease inhibitors was more effective than treatment with interleukin-10. However, to date, this approach has not been tested in humans.

Leptin

It has been estimated that approximately 30% of the North American and 20% of the European populations are overweight. Moreover, North Americans annually spend tens of billions of dollars on various weight reduction schemes, most of which are unsuccessful. However, it is possible that real weight reduction may be obtained by administration of the protein leptin. Leptin, the product of the *obese* (*ob*) gene, is a 167-amino-acid protein with a molecular mass of approximately 16 kDa. Leptin is synthesized as a precursor with a 21-amino-acid-long signal peptide that is removed when leptin is secreted. Treatment with recombinant leptin can reduce food intake and correct metabolic perturbations in (homozygous) leptin-deficient mice. Leptin administration also helps to overcome human congenital leptin deficiency. However, when it is introduced subcutaneously

Figure 5.22 Schematic representation of the unstable, native homodimeric interleukin-10 and its conversion from an unstable protein with two homodimers into a stable interleukin-10 derivative in which the two monomers are joined, head to tail, by a 15-amino-acid-long linker in a fusion protein.



(under the skin), leptin is not particularly effective in obese patients unless their serum leptin concentrations reach levels 20- to 30-fold higher than normal. This poor response has been attributed to the inefficient transport of leptin across the blood-brain barrier to the brain, where it is active. To overcome this problem, a scheme for the intranasal delivery of leptin has been devised. The nasal mucosa is highly vascularized, so delivery of a thin layer of medication across its broad surface area can result in rapid absorption of the medication into the bloodstream, bypassing the blood-brain barrier.

When leptin is produced in *E. coli,* it typically forms insoluble inclusion bodies that must be solubilized and renatured before the active protein is generated. This is a time-consuming, inefficient, and expensive process. To avoid this, in one study, the 462-bp cDNA for human leptin without its signal peptide was cloned and expressed under the control of an *L. lactis* signal peptide and the nisin promoter in *L. lactis* (Fig. 5.23), which is induced by nisin, a 34-amino-acid-residue polycyclic peptide from *L. lactis* with antibacterial activity. In *L. lactis*, leptin was produced efficiently without the formation of inclusion bodies and was secreted from the recombinant bacteria. When the leptin-producing *L. lactis* strain was administered intranasally in obese mice, the mice significantly reduced their food intake and body weight. This approach opens up the possibility that, if delivered properly, leptin might act as an effective weight loss treatment in humans.

It is usually essential that therapeutic proteins being delivered to a particular site be secreted from the producing lactic acid bacteria so that they are readily available and not maintained within the bacterial cell. Until recently, most researchers have relied upon the bacteria's secretory pathway to do this. However, a high level of expression of a foreign protein that is engineered to be secreted can overload the secretory pathway

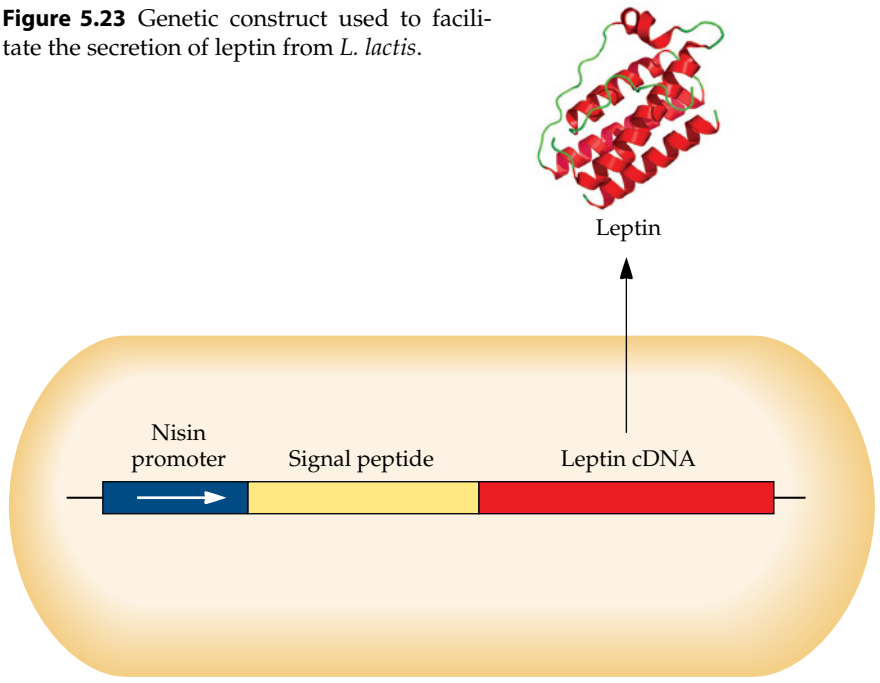


Figure 5.23 Genetic construct used to facilitate the secretion of leptin from *L. lactis*.

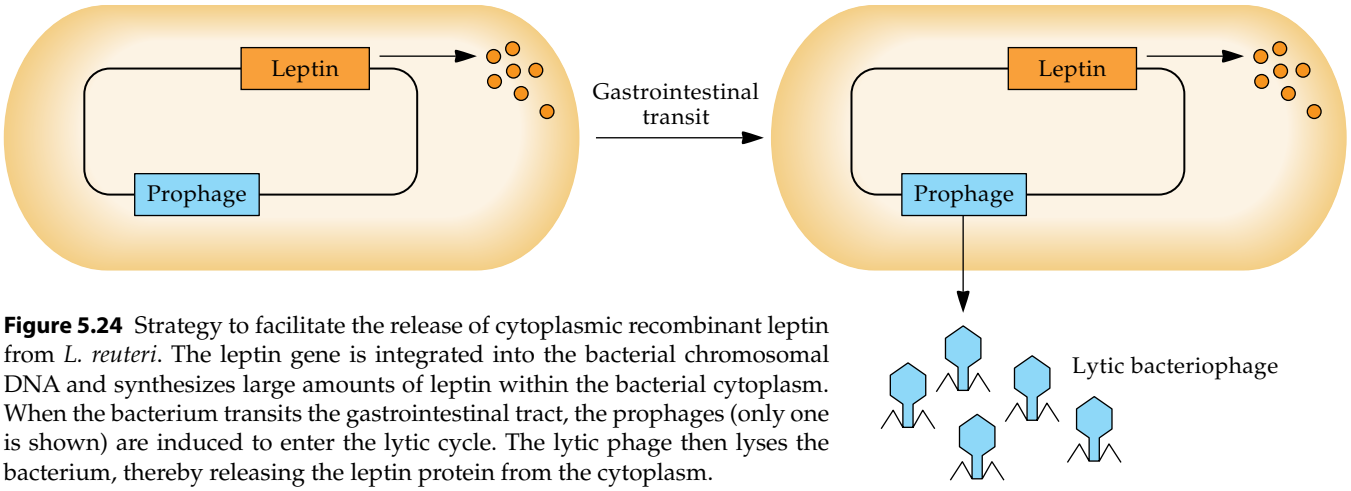


Figure 5.24 Strategy to facilitate the release of cytoplasmic recombinant leptin from *L. reuteri*. The leptin gene is integrated into the bacterial chromosomal DNA and synthesizes large amounts of leptin within the bacterial cytoplasm. When the bacterium transits the gastrointestinal tract, the prophages (only one is shown) are induced to enter the lytic cycle. The lytic phage then lyses the bacterium, thereby releasing the leptin protein from the cytoplasm.

of the delivery bacterium, thereby limiting the usefulness of this approach. In one recent experiment, scientists depended upon the natural presence of two active prophages in the *L. reuteri* genome, which are induced to become lytic during gastrointestinal transit of the lactic acid bacterium that was employed. First, the bacterium was engineered to synthesize but not secrete leptin within the bacterial cell. Following ingestion of the bacterium and the subsequent transit of the bacterium through the gastrointestinal tract, the prophages were induced to become lytic bacteriophages. When the bacteriophages lysed the *L. reuteri* bacterial host cells, the recombinant protein (leptin) that was inside the cells was released, thereby obviating the need to utilize the cellular secretory machinery (Fig. 5.24). While this method is claimed to be superior to the traditional approach of engineering leptin to be secreted by the engineered bacterium, a considerable amount of testing remains to be done before it can be used as a therapeutic delivery system.

An HIV Inhibitor

Worldwide, the predominant mode of HIV transmission is by heterosexual contact. One possible way to protect women, who currently comprise about half of all new cases of HIV/AIDS, against HIV infection is a topical microbicide, delivered by a live vaginal *Lactobacillus* strain, that prevents HIV infection directly at mucosal surfaces. This strategy seems reasonable because naturally occurring vaginal *Lactobacillus* strains create a low-pH environment and thereby play a protective role in preventing female urogenital infections.

The protein cyanovirin N, isolated from the cyanobacterium *Nostoc ellipsosporum*, blocks several steps of HIV infection, including virus entry into human cells. Consequently, cyanovirin N is a candidate for a topical microbiocide to prevent HIV (and possibly other viral) infections and for delivery by *Lactobacillus*.

Cyanovirin N is a cyanobacterial lectin, a protein that binds strongly to specific carbohydrate residues. This small, two-domain protein neutralizes HIV by specifically binding to high-mannose glycans

(oligosaccharides) on the viral surface envelope glycoproteins gp120 and gp41, as well as to high-mannose oligosaccharides found on the HIV envelope. The binding of cyanovirin N to these carbohydrate residues thereby prevents interaction of the virus with a mammalian host cell. In addition to its activity against HIV, cyanovirin N is also active against other enveloped viruses, including influenza, Ebola, hepatitis C, and some strains of herpesvirus.

To ensure that cyanovirin N is expressed at a sufficiently high level in a vaginal strain of *Lactobacillus jensenii*, the gene was chemically synthesized to reflect the codon usage found in the host bacterium rather than in the cyanobacterium *N. ellipsosporum*. Typically, the GC content of the lactobacillus genome is about 36%, while the GC content of the *N. ellipsosporum* genome is 44.5%, indicating that the codon usage of these two organisms is quite different. In addition, during the chemical synthesis of the gene, the codon for proline 51 was replaced by a codon for a glycine residue to stabilize the cyanovirin N protein, and four amino acids were added to the N terminus of the protein to ensure proper cleavage of the signal sequence (Fig. 5.25). The modified cyanovirin N gene was fused to a strong and constitutive *Lactobacillus* promoter. The final construct was integrated into the chromosomal DNA of a strain of *L. jensenii*, with the result that about 4 μg of cyanovirin N per ml was released into the culture medium. When the modified *L. jensenii* strain was tested for efficacy, it was found to be highly effective at preventing HIV infections in mice. The modified cyanovirin N and the wild type were tested for the ability to neutralize 33 different strains of HIV, with the modified strain being more potent than the wild type against all of the strains tested.

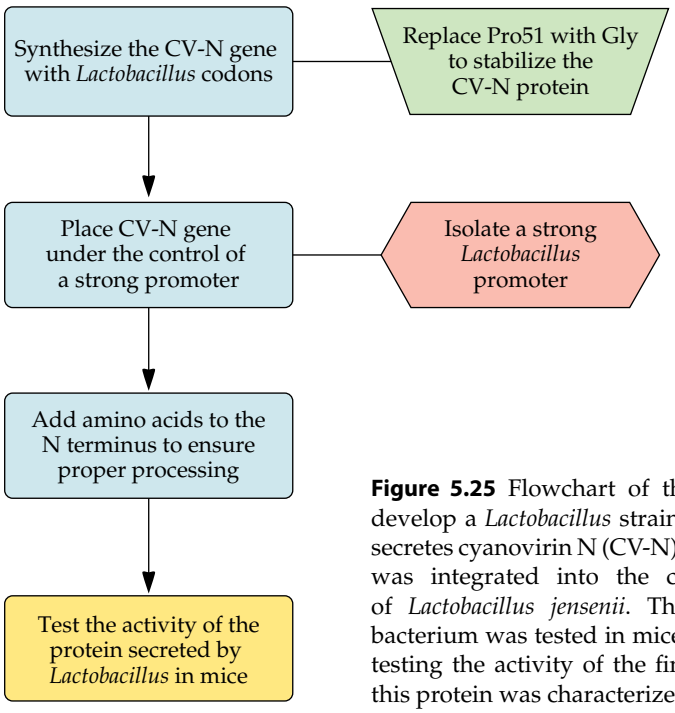


Figure 5.25 Flowchart of the scheme used to develop a *Lactobacillus* strain that produces and secretes cyanovirin N (CV-N). The final construct was integrated into the chromosomal DNA of *Lactobacillus jensenii*. The final engineered bacterium was tested in mice. Note that prior to testing the activity of the final protein in mice, this protein was characterized *in vitro*.

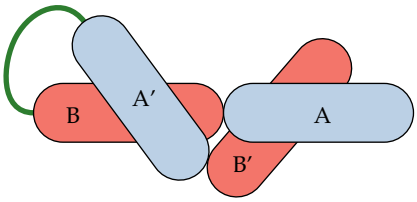


Figure 5.26 Schematic representation of head to tail linked molecules of cyanovirin-N with a peptide linker molecule between the monomeric subunits. In the dimeric form, domain A interacts with B′ and domain A′ interacts with B. The linker peptide is shown as a dark green line.

In solution, cyanovirin N is a monomeric protein, and each monomeric unit contains two symmetrically related structural domains (A and B), with each domain containing a carbohydrate binding site that specifically interacts with α(1-2) linked oligomannose residues. However, despite the fact that cyanovirin N is a monomer in solution, it forms a stable dimer when the protein is crystallized. This observation suggested to researchers that by covalently linking two cyanovirin N molecules together (i.e., synthesizing them as a single polypeptide) in a head to tail manner (encoded at the DNA level), it might be possible to stabilize the dimeric form in solution. In the dimer structure, domain A (from one molecule) physically interacts with domain B′ (from the other molecule) to form a "monomer-like unit," while domains A′ and B interact with one another (Fig. 5.26) to form a second "monomer-like unit." Most importantly, the linked dimers show enhanced HIV neutralization activity (up to 18-fold greater) compared to the wild-type form of this molecule. Of course, the next logical step in this work would be to express cyanovirin N linked dimers in a strain of *L. jensenii* as described above.

Insulin

Type I diabetes is a common metabolic disorder (it affects 40 to 45 million persons worldwide) in which the pancreas does not produce a sufficient amount of the hormone insulin to reduce blood sugar levels. Type I diabetics, who must be treated with insulin, account for approximately one-tenth of all diabetics worldwide, while type II diabetics, who do not usually require the provision of external insulin (except for those with advanced type II diabetes), make up the other 90% of diabetics. If left untreated, individuals with type I diabetes can expect to develop cardiovascular disease, stroke, kidney failure, foot ulcers, and damage to the eyes. For the majority of individuals with type I diabetes, a T-cell-mediated autoimmune attack leads to the loss of insulin-producing beta cells in the islets of Langerhans in the pancreas. Since insulin is needed for the regulation of carbohydrate and fat metabolism, people who are unable to synthesize sufficient insulin require (usually several times) daily subcutaneous injections of recombinant insulin.

Oral administration of insulin would greatly simplify the lives of most insulin-dependent diabetics. However, this has not been possible due to insulin's poor stability during its passage through the gastrointestinal tract. As an alternative to various unsuccessful strategies to produce encapsulated insulin that could be delivered orally, one group of researchers developed a single-chain insulin molecule that can be delivered to diabetic patients using the bacterium *L. lactis*.

The peptide hormone insulin is produced in animal pancreatic cells as a single polypeptide, preproinsulin, that is processed to proinsulin following its secretion and then to insulin following proteolytic cleavage (Fig. 5.27A). The insulin A chain includes 21 amino acid residues, while the B chain consists of 30 amino acid residues. Given the presence of two interchain and one intrachain disulfide bond (not shown in Fig. 5.27), it

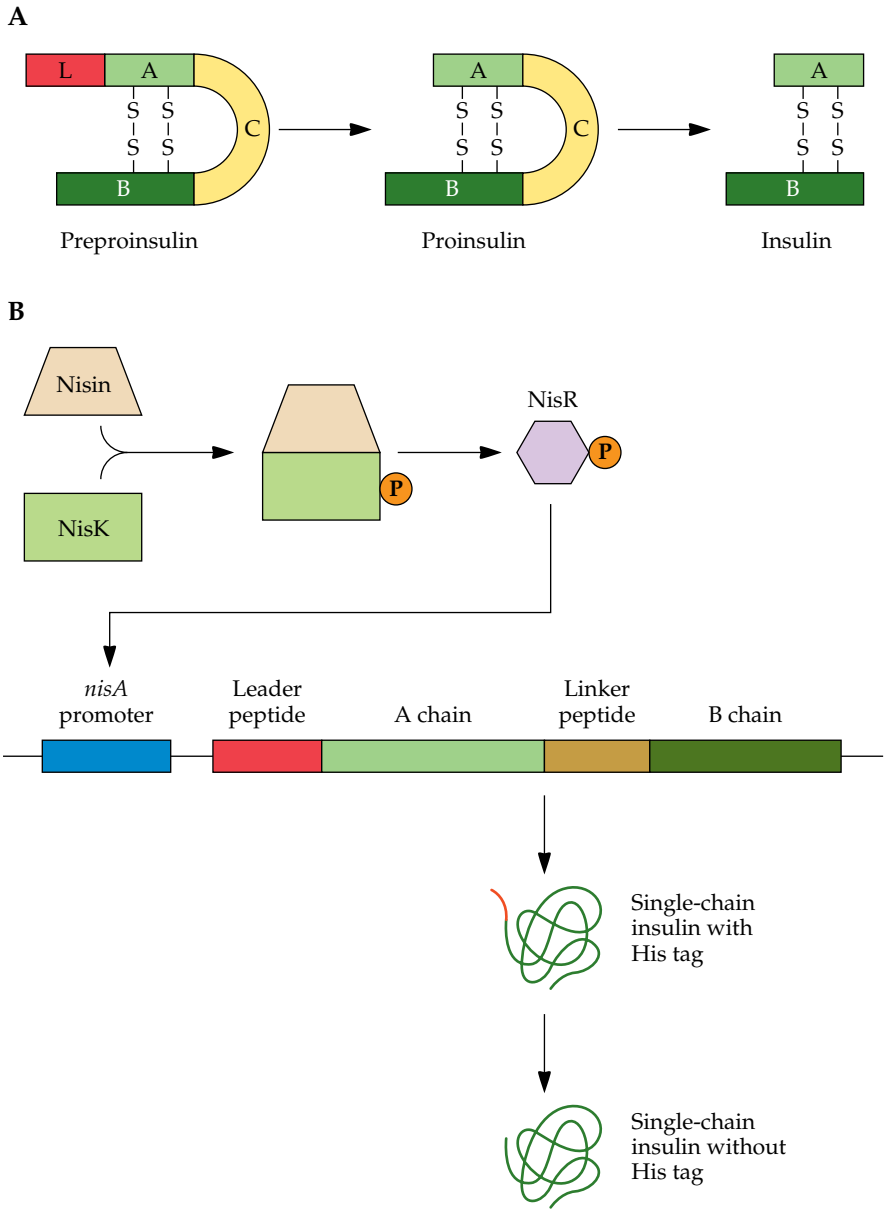


Figure 5.27 **(A)** Cleavage of inactive preproinsulin, first to proinsulin and then to yield active insulin. Proteases remove the leader peptide (L) and an internal peptide (C) to yield a peptide that consists of chains A and B. **(B)** Genetic regulation of the expression of a single-chain (sc) form of insulin in *L. lactis*. The binding of nisin, which is added to the medium, to NisK (a histidine kinase) causes NisK to autophosphorylate. NisK then transfers its phosphate to NisR, which activates the *nisA* promoter, inducing transcription of the single-chain insulin gene. The leader peptide ensures that the single-chain insulin will be secreted. Following secretion and removal of the leader peptide, a histidine (His) tag that is part of the protein facilitates its purification before the His tag is removed by proteolytic cleavage to yield the single-chain form of insulin.

is technically quite difficult to produce insulin in *L. lactis*. Therefore, it was decided to synthesize a single-chain insulin in which the A and B chains are joined covalently by a 6-amino-acid-long linker peptide. To ensure that the single-chain insulin was as active as native insulin, four additional amino acid changes were made to the protein (introduced at the DNA level). Insulin derivatives with two of these amino acid changes were already in clinical use when this work was undertaken, making the researchers confident that the proposed four-amino-acid changes would yield an active and stable form of insulin. The resulting single-chain insulin is more stable at high temperature and is less likely to aggregate than native insulin. The single-chain insulin gene was expressed in *L. lactis* (Fig. 5.27B) under the control of the *nisA* promoter, which is activated by a phosphorylated form of the transcription factor NisR in the presence of the antibacterial peptide nisin. The NisR protein is phosphorylated by the transfer of a phosphate residue from a NisK protein bound to the peptide nisin. To date, it has been demonstrated that genetically engineered *L. lactis* can produce and secrete a single-chain insulin that has *in vitro* insulin biological activity. Whether this system will function effectively as an insulin delivery system, replacing at least once-a-day and often several-times-daily insulin injections for human diabetics, remains to be determined. Several other groups of scientists have also synthesized single-chain insulin derivatives, all of which are active. It remains to be determined which of these insulin derivatives is most beneficial to type I diabetics.

Treatment of excessive amounts of blood sugar (hyperglycemia) with insulin may decrease an individual's blood sugar amount to a dangerously low level (hypoglycemia). To avoid this potentially unsafe side effect of insulin treatment, several labs are in the midst of developing insulin treatments that are responsive to an individual patient's blood sugar level. At this stage, it is too early to predict which, if any, of the current approaches might be successful. Nevertheless, it is important to understand how this might work (Fig. 5.28). The derivatized insulin would bind to a protein-binding element when blood glucose levels are low and subsequently be dissociated from this protein-binding element

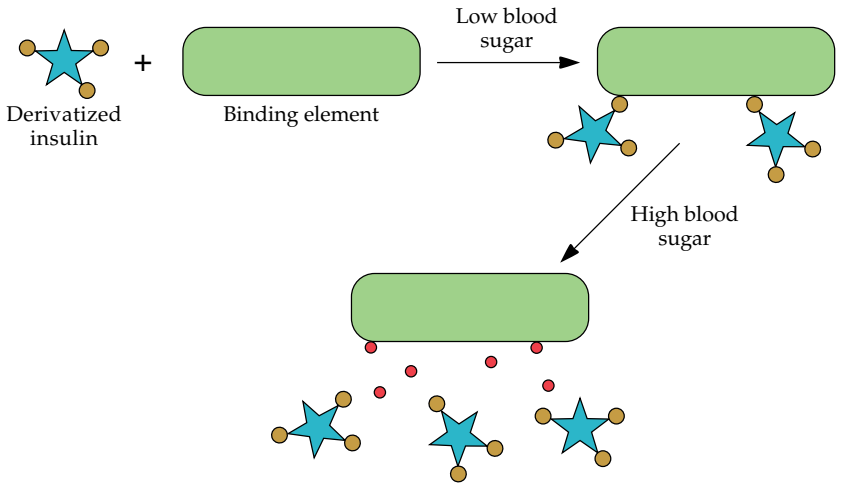


Figure 5.28 Schematic representation of a glucose-responsive insulin. The derivatized insulin binds to a protein-binding element when the blood sugar is low and prevents it from otherwise binding to insulin receptors which would further lower the blood sugar level. Following a high-sugar meal, the glucose molecules bind to the binding element and remove the bound derivatized insulin; the derivatized insulin molecules now act to lower the level of blood sugar.

when glucose levels increase. Once researchers have developed a workable system, it would be beneficial to express this system within a human-compatible strain of *L. lactis*.

Parkinson's Disease

Parkinson's disease is a long-term degenerative disorder of the central nervous system that affects approximately 10 million people worldwide. For more than 50 years, the primary treatment for this disease has been oral administration of the chemical compound L-DOPA. While L-DOPA provides some relief to Parkinson's disease sufferers, over time, this treatment (which is not a cure) becomes progressively less effective. This is a consequence of the L-DOPA being converted to dopamine by bacterial enzymes that are typically present in the human gut (and whose concentration is increased over time) and the dopamine being unable to cross the human blood-brain barrier. On the other hand, L-DOPA readily crosses the human blood-brain barrier, and once inside the brain, it is enzymatically converted to dopamine, which replaces the dopamine normally produced by brain cells but is present at a lowered level in Parkinson's patients. To address the decreasing effectiveness of L-DOPA treatment over time, scientists administered the compound alpha-fluoromethyltyrosine at the same time as L-DOPA and found that it inhibited the gut bacterial enzyme (L-DOPA decarboxylase from the bacterium *Enterococcus faecalis*) that prematurely converted L-DOPA to dopamine (Fig. 5.29). In the presence of this inhibitory compound, the L-DOPA is no longer prematurely converted to dopamine in the gut, and more of the drug crosses the blood-brain barrier. This is one example of a particular gut bacterium producing a specific enzyme that inactivates a therapeutic treatment. Notwithstanding the laboratory success of this experiment, a considerable amount of work remains to be done in terms of testing this approach in small animals and then in humans.

Cancer and Bacteria

Recent reports strongly suggest that the human intestinal microbiome can have a significant effect on the efficacy of various cancer treatments. For example, as discussed in some detail later in this chapter, one way to treat various human tumors is the "blockage of immune checkpoints," where immune checkpoint receptors are found on the surface of human T cells and are responsible for preventing the immune system from attacking an individual's own tissues and cells. While it is possible to target immune checkpoints and allow a strong antitumor response (see Fig. 5.49), some researchers have observed that patients that respond well to this treatment have a high abundance of *Faecalibacterium* spp., while patients that do not respond to this treatment have a high abundance of various *Bacteroidales* in their feces. These data are consistent with certain gut bacteria inactivating the immune checkpoint inhibitors. Thus, it may be possible to alter the composition of gut bacteria and thereby develop approaches that restore the efficacy of these antitumor agents.

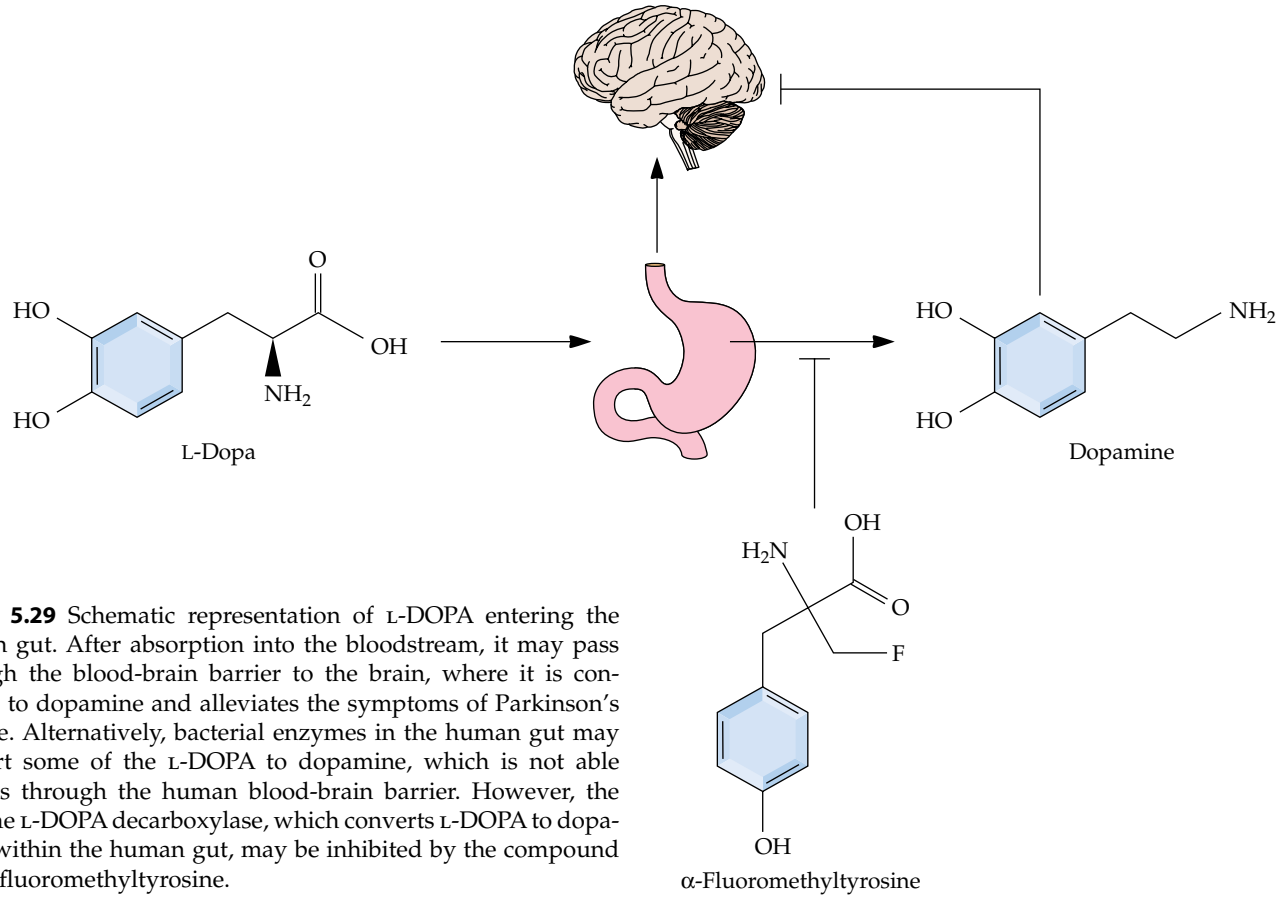


Figure 5.29 Schematic representation of L-DOPA entering the human gut. After absorption into the bloodstream, it may pass through the blood-brain barrier to the brain, where it is converted to dopamine and alleviates the symptoms of Parkinson's disease. Alternatively, bacterial enzymes in the human gut may convert some of the L-DOPA to dopamine, which is not able to pass through the human blood-brain barrier. However, the enzyme L-DOPA decarboxylase, which converts L-DOPA to dopamine within the human gut, may be inhibited by the compound alpha-fluoromethyltyrosine.

Recombinant Antibodies

Around 100 years ago, horses were inoculated with the bacterium *Corynebacterium diphtheriae*, which causes diphtheria in humans. Crude antiserum was then extracted from the horses and used to treat individuals who contracted this, often fatal, childhood disease. In those days, mortality from this disease sometimes reached 45%. *C. diphtheriae* infects the throat or tonsils and produces an exotoxin that is lethal to human cells. Moreover, this exotoxin enters the bloodstream and damages organs that are distant from the primary site of infection. The administration of horse antiserum containing antibodies against the exotoxin provided individuals with passive immunity, protecting the patient from an otherwise fatal outcome, as long as the antiserum was given within the first few days after the onset of infection.

Unfortunately, this kind of antibody therapy carries considerable risk and is not used today. Patients often develop antibodies against the foreign proteins of either whole or partially purified horse antiserum. After a second treatment, the sensitized patient may go into anaphylactic shock and die. As a result, the use of antibodies as therapeutic agents was generally considered too dangerous for patients. However, with the development of hybridoma technology (see chapter 4), antibodies have become common therapeutic agents. One reason for the renewed

interest in therapeutic antibodies is that it is now possible to engineer antibodies with a greatly reduced level of immunogenicity in humans. In addition, hybridoma technology can be used to maintain a continuous supply of pure monospecific antibody. However, the problems of cross-reactivity leading to an immune response and anaphylaxis have not been completely overcome. Thus, the recipient might still produce antibodies to a monoclonal antibody that carries mouse (murine) determinants since monoclonal antibodies are generally produced using mouse cells. To avoid this problem, human monoclonal antibodies with both specific immunotherapeutic properties and lowered potential for immunogenicity have been produced. In fact, a large number of monoclonal antibodies have been approved for treating human diseases (Table 5.5, Box 5.2, and Box 5.3). It should be pointed out that monoclonal antibodies are relatively stable and long-lived compared to other proteins; they typically have a half-life of around 21 days. This long lifetime facilitates the efficacy of monoclonal antibodies and typically provides sufficient time for the antibody molecules to recruit other components of the immune system in destroying the target cells.

box 5.2

Trastuzumab: the First Humanized Monoclonal Antibody Approved for the Treatment of Breast Cancer

In 25 to 30% of women with aggressive metastatic breast cancer, there is a genetic alteration in the *HER2* gene that results in the production of an increased amount of human epidermal growth factor receptor 2 (HER2) protein on the surface of the tumor. Overexpression of the HER2 protein can readily be determined by using an immunohistochemistry-based assay. Some years ago, researchers at Genentech isolated a mouse monoclonal antibody with high affinity for the HER2 protein and then (using procedures similar to those described in this chapter) humanized it. The humanized anti-HER2 monoclonal antibody, trastuzumab (Herceptin), contains human FRs and mouse CDRs and is produced commercially using mammalian (CHO) cells grown in suspension culture as the host for the expression of the antibody. Antibodies produced in CHO cells are glycosylated similarly to bona fide human antibodies. After humanization, trastuzumab bound to the HER2 protein with a dissociation constant of approximately 5×10^{-9} M, indicating that the high level of specificity for the substrate had been maintained through the process of humanization.

In the laboratory, and then in initial clinical trials with more than 800 patients, trastuzumab mediated antibody-dependent cellular cytotoxicity (i.e., it told the immune system to target the cancerous cells) and inhibited the proliferation of human tumor cells that overexpressed HER2 (i.e., it stopped the cancerous cells from growing). Trastuzumab was most effective when it was administered together with some of the chemicals that are currently used for the treatment (chemotherapy) of breast cancer, provided that the breast cancer was at a later stage of development. In two large clinical trials that included over 3,700 women, those who received trastuzumab and chemotherapy had a 52% higher chance that the cancer would not return than those who were treated with chemotherapy alone. Trastuzumab is provided by the manufacturer as a sterile white to pale yellow powder containing 440 mg per vial, and after reconstitution, it is typically administered intravenously over a period of 30 minutes and is taken weekly for 52 weeks. Since a small number of individuals treated with trastuzumab develop heart problems, it is necessary to carefully monitor the cardiac functions of all patients on this therapy, especially older patients and those with a family history of heart problems. In the relatively short time that it has been available, trastuzumab has become a blockbuster drug, with annual sales above $7.4 billion in 2017 and cumulative sales from 1998 to 2017 of ~$27.1 billion. In 2012, in the United States, trastuzumab treatment for one individual cost approximately $70,000 for the year.

Table 5.5 Some therapeutic monoclonal antibodies that have been approved for human use in either the United States or the European Union

Approval date	Antibody	Drug name	Antibody type	Therapeutic use
1986	Muromonab	Orthoclone	Murine	Acute kidney transplant rejection
1994	Abciximab	ReoPro	Chimeric	Blood clots
1997	Daclizumab	Zenapax	Humanized	Acute kidney transplant rejection
1998	Rituximab	Rituxan	Chimeric	Non-Hodgkin's lymphoma
1998	Infliximab	Remicade	Chimeric	Crohn's disease, psoriasis, rheumatoid arthritis
1998	Basiliximab	Simulect	Chimeric	Transplantation rejection
1998	Palivizumab	Synagis	Humanized	Viral infections in children
1998	Trastuzumab	Herceptin	Humanized	Metastatic breast cancer
2000	Gemtuzumab	Mylotarg	Humanized	Acute myeloid leukemia
2001	Alemtuzumab	Leukosite	Humanized	Chronic lymphocytic leukemia
2002	Adalimumab	Humira	Human	Rheumatoid arthritis
2002	Ibritumomab	Zevalin	Chimeric	Non-Hodgkin's lymphoma
2003	Efalizumab	Raptiva	Humanized	Severe plaque psoriasis
2003	Omalizumab	Xolair	Humanized	Severe persistent asthma
2003	Tositumomab	Bexxar	Murine + iodine-131	Non-Hodgkin's lymphoma
2004	Cetuximab	Erbitux	Chimeric	Various cancers
2004	Natalizumab	Tysabri	Humanized	Multiple sclerosis
2004	Bevacizumab	Avastin	Humanized	Various cancers
2006	Panitumumab	Vectibix	Human	Colorectal cancer
2009	Ofatumumab	Arzerra	Human	Treatment of chronic lymphocytic leukemia
2010	Denosumab	Prolia	Human	Osteoporosis
2010	Denosumab	Xgeva	Human	Bone cancer
2011	Belimumab	Benlysta	Human	Systemic lupus erythematosus
2011	Ipilimumab	Yervoy	Human	Late-stage melanoma
2011	Brentuximab vedotin	Adcetris	Chimeric	Hodgkin's lymphoma
2012	Pertuzumab	Perjeta	Humanized	HER2-positive metastatic breast cancer
2013	Obinutuzumab	Gazyvaro	Humanized	Chronic lymphoid leukemia
2014	Pembrolizumab	Keytruda	Humanized	Late-stage melanoma
2014	Ramucirumab	Cyramza	Human	Solid tumors
2014	Siltuximab	Sylvant	Chimeric	Various cancers
2014	Blinatumomab	Blincyto	Bispecific/humanized	Philadelphia chromosome-negative relapsed or refractory B-cell precursor acute lymphoblastic leukemia
2015	Dinutuximab	Unituxin	Chimeric	High-risk pediatric neuroblastoma
2015	Secukinumab	Cosentyx	Human	Psoriasis
2015	Evolocumab	Repatha	Human	High cholesterol
2015	Mepolizumab	Nucala	Humanized	Asthma
2015	Alirocumab	Praluent	Human	High cholesterol
2015	Daratumumab	Darzalex	Human	Multiple myeloma
2015	Elotuzumab	Empliciti	Humanized	Multiple myeloma

2015	Necitumuma	Portrazza	Human	Non-small cell lung cancer
2015	Idarucizumab	Praxbind	Humanized	Dabigatran-induced anticoagulation
2016	Ixekizumab	Talz	Humanized	Psoriasis
2016	Reslizumab	Cinqaero	Humanized	Asthma
2016	Olaratumab	Lartruvo	Human	Soft tissue sarcoma
2017	Bezlotoxumab	Zinplava	Human	*Clostridiodes (Clostridium) difficile*
2017	Atezolizumab	Tecentriq	Humanized	Bladder cancer
2017	Brodalumab	Siliq	Human	Psoriasis
2017	Avelumab	Bavencio	Human	Merkel cell carcinoma
2017	Dupilumab	Dupixent	Human	Atopic dermatitis
2017	Sarilumab	Kevzara	Human	Rheumatoid arthritis
2017	Inotuzumab	Besponsa	Humanized	Hematological malignancy
2017	Guselkumab	Tremfya	Human	Psoriasis
2018	Ocrelizumab	Ocrevus	Humanized	Multiple sclerosis
2018	Benralizumab	Fasenra	Humanized	Asthma
2018	Emicizumab	Hemlibra	Bispecific/humanized	Hemophilia A
2018	Durvalumab	Imfinzi	Human	Bladder cancer
2018	Burosumab	Crysvita	Human	X-linked hypophosphatemia
2018	Ibalizumab	Trogarzo	Humanized	HIV infection
2018	Tildrakizumab	Ilumya	Humanized	Psoriasis
2018	Erenumab	Aimovig	Human	Migraines
2018	Mogamulizumab	Poletigeo	Humanized	Cutaneous T cell lymphoma
2018	Lanadelumab	Takhzyro	Human	Hereditary angioedema
2018	Galcanezumab	Emgality	Human	Migraines
2018	Fremanezumab	Ajovy	Human	Migraines
2018	Cemiplimab	Libtayo	Human	Cutaneous squamous cell carcinoma
2018	Emapalumab	Gamifant	Human	Hemophagocytic lymphohistiocytosis
2018	Ravulizumab	Ultomiris	Humanized	Paroxysmal nocturnal hemoglobinuria
2019	Caplacizumab	Cablivi	Humanized	Acquired thrombotic thrombo-cytopenic purpura
2019	Romosozumab	Evenity	Humanized	Osteoporosis
2019	Risankizumab	Skyrizi	Humanized	Psoriasis
2019	Polatuzumab	Polivy	Humanized	B-cell lymphoma

In addition to the monoclonal antibodies listed here, a number of monoclonal antibodies have been approved for diagnostic and imaging purposes.

box 5.3

Rituximab, Ibritumomab, and Obinutuzumab: Therapeutic Monoclonal Antibodies That Treat Non-Hodgkin's Lymphoma

Non-Hodgkin's lymphoma is a malignant growth of B or T cells of the lymph system. It has been estimated by the American Cancer Society that in 2010, globally, there were more than 200,000 deaths directly attributable to non-Hodgkin's lymphoma. Currently, more than 6 million people worldwide have non-Hodgkin's lymphoma, and the incidence of the disease is growing. It is the fifth most common cancer (although there are about 29 different lymphomas in this category), with an individual's chance of developing the disease in their lifetime being about 1 in 50.

There are a variety of treatments for patients with non-Hodgkin's lymphoma, including radiation therapy, chemotherapy, immunotherapy, bone marrow transplantation, and "watch and wait" for slowly growing cases. In 1997, the FDA approved the use of rituximab (Rituxan) for the treatment of non-Hodgkin's lymphoma. Rituximab is a genetically engineered chimeric (murine/human) monoclonal antibody directed against the CD20 antigen (a protein on the surfaces of B lymphocytes). Following binding of the antibody to CD20, the body's defenses attack and kill the antibody-marked B cells. Stem cells in bone marrow lack CD20, so they are uninhibited by this treatment. Healthy B cells can regenerate from those stem cells, after the completion of the course of rituximab treatment (given once a week for 4 to 8 weeks), and return to normal levels within several months. In 2006, the FDA approved the use of rituximab in combination with CHOP (cyclophosphamide, doxorubicin, vincristine, and prednisone) and other anthracycline-based chemotherapy regimens. In addition, the use of rituximab in combination with the chemical compound methotrexate was approved for the treatment of moderately to severely active rheumatoid arthritis in patients who had been refractory to other treatments.

Notwithstanding some severe side effects in some patients, rituximab has been enormously successful. In 2017, sales of rituximab were approximately $7.8 billion, with cumulative sales from 1998 to 2017 reaching >$29.1 billion. Hundreds of thousands of people worldwide who did not respond well to conventional chemotherapy have been successfully treated with rituximab. In fact, while the incidence of non-Hodgkin's lymphoma continues to increase, since the introduction of rituximab, mortality from the disease in the United States has declined at a rate of approximately 2.3% a year. In 2002, the FDA approved the use of ibritumomab tiuxetan (Zevalin) together with rituximab. Ibritumomab is also a monoclonal antibody that targets B cells. However, ibritumomab is linked to a chemical chelator molecule (tiuxetan) that binds tightly to radioactive indium-111 or yttrium-90. Thus, a therapeutic regimen with ibritumomab tiuxetan targets tumor cells with a high dose of radiation. In 2011, treatment with ibritumomab tiuxetan was priced at approximately $25,000 per month, with treatments typically lasting 1 or 2 months. Treatment with ibritumomab tiuxetan is quite toxic, and around half of the treated individuals experience side effects. Therefore, ibritumomab tiuxetan is approved only for patients who have failed to respond to other treatments.

On 1 November 2013, the FDA approved the use of obinutuzumab (Gazyva or GA101) for the treatment of CD20-positive B-cell non-Hodgkin's lymphoma. Obinutuzumab is a fully humanized monoclonal antibody that, like rituximab, is directed against the CD20 protein found on B cells. Obinutuzumab is seen by the manufacturer as the successor to rituximab. In a direct comparison with rituximab, patients treated with obinutuzumab had significantly longer periods of remission, when they were disease free and did not require treatment.

Hybrid Human–Mouse Monoclonal Antibodies

The modular nature of antibody functions (see chapter 4) has made it possible to convert a mouse monoclonal antibody into one that has some human segments but still retains its original antigen-binding specificity. This hybrid molecule is called a chimeric antibody (Fig. 5.30A), or with more human sequences, a "humanized" antibody (Fig. 5.30B). The difference between a chimeric and a humanized mouse monoclonal antibody is the portion of the mouse antibody that has been removed. In a chimeric antibody, the portion of the mouse monoclonal antibody that was targeted for replacement with a human sequence was the mouse Fc fragment. This is because the mouse Fc fragment functions poorly as an effector of immunological responses in humans, and it is also the most

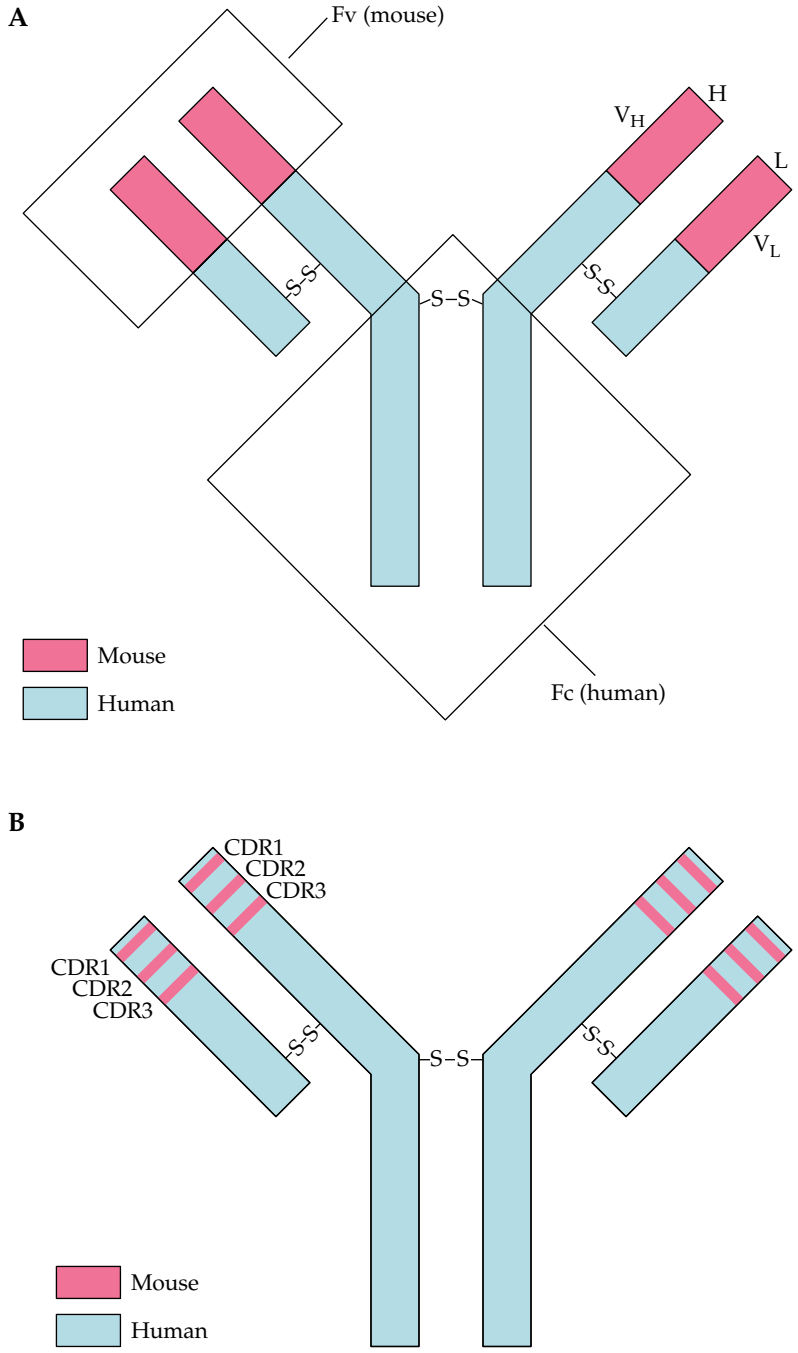


Figure 5.30 Genetically engineered antibodies. **(A)** Chimeric antibody. The V_L and V_H DNA regions from the immunoglobulin L and H genes that encode part of a mouse monoclonal antibody were substituted for the V_L and V_H DNA regions of a human immunoglobulin molecule. The product of the constructed gene is a chimeric (partially humanized) immunoglobulin with the antigen-binding specificity of the mouse monoclonal antibody and lowered both immunogenicity in humans and human Fc effector capabilities. **(B)** Humanized antibody. The CDRs (CDR1, CDR2, and CDR3) from the genes for H and L immunoglobulin chains of a mouse monoclonal antibody replace the CDRs of the genes for a human antibody. The product of this constructed gene is an immunoglobulin with the antigen-binding specificity of the mouse monoclonal antibody and all the other properties of a human antibody molecule.

likely fragment to elicit the production of human antibodies against it. To diminish immunogenicity and retain human Fc effector capabilities while introducing antigen-targeting capability, the DNA coding sequences for the Fv regions of both the L and the H chains of a human immunoglobulin were substituted for the Fv DNA sequences for the L and H chains from a specific mouse monoclonal antibody (Fig. 5.30A). Chimeric antibodies are composed of approximately 70% human and 30% mouse DNA sequences.

The humanizing of mouse monoclonal antibodies has been taken one step further than the formation of chimeric molecules by substituting into human antibodies only the complementarity-determining regions (CDRs) of the mouse monoclonal antibodies (Fig. 5.30B). Humanized antibodies consist of approximately 95% human and 5% mouse DNA sequences. Because these engineered humanized antibodies have antigen-binding affinities similar to those of the original mouse monoclonal antibodies, they are more effective therapeutic agents and are subsequently less likely to generate an undesirable immune response.

The humanizing of mouse monoclonal antibodies may be performed as follows (Fig. 5.31). Starting with a mouse hybridoma cell line, cDNAs for the L and H chains are isolated. The variable regions of these cDNAs are amplified by PCR. The oligonucleotide primers that are used for this amplification are complementary to the sequences at the 5′ and 3′ ends of the DNA encoding the variable regions that are involved in antigen recognition. From the nucleotide sequences of the cDNAs for the L and H regions (V_L and V_H), it is possible to delineate the limits of the CDRs. It is usually straightforward to determine where the CDRs begin and end, because these regions are highly variable in sequence, while the sequences of the framework regions (FRs) are relatively conserved. On the basis of the sequences of the DNAs encoding the rodent CDRs, six pairs of oligonucleotide PCR primers are synthesized. Each pair of primers is designed to initiate the synthesis of the DNA for one of the six mouse CDRs—three from the L chain and three from the H chain. In addition, each primer includes an extra 12 nucleotides at its 5′ end, complementary to the flanking regions within the human framework DNA into which the DNA for the mouse CDRs is targeted. Overlap extension PCR (see chapter 3) is then used to replace, one at a time, the complete DNA sequence for each of the human CDRs with the amplified DNA for the mouse CDRs. Thus, it is necessary to carry out six cycles of overlap extension PCR, one cycle to replace each CDR. This procedure, in effect, "grafts" the mouse CDRs onto the human antibody framework. The humanized variable-region cDNAs are then cloned into expression vectors, which are then introduced into appropriate host cells, usually Chinese hamster ovary (CHO) cells, for the production of antibodies.

To date, more than 50 different monoclonal antibodies have been humanized, the time and expense notwithstanding. For example, on 1 November 2013, the FDA approved the use obinutuzumab (Gazyva or GA101) for the treatment of CD20-positive B-cell non-Hodgkin's lymphoma. Obinutuzumab is a fully humanized monoclonal antibody that, like rituximab (a previously approved monoclonal antibody), is directed against the CD20 protein found on B cells. Obinutuzumab is seen by Genentech, the manufacturer of rituximab, as the commercial successor to rituximab, a chimeric monoclonal antibody with the same target.

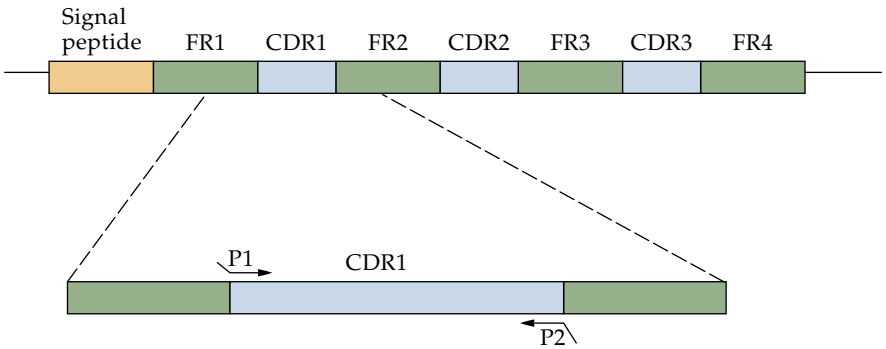


Figure 5.31 PCR amplification of CDR1 from a rodent monoclonal antibody L-chain cDNA. The PCR primers P1 and P2 contain oligonucleotides complementary to the rodent CDR1 DNA. In addition, P1 and P2 each contain 12 nucleotides at their 5′ ends that are complementary to the FRs of human monoclonal L-chain cDNAs. Using six separate pairs of oligonucleotide primers—three for the VL region and three for the VH region—each of the rodent CDRs is separately amplified by PCR. Then, by PCR, the amplified rodent CDRs are spliced into human antibody genes in place of the resident CDRs. This grafting is made possible by the presence of DNA complementary to the human FRs on the amplified rodent CDR DNAs.

Obinutuzumab targets and destroys CD20-positive B cells by directly binding to the tumor cell surface-exposed CD20 molecules and initiating antibody-dependent cytotoxicity (ADCC), where the antibody recruits the immune system to attack the CD20-positive B cells (Fig. 5.32). When obinutuzumab was compared directly with rituximab, patients treated with obinutuzumab had significantly longer periods of remission when they were disease free and did not require treatment. In addition, obinutuzumab has been glycol-engineered (i.e., its complement of surface carbohydrates has been modified) so that it does not contain any fucosylated sugar molecules. Fucose is a hexose deoxy sugar that is found on the surface of many eukaryotic cells (including mammals). It is thought that the lack of fucose moieties may result in an increase in the ADCC activity of the antibody.

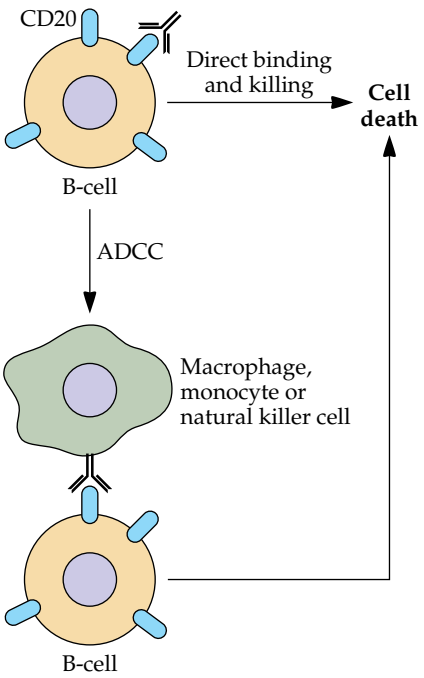


Figure 5.32 Schematic representation of the mode of action of the monoclonal antibody obinutuzumab, either directly targeting CD20-positive B cells or recruiting the immune system to attack and destroy CD20-positive B cells.

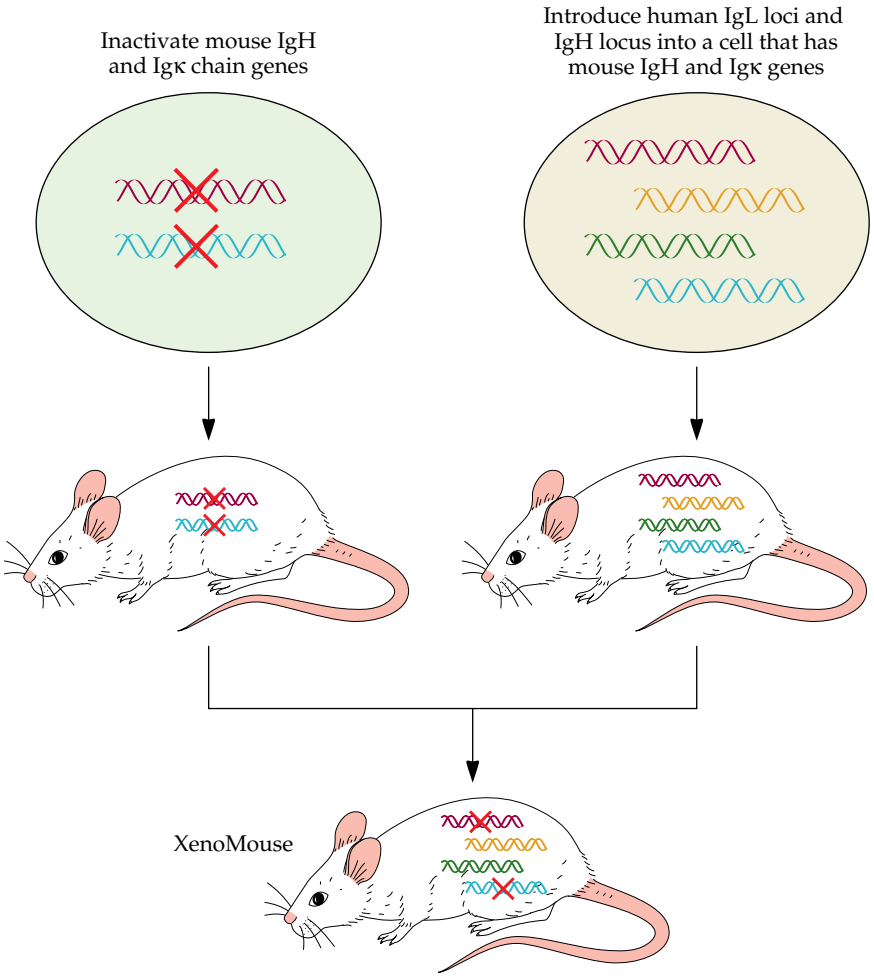
Human Monoclonal Antibodies

Although most of the immunotherapeutic agents that have been developed have been effective, there are drawbacks to the use of monoclonal antibodies that contain nonhuman sequences. For example, if multiple treatments are required, which is often the case, it is desirable that the antibody contain no, or only a very limited amount of, nonhuman sequences to prevent immunological cross-reactivity and sensitization of the patient. Unfortunately, it is very difficult to create human monoclonal antibodies, for a number of technical and ethical reasons. For example, the human chromosomes of fused human lymphocyte–mouse myeloma cells during hybridoma formation are unstable, and cells that produce a human monoclonal antibody are rarely formed. To date, no human myeloma cell line has been discovered that can replace the mouse myeloma cell line in this procedure. Even if it were possible to form human hybridoma cell lines, it is contrary to accepted norms of

medical research to inject humans with a specific antigen for nontherapeutic purposes and to perform a partial splenectomy (spleen removal) to collect antibody-producing cells. Therefore, it has been necessary to devise other approaches for obtaining human monoclonal antibodies.

To address this need, researchers constructed a "XenoMouse" in which (i) the mouse antibody production genes are deleted and (ii) all of the human immunoglobulin loci (encoding both L and H chains) are integrated into a mouse chromosome (Fig. 5.33). The human H-chain genes and the human L-chain genes (both κ and λ, which are different classes of L-chain genes)

Figure 5.33 Generation of a XenoMouse. Mouse antibody genes are inactivated by specific deletions in embryonic stem cells, which are subsequently used to generate transgenic mice unable to make antibodies. The human genes encoding immunoglobulin light and heavy chains are introduced on a yeast artificial chromosome (YAC) vector into mouse embryonic stem cells. These cells are used to generate transgenic mice able to synthesize both mouse and human antibodies. The mice generated from these two types of manipulation are crossbred, and mice that can synthesize only human immunoglobulins are selected, immunized, and used to make hybridomas producing human antibodies. Purple- and blue-colored DNA schematically represent mouse DNA, while yellow- and green-colored DNA represent human DNA. Inactivated genes are depicted by a red X on top of the mouse DNA structures.

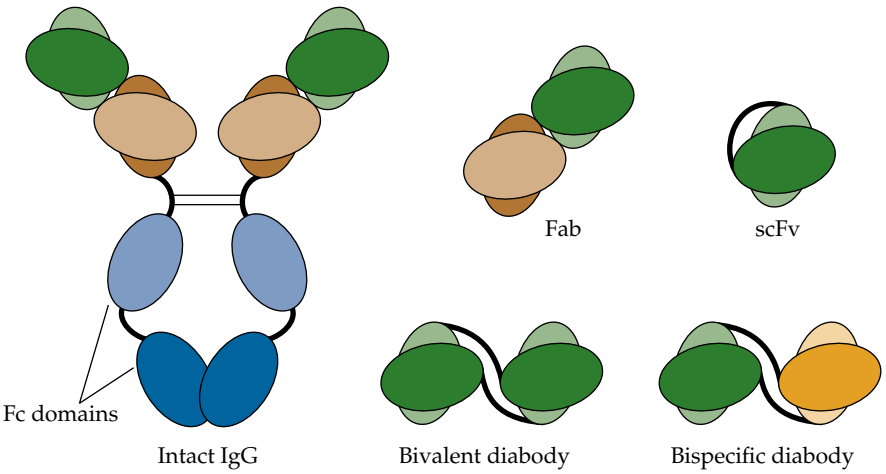


were cloned into a yeast artificial chromosome (YAC) that can carry very large amounts of foreign DNA (Fig. 3.29). The YAC with the human immunoglobulin genes were then introduced into mouse embryonic stem cells by fusing YAC-containing yeast spheroplasts (cells from which the cell wall has been removed) with the embryonic stem cells. This procedure yielded a large number of embryonic stem cells in which all of the introduced human immunoglobulin genes have become stably integrated into the chromosomal DNA. These transfected cells were used to generate mice containing human (as well as mouse) immunoglobulin gene loci. Crossbreeding of two mouse lines, one carrying both mouse and human immunoglobulin genes and the other carrying only the deleted mouse immunoglobulin genes, produced a mouse strain (called XenoMouse) that expresses only human immunoglobulins. It is now possible, after immunization of a XenoMouse with a particular antigen, to produce a fully human immunoglobulin, many of which have been approved for human use (see Table 5.5), with many more at various stages of clinical development.

Antibody Fragments

Naturally occurring antibodies are highly specific targeting proteins that provide animals with a powerful means of defending themselves against a wide range of pathogenic organisms and toxins. Immunoglobulin G (IgG) is the main antibody found in mammalian serum, and it is the native form that is almost exclusively used in therapeutic antibodies (Fig. 5.34). The fact that IgG molecules have two identical sites (Fv regions) that bind to two identical antigens (i.e., they are bivalent) generally increases their effectiveness *in vivo*. While the Fc portion of the IgG molecule is important in recruiting cytotoxic effector functions through complement (an amplifying cascade of proteins that facilitates pathogen destruction) or interaction with specific receptors, Fc-mediated effects are not necessary for all applications and may even sometimes be undesirable. By manipulating portions of the IgG L- and H-chain cDNAs, researchers have constructed a variety of IgG derivatives or fragments that may be used instead of whole antibody

Figure 5.34 Schematic representation of several different active antibodies and antibody fragments.

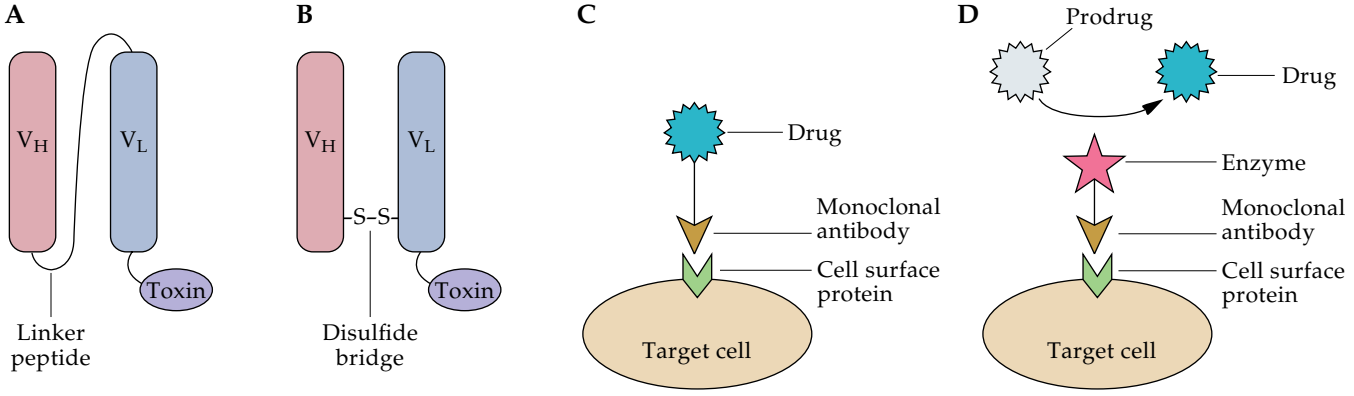


molecules (Fig. 5.34). Some of these molecules, because of their small size, bind more efficiently to targets that are inaccessible to conventional whole antibodies. Others have multiple sites for binding to the same antigen, while others still have binding specificities for two or more target antigens.

Antigen-binding single-protein chains (scFv), consisting of only V_L and V_H domains, may be used for a variety of therapeutic and diagnostic applications in which the Fc effector functions are not required and when small size is an advantage. Single-chain antibodies have a molecular mass of approximately 27 kilodaltons (kDa), compared with approximately 150 kDa for IgG molecules. Because of their small size, single-chain antibodies can penetrate and distribute in large tumors more readily than intact antibodies. In addition, a protein-coding sequence can be linked to a single-chain antibody sequence to create a dual-function molecule that can both bind to a specific target and deliver a toxin or some other specific activity to a cell (Fig. 5.35).

Computer simulations of the three-dimensional structure of a potential single-chain antibody showed that the V_L and V_H domains have to be separated by a linker peptide to assume the correct conformation for antigen binding (Fig. 5.35A). Thus, DNA constructs of V_L and V_H sequences from a cDNA of a cloned monoclonal antibody were each ligated to a chemically synthesized DNA linker fragment in the order V_L–linker–V_H. After expression in *E. coli*, the single-chain protein was purified, and both its affinity and specificity were found to be equivalent to those of the original intact monoclonal antibody. Moreover, instead of linking the V_H and V_L chains with a short peptide, amino acids in the FR can be replaced with cysteine residues that will form a disulfide linkage between the two peptides (Fig. 5.35B). The effectiveness of this disulfide-stabilized Fv molecule (V_L-S-S-V_H) coupled to a cancer cell toxin was compared with that of an scFv molecule coupled to the same toxin. The disulfide-stabilized and scFv immunotoxins had the same activity and specificity. However, the disulfide-stabilized immunotoxin was severalfold more stable than the scFv molecule. This result suggests

Figure 5.35 Schematic representation of the use of monoclonal antibody fragments to deliver other molecules to target cells. A toxin is coupled to an scFv molecule **(A)** and to a disulfide-stabilized Fv molecule **(B).** A monoclonal antibody-based drug delivery system. The drug is coupled directly to a monoclonal antibody or fragment **(C)**. An enzyme that converts an inactive prodrug to an active drug is attached to a monoclonal antibody or fragment. The active drug is formed only in the immediate vicinity of the target cells. **(D)** In all cases, the monoclonal antibody binds to a specific protein on the surface of the target cell.



that disulfide-stabilized Fv molecules may be more useful than scFv molecules in some therapeutic applications. On the other hand, single-peptide scFv fragments are somewhat easier to synthesize than are disulfide-stabilized Fv molecules that require the synthesis of two separate peptides and their subsequent precise joining. The two types of molecules have been used in different ways. For example, by altering the number of amino acids in the linker in an scFv molecule (usually to five or fewer amino acids), it is possible to direct the self-assembly of these molecules into either bivalent dimers, called diabodies (Fig. 5.34), trimers (triabodies), or tetramers (tetrabodies). Shortening the linker affects not only the multimerization, but also the stability of the molecule, with molecules with a shorter linker tending to be more stable. In addition, it is possible to combine two different antigen specificities into a single bispecific diabody (Fig. 5.34).

Drugs that are effective when tested in cell culture are often much less potent in a whole organism. This difference is typically due to the inability of the drug to reach the targeted site in the whole animal at a concentration sufficient to be effective. Increasing the dose of a drug is usually not the answer to this problem, because high drug concentrations often cause problematic side effects. One strategy to enhance the delivery of a drug to its target site includes coupling the drug to a monoclonal antibody fragment that is specific for proteins found only on the surfaces of certain cells, e.g., tumor cells (Fig. 5.35C). Alternatively, a specific enzyme that converts an inert prodrug (precursor) to the active form of the drug may be coupled to a monoclonal antibody directed against a specific cell surface antigen (Fig. 5.35D). To ensure that the drug is released only in the vicinity of the target cells, the monoclonal antibody or single-chain antibody that is complexed with the prodrug-converting enzyme must bind to a protein that is highly specific to the target cell and be stable under physiological conditions but cleared rapidly from circulation. With this approach, only specifically targeted cells are exposed to the drug, permitting the use of a much lower concentration than if it was administered directly.

It is sometimes advantageous to use an antibody fragment (particularly one that is conjugated to a toxin or radiochemical) to target a tumor or pathogen cell, since the Fc portion of the full-length antibody may impede or prevent the rest of the molecule from binding to relatively inaccessible antigens. However, despite the usefulness of antibody fragments in a variety of applications, their major limitation as therapeutic agents is that, since they lack the Fc portion of the molecule, they are unable to mediate a complete immune response.

One example of coupling a toxin to an antibody fragment is the fusion of a portion of *Pseudomonas* exotoxin A to an scFv. *Pseudomonas* exotoxin A is a 66-kDa protein with three separate domains; domain I is responsible for cell binding, domain II for translocation of the protein into the cell, and domain III for ADP-ribosylation, where the latter activity is responsible for the toxicity of the molecule (Fig. 5.36A). An immunotoxin is generally synthesized by replacing the N-terminal target-binding domain of the toxin, e.g., *Pseudomonas* exotoxin A domain I, with a single-chain antibody sequence, thereby creating a molecule very similar in size to the original toxin with the ability to bind, enter, and kill a specific cell (Fig. 5.36B). A number of immunotoxins that have antitumor activity *in vitro* and in animal models have been constructed. These include antibodies directed against some cancer cell surface proteins such as the p55 subunit of the interleukin-2 receptor, the transferrin receptor,

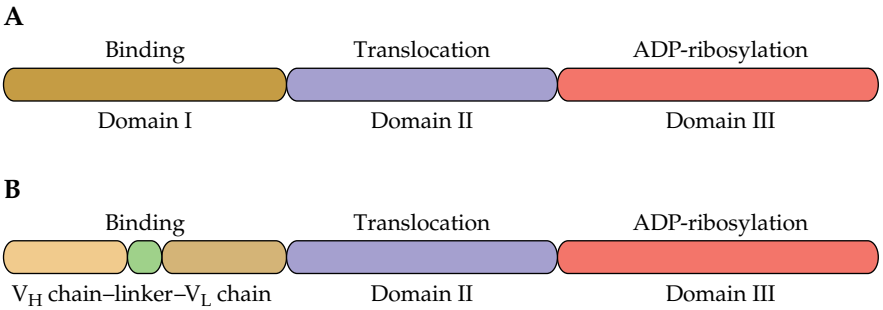
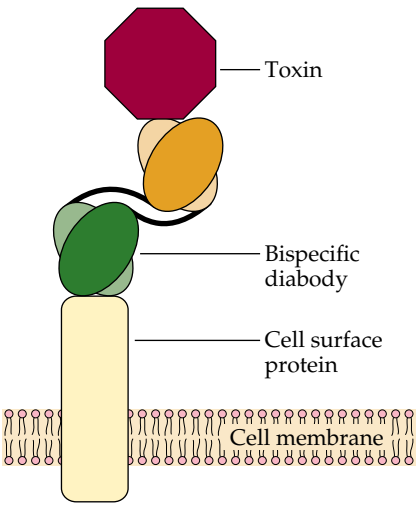


Figure 5.36 Domain structure of *Pseudomonas* exotoxin A **(A)** and a single-chain antibody–*Pseudomonas* exotoxin A **(B)**. The functions of the various domains are shown.

Figure 5.37 Schematic representation of the binding of a diabody to a protein molecule on the surface of a cancerous cell, as well as the binding of a toxin protein molecule to the other portion of the diabody.



carbohydrate antigens, and the epidermal growth factor receptor. It is also possible to direct toxin molecules to cancer cells by using a bispecific diabody that is engineered to bind to a surface-specific tumor-associated antigen and then to a toxin molecule, thereby directing the toxin molecule to the tumor (Fig. 5.37). Several different engineered immunotoxins are currently in clinical trials.

Sometimes, it is therapeutically advantageous to couple a whole antibody molecule to a drug/toxin. At the present time, most of the whole antibody–drug conjugates that are in use are designed to enter into lysosomes (where lysosomes are membrane-bound organelles that contain hydrolytic enzymes that act to break down various biomolecules) following their internalization into target cells. Importantly, prior to the experiment described below, human epidermal growth factor receptor 2 (HER2) monoclonal antibodies were already in clinical use (as pertuzumab) for the treatment of HER2 breast cancers. To improve upon the outcomes with this monoclonal antibody, a monoclonal antibody-drug conjugate directed at HER2-positive breast cancer cells, i.e., cells that contain the protein human epidermal growth factor receptor 2 on the tumor cells' surface, was modified by engineering the monoclonal antibody to reduce its affinity for its target at acidic endosomal pH relative to near neutral pH (Fig. 5.38). The modified antibody retained its specificity for HER2 but has an altered isoelectric point. This small change in the monoclonal antibody resulted in the subsequently formed antibody-drug conjugate being able to more readily dissociate from its target, thereby facilitating entry of the target into the lysosome. It is thought that this type of modification should eventually increase the therapeutic efficiency of HER2–drug conjugates.

It is possible to create peptides that are smaller than scFvs and still retain the ability to bind to a specific antigen. The rationale for developing smaller antibody–toxin complexes is that they are more likely to penetrate a tumor and thereby more completely stop tumor growth. It is well established that antibody-binding specificity resides within the six hypervariable loops called CDRs, three from the variable region of the heavy chain and three from the variable region of the light chain (Fig. 5.39). In all antibody molecules, the CDRs are flanked by FRs. Hence, it was speculated, at least for some antibodies, that the major portion of the antigen-binding site might reside primarily within two CDRs, one from the heavy chain and the other from the light chain, and not require significant contributions toward antigen binding from the four other CDRs. To test this possibility, starting

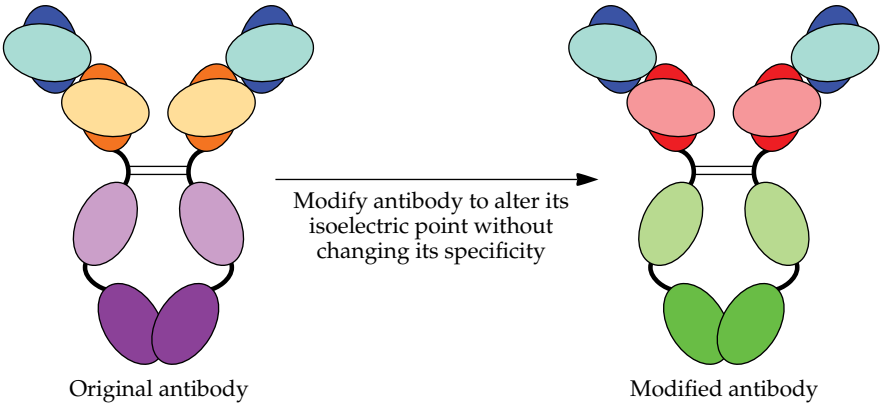
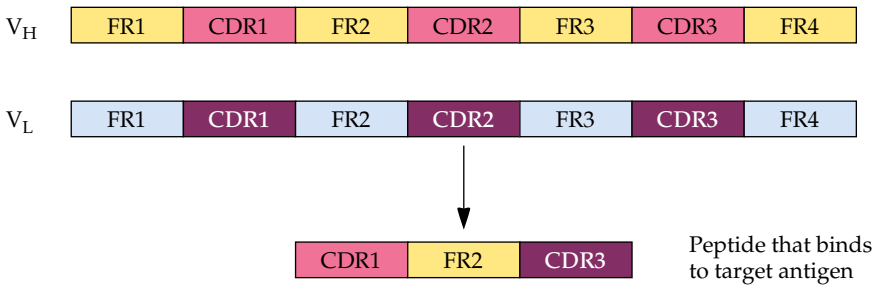


Figure 5.38 Schematic representation of the modification of a monoclonal antibody directed against HER2 that retains the antibody specificity but alters the protein's isoelectric point.

with genes for the variable portion of a monoclonal antibody against a surface protein from Epstein-Barr virus, eight different peptide combinations were synthesized. Each peptide contained at least one CDR3 loop (known to be the major antigen-contacting segment), as well as one other CDR loop and an FR spacer (which acts as a linker peptide). All eight of these peptides were tested *in vitro* for the ability to compete with the parental full-length antibody for binding to the surface protein of the Epstein-Barr virus (thought to be the causative agent of Burkitt lymphoma and other cancers). One of the peptide combinations, V_HCDR1–V_HFR2–V_LCDR3 (Fig. 5.39), appeared to be promising. The short peptide (28 amino acids long, or ~3 kDa) apparently retained the binding specificity of the whole monoclonal antibody (~150 kDa) from which it was derived.

Next, this peptide was coupled to a toxin molecule, colicin Ia, and the combination was tested in cultured cells and in mice. Colicin Ia has a molecular mass of approximately 69 kDa. When it was injected into mice, the peptide–colicin adduct efficiently traveled through the circulation and then found and killed the tumor cells expressing the target antigen. Colicins may kill tumor cells in several different ways, including depolarization of the cytoplasmic membrane, or through DNase or RNase activity. Importantly, without the antibody fragment, colicin Ia by itself does not affect these tumors to any significant extent. Also, the original monoclonal

Figure 5.39 Organization of V_H and V_L regions of a monoclonal antibody and the development of a peptide, based on a portion of the CDR and FR of the V_H and V_L regions of the antibody molecule, with a similar binding specificity.



antibody is unable to penetrate into the tumor. On the other hand, the peptide–colicin adduct accumulated at the cores of the targeted tumors. This exciting work is at an early stage of development, so a large number of technical issues remain to be addressed before it can become an effective human therapeutic measure. Nevertheless, the demonstration that a small peptide can mimic the binding specificity of an entire antibody molecule and successfully deliver a cellular toxin to targeted cells may provide the basis for a whole new approach for treating tumors.

Combinatorial Libraries of Antibody Fragments

Hybridoma cells (producing monoclonal antibodies), like most other animal cells in culture, grow relatively slowly, do not attain high cell densities, and require complex and expensive growth media. As a consequence, the cost of monoclonal antibody production is quite high and is an impediment to their more widespread use as therapeutic agents. In addition, various monoclonal antibodies that are directed against the same target protein may be different from one another in that they bind to different epitopes on the target protein and with different affinities. To circumvent these problems, attempts have been made to genetically engineer bacteria, plants, and animals to act as "bioreactors" for the production of monoclonal antibodies and antibody fragments.

An elaborate series of manipulations makes it possible to select, as well as produce, functional fragment antibodies in *E. coli*. Following cDNA synthesis from mRNA isolated from mouse antibody-producing cells (B lymphocytes), the variable portions of the H- and L-chain sequences in a cDNA preparation are amplified separately by PCR. The amplified cDNA sequences of the H and L chains are each cloned, separately, into a bacteriophage λ vector that can carry relatively large fragments of DNA (Fig. 5.40A and B). Subsequently, the cDNAs of one H and one L chain are cloned into a single "combinatorial" vector, thereby enabling the bacteriophage-infected *E. coli* to coexpress both chains, thus forming an assembled antibody Fv fragment (Fig. 5.40C). The H and L chains are expressed in *E. coli* during the lytic cycle of bacteriophage λ, and the library of combinatorial bacteriophage clones can be screened for the presence of specific antigen-binding activity.

The step in which L and H chain cDNAs are combined onto one vector creates a vast array of diverse antibody genes, some of which encode unique target-binding sites whose isolation would never have been possible by standard hybridoma procedures. The mammalian antibody repertoire has the potential to produce approximately 10^8 to 10^{10} different antibodies (from ~10^4 L chains and the same number of H chains). A phage library contains approximately this number of clones, so one combinatorial library can be expected to produce approximately as many different antibodies (in this case, Fv molecules) as any mammal. In addition, once an initial combinatorial library has been constructed, it is possible to shuffle the cDNAs encoding the L and H chains to obtain Fv molecules that recognize unusual epitopes. Furthermore, even greater variation may be achieved by random mutagenesis of the DNAs that are contained in the combinatorial library. Moreover, because millions of bacteriophage plaques can be screened in a relatively short period, the identification of Fv molecules with the desired specificity takes only about 7 to 14 days.

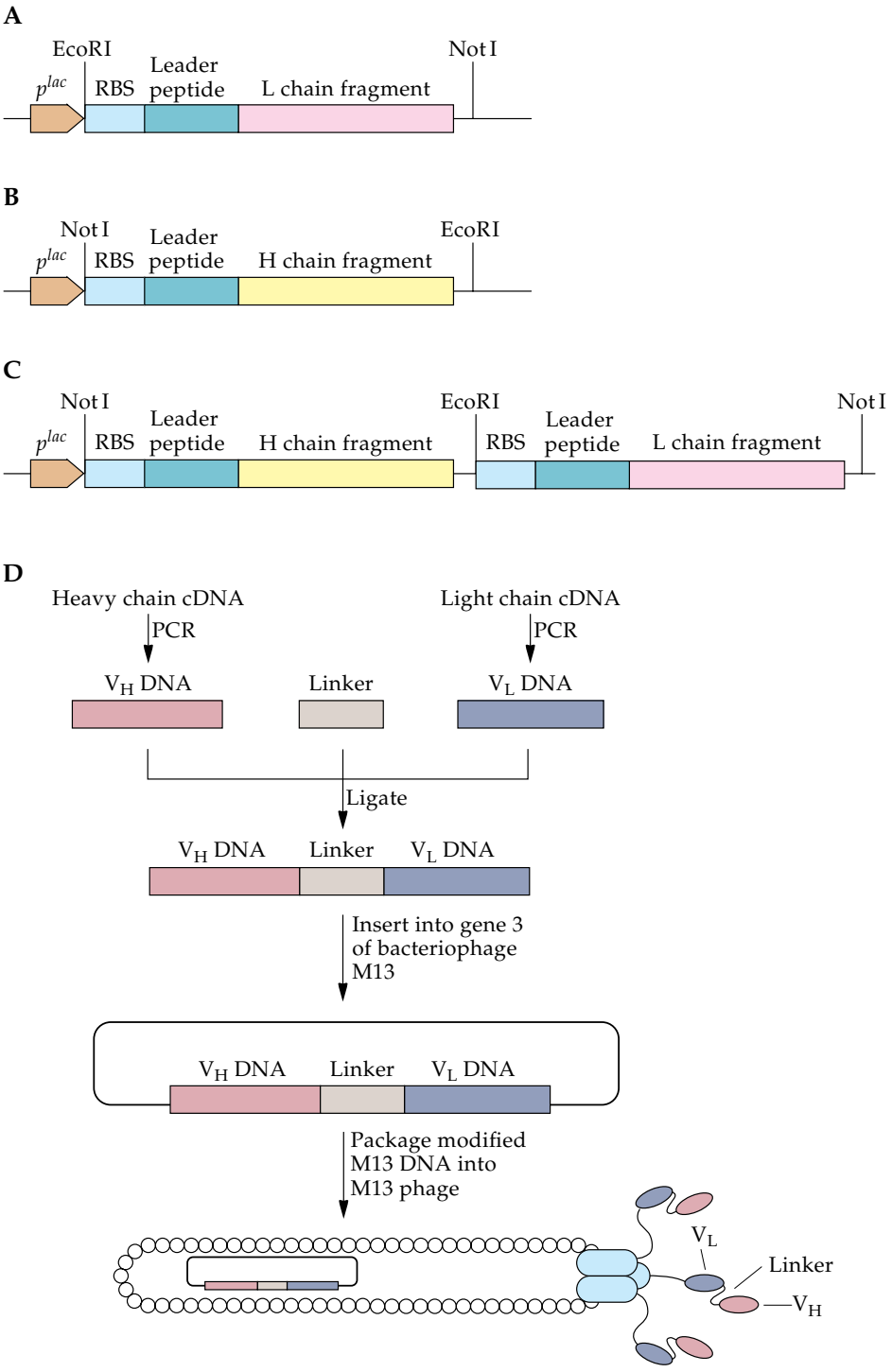


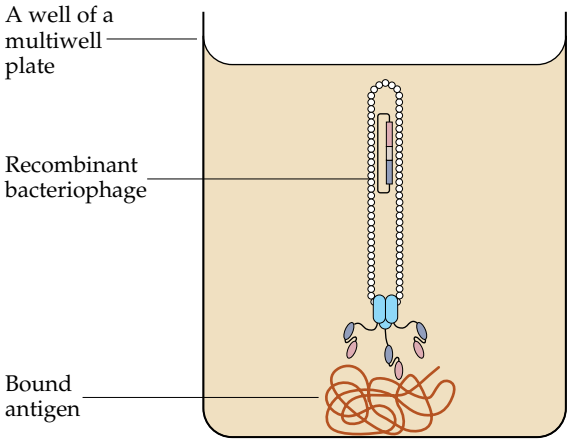
Figure 5.40 DNA constructs of an Fv combinatorial gene library cloned into bacteriophage λ DNA. **(A and B)** Portions of the cDNAs of the L **(A)** and H **(B)** chains are separately cloned into bacteriophage λ vectors. **(C)** The sequences encoding the fragments from the H-chain library (following EcoRI digestion) are ligated to the sequences encoding the fragments from the L-chain library, thereby creating a combinatorial library that contains all possible combinations of L- and H-chain fragments. **(D)** Formation of an Fv antibody combinatorial library in the filamentous bacteriophage M13. Following extraction of mRNA and its conversion to cDNA, the cDNAs for the VL and VH regions are amplified (separately) by PCR and then ligated to DNA that encodes a short linker peptide. Each single-chain antibody–DNA construct is cloned into gene 3 of bacteriophage M13. Three molecules of the phage gene 3 protein are produced on the surface of each M13 bacteriophage; these are expressed after the modified bacteriophage is used to infect *E. coli*. p^{lac}, *E. coli lac* promoter; RBS, ribosome-binding site.

Because they lyse bacterial host cells, bacteriophage λ vectors are not particularly useful for the production of large quantities of protein. To overcome this drawback, the bacteriophage λ vector was engineered so that the H- and L-chain DNA sequences were inserted into a site that was flanked by plasmid DNA sequences. This plasmid DNA, containing an H- and L-chain DNA combination, can be excised from the bacteriophage λ vector and transformed into *E. coli*. As part of a plasmid, large numbers of Fv fragments can be produced in *E. coli* cells.

As an alternative to the use of bacteriophage λ, filamentous bacteriophages, such as M13 and fd, have been used for the production of combinatorial libraries (Fig. 5.40D). Thus, following infection of *E. coli* by the filamentous bacteriophage combinatorial library, the antibody fragment is synthesized as part of a fusion protein on the surface of the bacteriophage. Combinatorial libraries of antibody fragments on the surface of a filamentous bacteriophage can be screened by an enzyme-linked immunosorbent assay-like system (Fig. 5.41). Briefly, samples (aliquots) of the library are added to the wells of a multiwell plate that are coated with the target antigen, and the bacteriophage particles that produce a specific Fv fragment bind to the target antigen. The wells are rinsed thoroughly to remove any unbound bacteriophage. Next, an antibody that binds to the bacteriophage coat protein, and is covalently conjugated with an enzyme, is added to each well. The wells are rinsed to remove any unbound antibody–enzyme complex. When the bacteriophage expresses an antibody fragment that binds to the adsorbed (to the microtiter plate well) target antigen, the presence of this complex is readily detected by assessing the enzymatic activity of the enzyme attached to the antibody bound to the bacteriophage coat protein to convert a colorless substrate into a colored compound.

This approach is easier than using plaque assays with bacteriophage λ to select and subsequently purify a bacteriophage producing an antibody fragment that binds to a specific antigen. Using either bacteriophage λ or M13, once a desired antibody fragment-producing bacteriophage has been identified, the DNA can be isolated and subcloned into an expression vector. These procedures may be used to produce mouse, chimeric, or humanized antibodies.

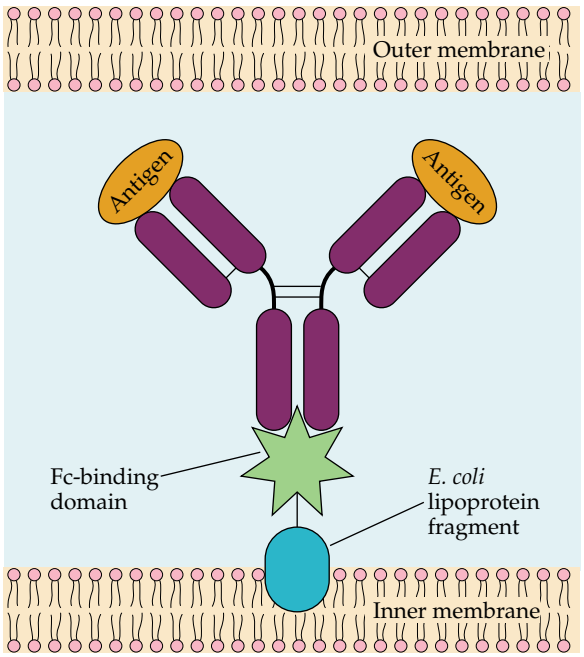
Figure 5.41 Immunological screening of a bacteriophage M13 combinatorial library.



A Combinatorial Library of Full-Length Antibodies

All of the early combinatorial libraries of antibodies included either single-chain antibodies or Fab fragments and not full-length antibodies. However, for many applications it is advantageous for therapeutic antibodies to be full length. With this in mind, researchers have also cloned and expressed complete antibody molecules in *E. coli*. To do this, a combinatorial library of full-length antibodies was generated and engineered so that the antibodies were secreted into the periplasm between the inner and outer membranes. In addition, prior to the expression of the antibody molecules, the host *E. coli* was engineered to express a fusion protein that became anchored within the inner membrane (lipoprotein fragment) and also contained a portion of a protein from the *Staphylococcus aureus* protein A that binds specifically to Fc antibody regions (Fig. 5.42). When an IgG molecule is secreted into the periplasm, the fusion protein binds to the Fc region. The tightly bound IgG–fusion protein complex remains intact when the *E. coli* cells are treated with EDTA and lysozyme to remove a portion of the outer membrane. Then, a fluorescently labeled target antigen is added to detect those cells that have expressed an antibody directed against that target antigen. By cell sorting, fluorescently labeled cells are selected. Then, the plasmid DNA that encodes the selected IgG molecules is isolated and expressed in an *E. coli* strain that does not synthesize the membrane-anchored fusion protein. This simple, yet powerful, technique facilitates the selection and production in *E. coli* of full-length monoclonal antibodies.

Figure 5.42 Schematic representation of the selection of a full-length antibody that binds to a fluorescently labeled target antigen. The Fc portion of the antibody is bound by a fragment of *S. aureus* protein A fused to a portion of an *E. coli* lipoprotein anchored in the inner membrane.



Shuffling CDR Sequences

Very large libraries of single-chain antibodies, maintained in *E. coli*, can be the sources of a wide range of highly specific human antibodies so that one does not have to resort to using mice for the initial production of hybridomas. Theoretically, such libraries may contain a greater diversity of antibodies than is normally found in the human immune system. To construct a library of this sort, B cells and other antibody-producing cells, from different, nonimmunized individuals and from different tissues and organs, are collected and pooled before the mRNA is isolated (Fig. 5.43). The isolated mRNA is used to synthesize cDNA, which is a template for the specific PCR amplification of each encoded CDR (separately). The amplified V_L and V_H CDR DNA sequences are mixed with oligonucleotides encoding FR regions and a linker. Overlap extension PCR is used to join and amplify the sequences for CDRs separated by FRs. In addition to being entirely human, the 2×10^9 different single-chain antibodies that can be produced in this way have a very wide range of specificities, reflecting the fact that the CDR sequences have been incorporated in random order and from a variety of sources; i.e., they have been shuffled. The dissociation constants (*Kd*s) for the interaction of the selected antibodies with their target antigens range from 0.9×10^{-9} to 420×10^{-9} M^{-1} (where tight binding is represented by a low number). For antibodies, *Kd*s of around 0.1×10^{-8} to 10×10^{-8} M^{-1} reflect a very high degree of specificity for the target antigen. The fact that a very large library of single-chain antibodies, all with a high affinity for their target antigens, could easily be generated by this procedure means that the goal of being able to select virtually any antibody from a nonimmunized library has been achieved.

Dual-Variable-Domain Antibodies

Despite the usefulness of antibody fragments in a variety of applications, a major limitation of using them as therapeutic agents is that, since they lack the Fc portion of the molecule, they are unable to mount a complete immune response. To increase the utility of naturally existing full-length antibodies, as well as to ensure that they are effective initiators of a complete immune response, researchers have created what they have termed "dual-variable-domain immunoglobulins" (Fig. 5.44). These constructs are essentially IgG molecules containing two tandem Fv regions, each with a different specificity. Dual-variable-domain immunoglobulins are bispecific and tetravalent, and they consist entirely of human immunoglobulin sequences. That is, each molecule contains four Fv regions, two identical Fv regions directed against one antigen and two Fv regions directed against another antigen. For example, a dual-variable-domain immunoglobulin specific for both interleukin-12 and interleukin-18 produced in CHO cells showed binding to both of these cytokines, binding each with an affinity similar to that of the original monospecific antibody. In addition, in a biological assay, using an immunodeficient mouse model, the interleukin-12/interleukin-18 antibody was as effective at inhibiting induced IFN-γ production as was a combination of the original two monospecific antibodies. Using this strategy, it should be possible to generate full-size bispecific antibodies for a variety of therapeutic purposes.

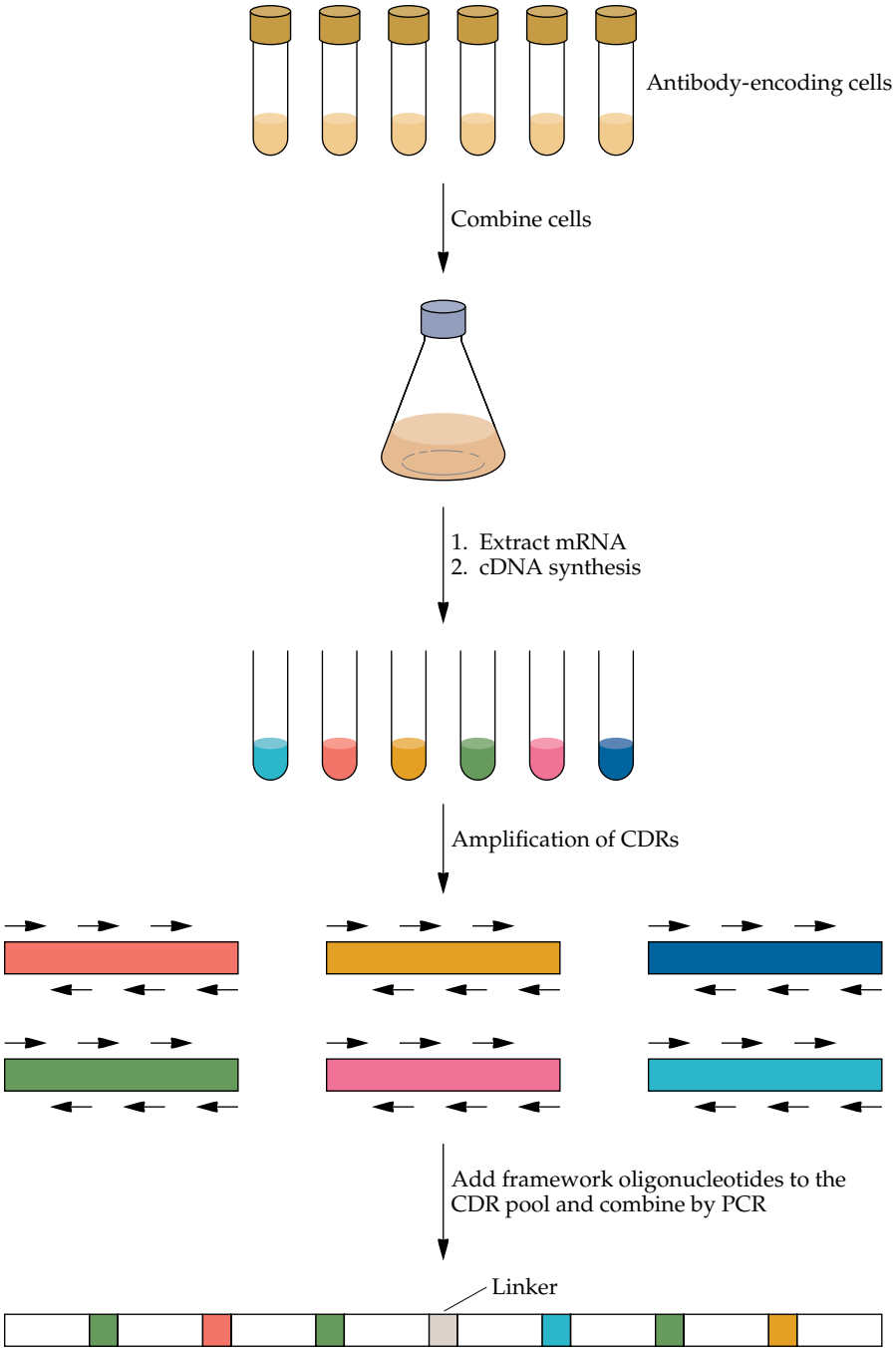


Figure 5.43 Construction of a large library of single-chain antibodies containing multiple CDRs. B cells from several nonimmunized individuals were collected and pooled, and the mRNA was isolated and used to synthesize cDNA. Oligonucleotide primers containing DNA sequences that included small portions of the FR sequence were added to amplify in separate reactions the six CDRs in the L and H chains in the cDNAs. The amplified CDRs from all six PCRs were mixed together with oligonucleotides encoding the FRs and the linker, and genes encoding random combinations of CDRs separated by FRs were synthesized by overlap extension PCR. One of the many possible single-chain antibodies that were synthesized is shown.

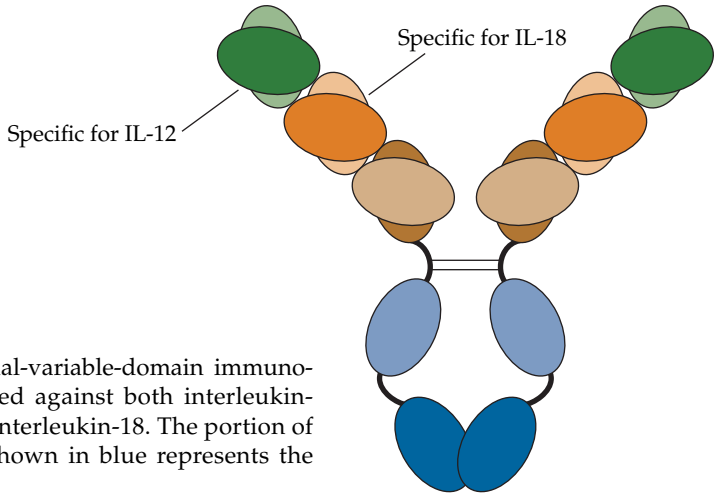
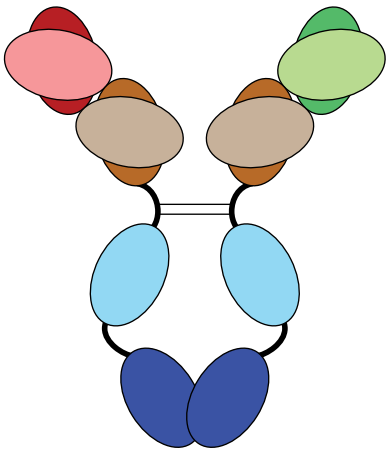


Figure 5.44 Dual-variable-domain immunoglobulin directed against both interleukin-12 (IL-12) and interleukin-18. The portion of the molecule shown in blue represents the Fc region.

Figure 5.45 A bispecific antibody that binds to both human blood coagulation factor IXa (Fv region shown in red) and human blood coagulation factor X (Fv region shown in green). The portion of the molecule shown in blue represents the Fc region.



Bispecific Antibodies against Hemophilia

A simpler variant of the above-mentioned dual-variable-domain antibodies is bispecific antibodies. In this case, the two Fv regions have different specificities, while the remainder of the molecule is identical in each of the two halves (Fig. 5.45). As an example, one Fv region binds to blood coagulation protein factor IXa (the activated version of factor IX), while the other Fv region binds to blood coagulation factor X, facilitating the interaction of the two blood coagulation factors, which activates factor X. This bispecific antibody may be used to treat patients with hemophilia A who are unable to activate factor X because they are either missing blood coagulation factor VIII or have a gene that produces a defective factor VIII protein. Factor VIII is normally responsible for bringing together factors IXa and X in order to activate factor X as part of the intrinsic human blood clotting cascade (Fig. 5.46). Individuals with 1 to 5% of normal factor VIII activity levels are said to have severe hemophilia and bleed excessively after what is otherwise a slight injury, while those with factor VIII levels from 5 to 40% of normal are considered to have mild hemophilia (and generally only bleed excessively following surgery). Currently, bleeding episodes in hemophilia patients are often treated by administering either activated prothrombin, thereby bypassing the blockage in the clotting cascade, or recombinant factor VIII. However, neither of these treatments is entirely satisfactory and without problems. Therefore, scientists developed the bispecific monoclonal antibody, emicizumab, which effectively replaces factor VIII. Following several years of clinical trials, emicizumab was approved for clinical use in both children (as young as 1 year old) and adults in 2018. The results to date have been quite encouraging.

Anti-HIV Antibodies

Acquired immune deficiency syndrome (AIDS) is a disease that is caused by the human immunodeficiency virus (HIV). As the disease progresses, it interferes with the body's immune system so that an infected individual typically finds it difficult to fight a wide variety of opportunistic infections, some of which are rare in individuals with normal

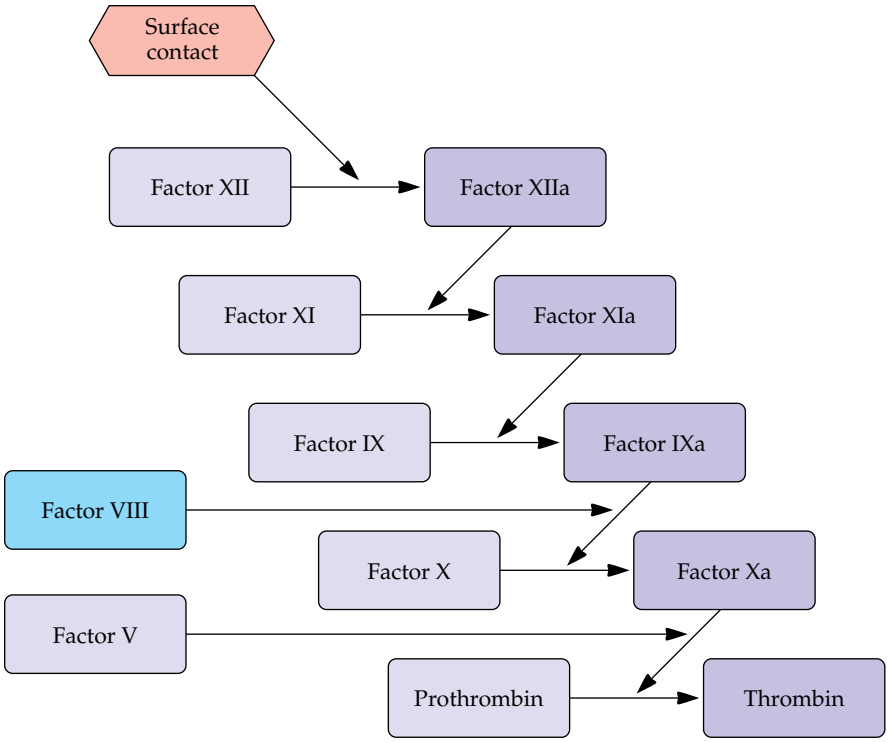


Figure 5.46 A portion of the intrinsic human blood coagulation cascade. The boxes represent protein factors that are involved in this cascade. The dark purple boxes indicate factors that have been activated. Factor VIII, which normally brings factors IXa and X together for the activation of factor X, is expressed at a much lower level than normal in individuals with hemophilia A.

immune system functioning. In 2018 it was estimated that ~38 million people worldwide were infected with HIV, with more than three-quarters of a million people dying as a direct consequence of this infection during that year. The retrovirus (specifically two species of *Lentivirus*) that causes this disease is generally spread through the exchange of bodily fluids during unprotected sex, from contaminated blood transfusions or hypodermic syringes, and between mother and child during pregnancy, delivery, or breastfeeding. At the present time, there is neither a cure nor a vaccine for HIV infection. However, it is possible to slow down the progression of the disease through the ongoing administration of numerous retroviral compounds. In the absence of antiviral treatment, the average survival time from the initial infection is ~11 years. On the other hand, with antiviral compound treatment, many individuals may have a near-normal life expectancy. Also, with treatment with antiviral compounds, there is ~70% decreased risk of infected individuals acquiring tuberculosis. Importantly, even with unprotected sex, the possibility of passing on HIV through sex is extremely low in situations where the HIV-positive partner has a consistently undetectable viral load. In the United States, individuals with AIDS and receiving antiviral therapy spent ~$22.5 billion for these drugs in 2018. However, it should be noted that individuals in low-income countries pay much less for antiviral compounds than individuals in high-income countries.

The first human monoclonal antibodies directed against HIV were developed in the 1990s. However, because HIV has a high mutation rate, many different strains of the virus exist, and although some of the monoclonal antibodies bound quite strongly to their targets, each was effective against only a limited number of HIV strains. More recently (~2013), two CD4-binding antibodies were developed and are currently (2020) in phase II clinical trials. CD4 cells are white blood cells that play a key role in responding to human infections and are targeted by HIV. The CD4 receptor on the surface of white blood cells is bound by the envelope glycoprotein gp120 on the surface of the HIV envelope, which is essential for attachment of the virus. The two anti-CD4 monoclonal antibodies prevent HIV from binding to the CD4 receptor and have a much broader specificity than earlier antibodies; they neutralize 90 to 95% of known HIV strains. In addition, about a dozen other broadly neutralizing monoclonal antibodies that recognize most of the available surfaces of the HIV envelope glycoprotein have been selected and tested. Unfortunately, none of these has shown both the potency and the breadth required to be an effective therapeutic agent, especially given the high HIV mutation rate.

In an attempt to overcome the limitations of using a single monoclonal antibody directed against HIV, researchers have developed bi- and trispecific antibody constructs that have greater breadth against HIV strains than the broadly neutralizing antibodies from which they were derived. Unfortunately, most of these constructs are not sufficiently potent to be clinically useful. Researchers reasoned that they could construct a more efficient antibody by creating a molecule that targeted two different HIV envelope epitopes simultaneously (Fig. 5.47). To start, they isolated and characterized two human monoclonal antibodies that, based on computer simulations, bound to nearby but distinct regions of the HIV envelope protein. One of these monoclonal antibodies was engineered to be a single-chain variable region attached to an Fc region (i.e., scFv-Fc), and this molecule was then further modified to form, together with a complete monoclonal antibody, a bispecific antibody. Importantly, the bispecific antibody neutralized more HIV isolates and was more potent than either of the antibodies from which it was derived. This activity was attributed to the ability of the bispecific antibody to simultaneously bind to two separate sites on the envelope protein. However, whether this construct eventually becomes a widespread effective therapeutic agent for the treatment of individuals infected with HIV remains to be determined.

Anticancer Antibodies

A number of therapeutic antibodies are directed against antigenic proteins that are overexpressed on the surfaces of cancer cells compared to noncancerous cells. Sometimes, merely binding antibodies to the cell surface can engage the immune response and thereby impair the functioning of the cancer cells. For example, women with a type of breast cancer that overproduces the HER2 protein on the surface of the tumor may be effectively treated with the humanized anti-HER2 monoclonal antibody trastuzumab (Herceptin) (Box 5.2). Another example of a highly successful chimeric anticancer antibody is rituximab (Rituxan) (Box 5.3), which is used to treat patients with non-Hodgkin's lymphoma. Unfortunately, with this approach, (i) antibodies directed against these proteins may also

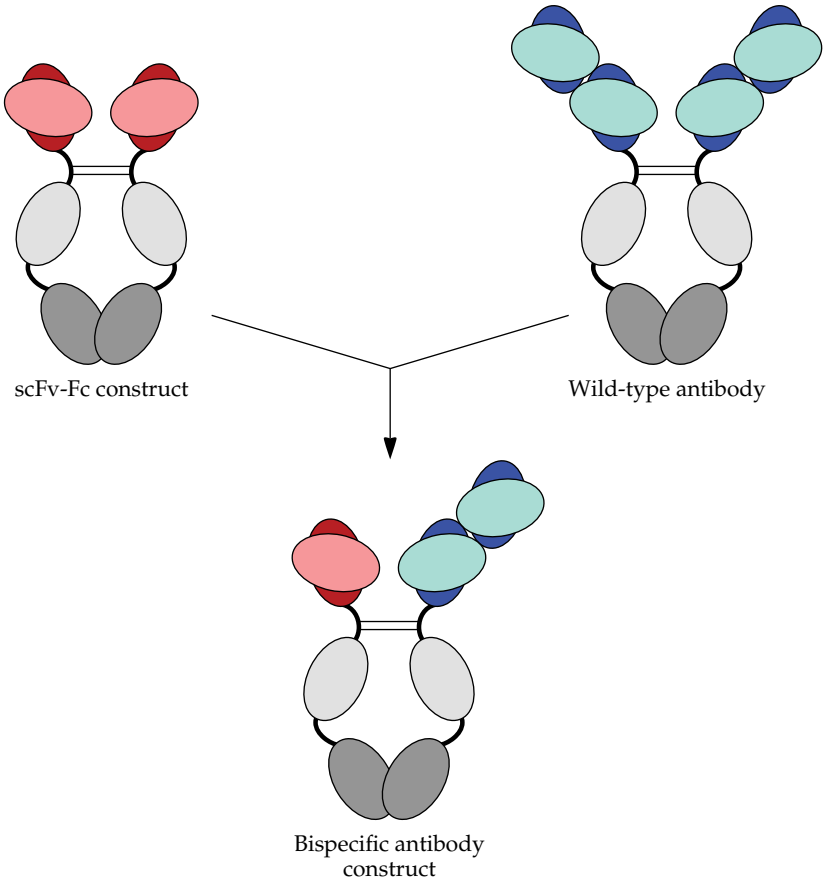


Figure 5.47 A schematic representation of the construction of a bispecific antibody directed against two separate epitopes that are both found on the surface of the HIV envelope protein. The Fc regions are shown in gray. Following the selection and characterization of two human monoclonal antibodies directed against separate regions of the envelope protein, one of these antibodies was engineered to form an scFv-Fc construct, which was then combined with the other monoclonal antibody to form the bispecific antibody construct shown.

bind to some noncancerous cells expressing the same or a similar antigen, and (ii) overexpressed surface proteins offer only a limited number of targets for therapeutic antibodies. One way to select for additional cell surface targets would be to identify proteins whose expression is selectively induced in tumor cells exposed to chemotherapeutic drugs. For example, when colorectal cancer cells were treated with the drug irinotecan, which is a topoisomerase enzyme inhibitor and is commonly used to treat this type of cancer, several newly synthesized proteins were found on the surfaces of those cells. (Irinotecan itself is a major chemotherapeutic agent, approved for human use in the United States in 1996 with sales of >$2 billion in 2018. Topoisomerases are enzymes that unwind DNA during either DNA replication or mRNA transcription.) The new cell surface proteins were expressed early, after exposure to the chemotherapeutic compound and prior to any major effects on the viability of treated cells. Monoclonal antibodies directed against one newly synthesized cell surface protein (called LY6D/E48) were generated, and then the antibodies were conjugated with

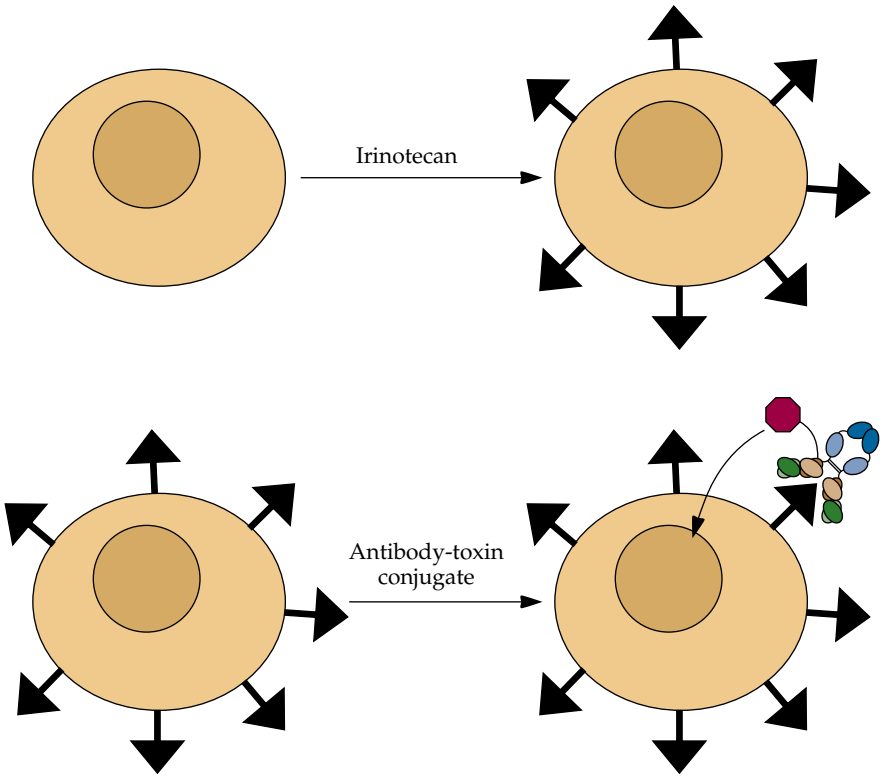


Figure 5.48 Targeting tumor cells for destruction by monoclonal antibody–toxin conjugates. Tumor cells are first treated with the chemotherapeutic agent irinotecan, which induces the synthesis of a unique cell surface protein. Then, a monoclonal antibody directed against the newly synthesized cell surface protein and conjugated to a toxin molecule is added. After binding of the antibody to the cell surface, the toxin is internalized, thereby killing the tumor cell.

the cellular toxin auristatin E. The antibody–toxin conjugate was then used to treat tumor cells that were first treated with irinotecan (Fig. 5.48). Following binding to the cell surface protein, the antibody–toxin conjugate was internalized inside the tumor cell. With this strategy, in six out of eight mice, tumors disappeared entirely, while in the other two mice, the tumors were dramatically decreased in size. This approach is currently being tested in clinical trials. In addition, provided that it is possible to identify one or more proteins that are specifically induced by chemotherapeutic agents and are not found on the surfaces of nontumor cells, this procedure could become a strategy that is used to treat a variety of different types of human cancer.

One recent promising approach to treating a variety of different types of human tumors is the "blockade of immune checkpoints." The term "immune checkpoint" refers to components of the immune system that are essential for maintaining self-tolerance. That is, immune checkpoints prevent the immune system from attacking an individual's own tissues and cells. Immune checkpoints also modulate the duration and amplitude of physiological immune responses so that any damage to peripheral tissues is minimized. Normally, immune T cells are able to identify and subsequently attack tumor cells based on the presence of specific proteins on the

surface of the tumor cells (Fig. 5.49A). However, sometimes tumor cells can get around the immune system by synthesizing surface proteins that bind to specific checkpoint receptor proteins on the surface of the immune system T cells (Fig. 5.49B). In this way, the tumor cells trick the immune system into switching off the attack against the tumor cells. As a result, the tumor's growth and development become resistant to inhibition by the immune system. This is particularly important for T cells that are specific for tumor antigens. However, because many of the immune checkpoints are initiated by ligand–receptor interactions, the tumor can be prevented from thwarting T cell function by the addition of specific antibodies directed toward the checkpoint receptor on the T cell (Fig. 5.49C). These added foreign antibodies bind directly to the checkpoint receptor on the T cell and prevent the tumor cells from binding to the checkpoint receptor. In this case, the T cells are no longer prevented from attacking the tumor. This novel therapeutic approach is still in a relatively early stage of development. However, the use of cytotoxic T-lymphocyte-associated antigen 4 (CTLA4)-specific monoclonal antibodies has been approved by the U.S. FDA as a component of such a strategy. The protein CTLA4 is normally expressed on the surface of helper T cells, sending an inhibitory signal to T cells and thereby acting as a checkpoint receptor. In addition, a monoclonal antibody that is directed against the negative immunoregulatory human cell surface receptor PD-1 (programmed death-1) so that it binds to and blocks the activation of PD-1 was first approved by the U.S. FDA in December 2014; there are now three different monoclonal antibodies directed against PD-1 (pembrolizumab, nivolumab, and cemiplimab). Binding of the antibody to PD-1 activates T cells and cell-mediated immune responses against tumor cells or pathogens; it is used in the treatment of advanced melanoma and non-small cell lung cancers. The success rate of using the above-mentioned monoclonal antibodies is around 15 to 20% of patients who receive this treatment. However, there is now some evidence that the combination of monoclonal antibodies with direct-acting, immunostimulatory oligonucleotides can enhance the responsiveness to checkpoint inhibitors. In addition, nanoparticles previously designed to deliver small interfering RNA (siRNA) further increase the efficacy of using immunostimulatory oligonucleotides together with checkpoint inhibitors.

Some types of immunotherapy to fight various forms of cancers use specific monoclonal antibodies and antibody fragments to target certain proteins that help cancer cells to grow. By binding to these cancer cell surface proteins, the antibodies may inhibit the growth of cancer cells or possibly even kill the cancer cells. For example, although most ovarian cancer patients generally initially respond well to treatment with the chemical agents cisplatin and paclitaxel, the vast majority of these patients typically relapse after a few years. This relapse is thought to be the consequence of the survival of a small number of chemotherapy-resistant ovarian cancer stem cells that subsequently proliferate, repopulate the tumor, and spread to distal sites. Therefore, to more effectively treat ovarian cancer, it would be important to inhibit the proliferation of ovarian cancer stem cells. To do this, researchers first identified a number of proteins that are found uniquely on the surface of these ovarian cancer stem cells. They then systematically inhibited the expression of one or more of these proteins to identify those that might be therapeutic targets for efforts to impair ovarian cancer stem cell proliferation. One target is a type I orphan receptor

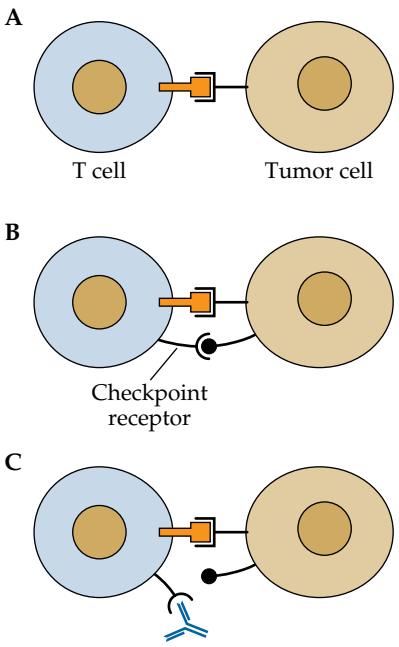


Figure 5.49 Schematic representation of cancer immunotherapy. **(A)** The T cells identify and attack the tumor cells based on the presence of cell surface proteins carried by the tumor cells. **(B)** The tumor cells bind to the checkpoint receptors on the surface of the T cells and trick the T cells into switching off the attack against the tumor cells. **(C)** Exogenous monoclonal antibodies that bind to the T cell checkpoint receptor protein are added to prevent the tumor cells from binding to the checkpoint receptor; in this way, the T cells retain the ability to attack and destroy the tumor cells.

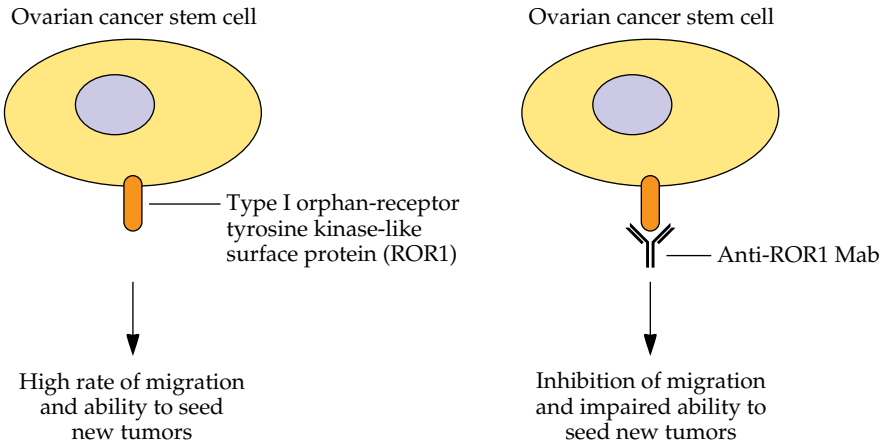


Figure 5.50 Schematic representation of ovarian cancer stem cells that produce the protein receptor tyrosine kinase-like orphan receptor 1 (ROR1) on their surface and the subsequent inhibition of the ability of those cells to migrate and cause the proliferation of the ovarian cancer stem cells by the addition of a monoclonal antibody (Mab) directed against the ROR1 protein.

tyrosine kinase-like surface protein (ROR1) (Fig. 5.50). Monoclonal antibodies that are directed against this ovarian cancer stem cell surface protein inhibit the cell's ability to migrate and to seed new tumors, suggesting that these monoclonal antibodies might provide an effective means of treating ovarian cancer patients. However, it remains to be demonstrated that this approach is as effective in patients with ovarian cancer as it is in immune-deficient mice.

A strategy similar to the one described above for ovarian cancer was developed and shown to be effective, in mice, for treating breast cancer (Fig. 5.51). That is, most of the breast cancer cells were killed following chemotherapy with paclitaxel; however, some cancer cells remained, and some of these became cancer stem cells. The cancer stem cells were responsible for the regrowth of the tumor and the metastasis (spreading to secondary sites) of the cancer. Because the cancer stem cells expressed a high level of type I orphan-receptor tyrosine kinase-like surface protein (ROR1), humanized monoclonal antibodies directed against ROR1 were highly effective in killing the vast majority of the remaining cancer cells (i.e., the cancer stem cells). While it appears that the approach used to treat ovarian cancer and breast cancer in mice is quite effective and may also be effective in the treatment of other types of cancers, it is necessary to demonstrate that this approach is as effective in humans as in mice.

Notwithstanding the enormous benefit that modern medicine has derived from the therapeutic use of monoclonal antibodies, the use of these molecules still has a number of limitations. For example, receptor tyrosine kinases (high-affinity cell surface receptors for a large number of hormones, peptides, and cytokines) are typically activated by the binding of specific ligands (such as hepatocyte growth factor) that cause the receptors to dimerize. Activation of these receptor tyrosine kinases by dimerization regulates normal cell processes but may also contribute to malignancy. Thus, one way to prevent the progression of many types of cancer would be to prevent the kinase activation (Fig. 5.52A). Therapeutic antibodies

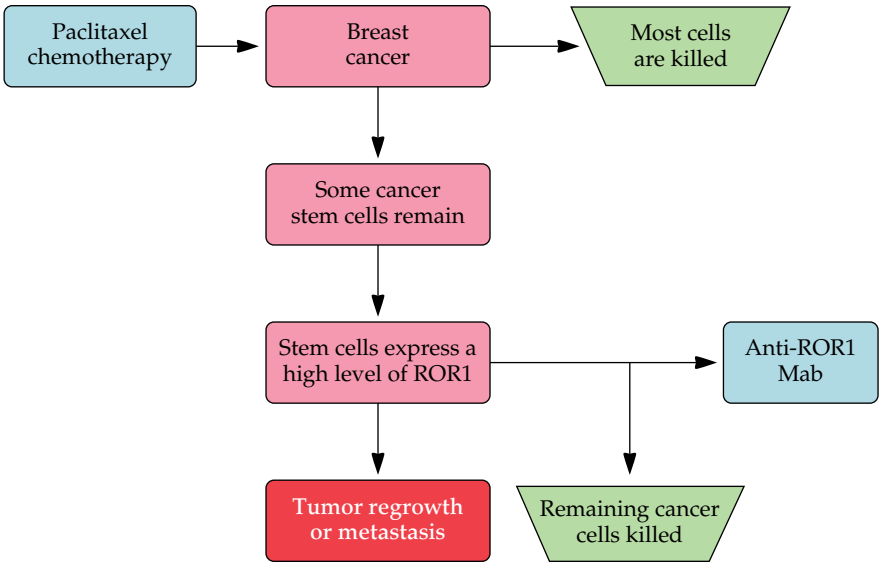


Figure 5.51 Schematic representation of the protocol used to treat breast cancer cells that are resistant to the chemotherapeutic agent paclitaxel. Some of the cancer cells that are not killed by paclitaxel become cancer stem cells that express a high level of the protein ROR1 on their surface. To prevent these cancer stem cells from facilitating tumor regrowth and metastasis, they may be treated with humanized monoclonal antibodies that are directed against ROR1.

may be used to block the interaction of the receptor tyrosine kinase with the activating ligand. However, sometimes a problem arises due to the bivalent nature of monoclonal antibodies; they may bind to two receptors, promoting their dimerization and thereby causing exactly the opposite of the intended effect. To overcome the problem of inadvertent dimerization caused by monoclonal antibodies against receptor tyrosine kinases, one group of scientists developed a monovalent version of a monoclonal antibody that has only one receptor-binding site and therefore cannot promote receptor dimerization (Fig. 5.52B). For a start, the scientists developed a mouse monoclonal antibody directed against the extracellular portion of the human hepatocyte growth factor receptor (also called MET) that has tyrosine kinase activity. The MET protein is essential for embryonic development and wound healing; however, its abnormal activation typically triggers events such as tumor growth, formation of new blood vessels that provide the tumor with nutrients, and the spread of the cancer to other organs. The expression of MET is activated in many types of human cancers, including cancers of the kidney, liver, stomach, breast, and brain. The selected mouse monoclonal antibody was then humanized before the antibody cDNA was cloned into *E. coli* and used to generate a monovalent monoclonal antibody molecule (Fig. 5.52B). The final, purified, monovalent monoclonal antibody molecule was tested extensively for its biological activity. The entire Fc portion of the original humanized monoclonal antibody molecule was included in an effort to increase the *in vivo* half-life of the final monovalent antibody; this is because antibody fragments, as opposed to complete antibody molecules, typically have short *in vivo* half-lives. To favor the construction of exclusively monovalent forms, mutations

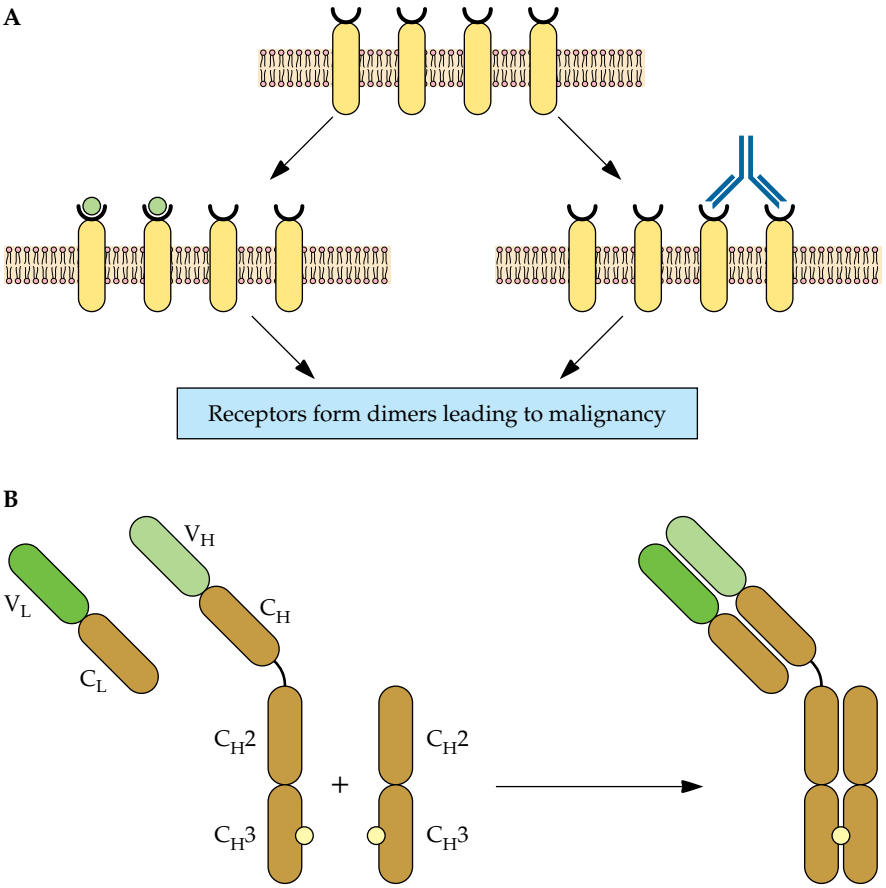


Figure 5.52 (A) Schematic representation of the binding of either specific ligands (such as hepatocyte growth factor) or monoclonal antibodies to a receptor tyrosine kinase, resulting in both cases in the dimerization of the receptors and the eventual development of malignancy. **(B)** A monoclonal antibody cDNA was expressed in *E. coli*, and portions of the DNA were used to generate a gene encoding a monovalent monoclonal antibody. Only the encoded proteins are shown in this figure. The yellow dots represent regions of the Fc portion of the molecule that were mutated to facilitate the specific binding of the two different C_H3 regions.

were created in the C_H3 Fc domains (shown as a yellow circle in Fig. 5.52B) to favor heterodimerization between the two distinct Fc domains. Based on the positive results in blocking receptor tyrosine kinase that were obtained with this approach, the researchers who were involved in this work envision that in the future there could be enormous therapeutic potential for different types of monovalent monoclonal antibodies in the treatment of various types of cancers.

One of the problems with the employment of monoclonal antibodies (whether or not they are conjugated to drugs) to treat many forms of cancer is that the monoclonal antibody may bind to antigens on normal healthy tissue, sometimes resulting in severe side effects. To avoid this problem, some researchers have derivatized monoclonal antibodies so that their ability to bind to their target is masked until they are within the tumor environment. More specifically, some leucine zipper coiled coils (the masking domain) were fused using a short linker to the N terminus of a

monoclonal antibody, conjugated or not to a drug (Fig. 5.53). Binding of the masked antibody to its target decreased by approximately 300- to 750-fold. However, when the masked antibody was exposed to specific tumor-associated proteases, the mask was cleaved from the antibody, and binding of the antibody to its target was restored. It is argued that this approach using the same mask may be used with a range of different antibodies, thereby improving the efficacy of the monoclonal antibody–drug conjugate. This approach remains to be tested on mice and then in clinical trials.

Several bispecific antibodies have been engineered so that they can simultaneously bind to cancer cells and to immune system T cells. This facilitates the killing of the cancer cells by the T cells, and these bispecific antibodies are often considerably more effective than conventional monoclonal antibodies (which must recruit immune system cells to kill cancer cells). With this in mind, one group of researchers recently developed a trispecific monoclonal antibody that can simultaneously bind to cancer cells and to immune system T cells and then activate the T cells so that they are potentially more lethal to the cancer cells (Fig. 5.54). This was done to target multiple myeloma cells (cancerous blood plasma cells), which accumulate in the bone marrow. The trispecific antibody was constructed by reconfiguring the antigen-recognition domains of two different antibody molecules and showing that the final construct was effective in a humanized mouse model. It is hoped that the increased efficacy of this engineered antibody will eventually become a more effective human therapeutic agent.

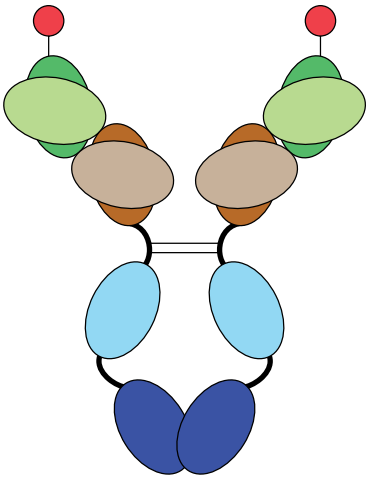
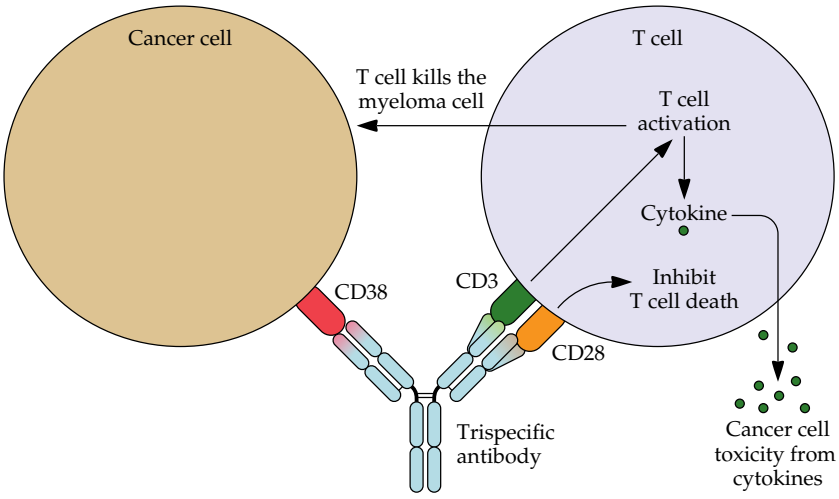


Figure 5.53 Schematic representation of a monoclonal antibody that has been modified so that its binding to its (tumor cell) target is masked (shown in red) and thereby prevented. In the vicinity of the target tumor cells, specific proteases remove the protein mask so that the antibody can now bind (with a much higher affinity) to its target cells. For simplicity of presentation, a drug conjugated to the monoclonal antibody is not shown.

Antibodies against Various Diseases

In comparison to the multiplicity of complex and often unintended effects in the treatment of various types of cancer with monoclonal antibodies, the effective treatment of a number of other diseases with monoclonal antibodies may be somewhat more straightforward. In order to

Figure 5.54 Schematic representation of a monoclonal antibody that has been engineered to simultaneously bind to a cancer (myeloma) cell (surface protein CD38), T cell surface protein CD3, which generates a signal that activates T cells, and T cell surface protein CD28 which increases both the binding of the antibody to T cells and the lifetime of the T cells.



limit the physical debilitation to individuals who have been infected with herpes simplex virus types 1 and 2 (HSV-1 and HSV-2), since the early 1990s HSV infections have been treated with chemotherapeutic agents such as acyclovir (Fig. 5.55A). However, in recent years, the resistance to these compounds has been increasing, especially in immunocompromised individuals. Resistance to acyclovir is typically the consequence of mutations in the HSV thymidine kinase gene, thereby precluding the use of other thymidine kinase-dependent nucleoside analogs. Fortunately, several monoclonal antibodies have recently been raised in mice against the HSV envelope (surface) glycoproteins gH, gD, and gB. In particular, it was shown that a gB-specific monoclonal antibody exhibited both strong HSV-neutralizing capacity as well as the ability to prevent viral cell-to-cell spread of HSV in animal models (Fig. 5.55B). For this antibody to be an effective human therapeutic agent, it was necessary to humanize the starting mouse monoclonal antibody. The humanized version of this monoclonal antibody is highly effective against nucleoside analog-resistant strains of HSV, both *in vitro* and *in vivo* in immunocompromised mice. Phase II clinical trials of humanized monoclonal antibodies directed against HSV-2 began in late 2019.

Rheumatoid arthritis is a common, chronic, inflammatory disorder that typically affects the small joints in the hands and feet of affected individuals. Rheumatoid arthritis affects the lining of the joints, causing painful joint swelling that can eventually result in bone erosion and joint deformity. In addition to causing joint problems, rheumatoid arthritis sometimes can affect other organs of the body, such as the skin, eyes, lungs, and blood

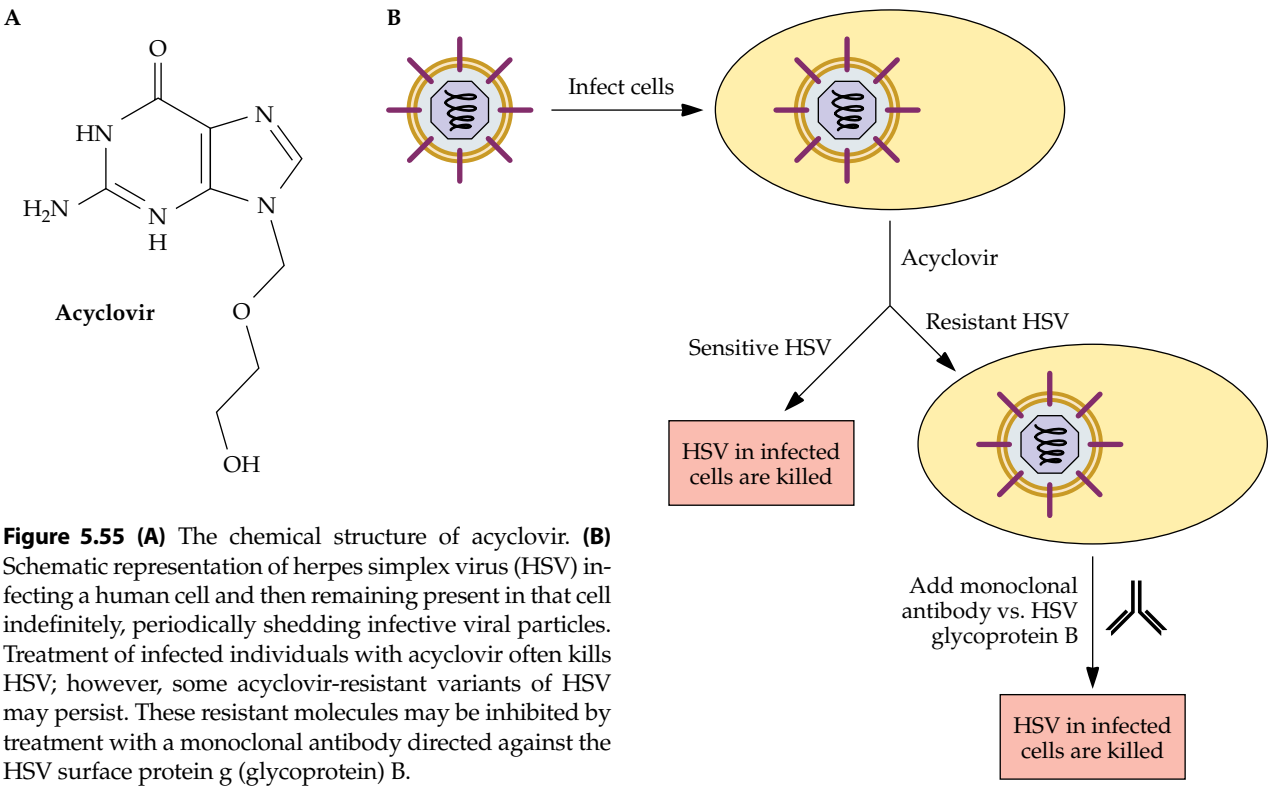


Figure 5.55 (A) The chemical structure of acyclovir. **(B)** Schematic representation of herpes simplex virus (HSV) infecting a human cell and then remaining present in that cell indefinitely, periodically shedding infective viral particles. Treatment of infected individuals with acyclovir often kills HSV; however, some acyclovir-resistant variants of HSV may persist. These resistant molecules may be inhibited by treatment with a monoclonal antibody directed against the HSV surface protein g (glycoprotein) B.

vessels. Moreover, there is no known cure for rheumatoid arthritis. Some cytokines play an important role in human inflammatory processes, and therefore, monoclonal antibodies that block the interaction of certain cytokines, such as tumor necrosis factor and interleukin-6, with their cellular receptors may provide a benefit to some rheumatoid arthritis patients. On the other hand, interleukin-4 has previously been shown (in animal models) to prevent some of the bone loss that often accompanies rheumatoid arthritis. A fusion product of interleukin-4 and a diabody (Fig. 5.56) that is specific for the tissues that are inflamed by the disease was found to localize at sites of arthritis (e.g., inflamed toes and paws) in the mouse. The antibody portion of the molecule was directed against the alternatively spliced extra domain A of the protein fibronectin (a high-molecular-weight glycoprotein that has a major role in cell adhesion), which is typically highly overexpressed at sites of arthritis. Treatment of mice with this fusion protein decreased the loss of body weight in arthritic mice, and it was the first time that researchers observed a regression of rheumatoid arthritis in mice. Following more extensive tests in animals, it may be possible to utilize the fusion protein as a therapeutic agent in human clinical trials.

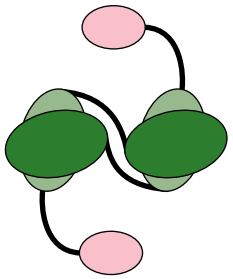


Figure 5.56 Schematic representation of a fusion protein between interleukin-4 (shown in pink) and a homodimeric single-chain Fv fragment (i.e., a diabody). The V_L chain is shown in light green, while the V_H chain is in dark green. In both cases the interleukin-4 is at the C-terminal end of the construct.

Anthrax is a lethal disease in animals and humans that is caused by *Bacillus anthracis*, a spore-forming, Gram-positive, rod-shaped bacterium. *B. anthracis* produces an exotoxin (anthrax toxin) made up of three proteins: (i) protective antigen (PA), (ii) edema factor (EF), so-called because it causes edema, an abnormal and excess accumulation of fluids in bodily tissues, and (iii) lethal factor (LF). These three proteins act together to impart their physiological effects. Assembled anthrax toxin complexes are taken up in an endosome (a cellular vesicle) of a target cell before they translocate into the cytoplasm. Once in the cell cytosol, the enzymatic components of the anthrax toxin disrupt cell signaling pathways that normally recruit immune cells, allowing the bacteria to evade the immune system, proliferate, and kill the host animal. The PA binds to cellular receptors and forms a pore through which it delivers LF and EF into the cytosol. LF is a Zn-dependent protease that cleaves mitogen-activated protein kinase kinases (resulting in lysis of macrophages, white blood cells that engulf and digest cellular debris). EF is a Ca–calmodulin-dependent adenylate cyclase (causing local inflammation and edema). PA together with LF results in the formation of lethal toxin, while PA with EF forms edema toxin. PA is both the major component in the current anthrax vaccine and the target antigen for most monoclonal antibodies that are used to protect individuals against anthrax, which has some researchers concerned about the long-term efficacy of this approach. Thus, it was decided to develop anti-EF monoclonal antibodies with the idea of using them either alone or in concert with anti-PA monoclonal antibodies in the treatment of anthrax infection.

Following the immunization of chimpanzees with purified EF protein, a phage combinatorial cDNA library of chimpanzee antibody genes was constructed (Fig. 5.57); cDNA is generated from expressed genes and includes transcripts encoding antibodies that respond to EF. The library was screened immunologically (see chapter 2), and four unique clones encoding Fab fragments specific for EF were selected. The Fab fragments were then converted to chimeric IgG molecules (Fig. 5.30A), with the remainder of the molecule being of human origin. One of the four selected antibodies

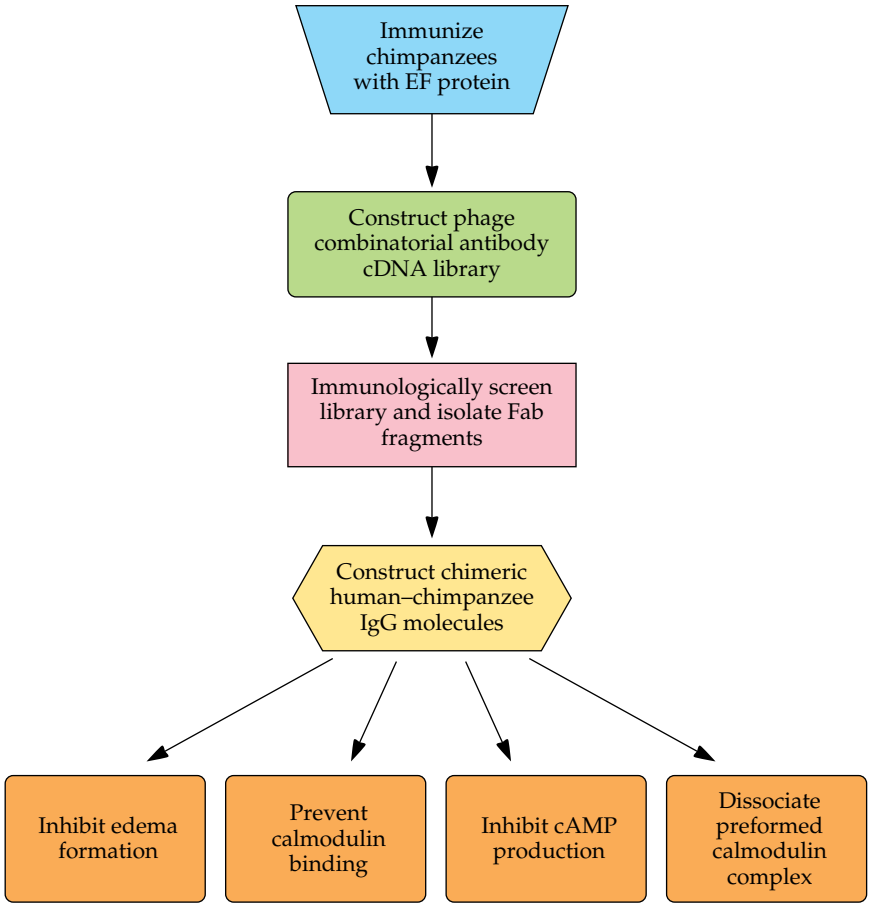


Figure 5.57 Flowchart of the development of a humanized monoclonal antibody against anthrax edema toxin.

inhibited EF-mediated cyclic AMP production (adenylate cyclase activity). The same antibody also inhibited edema formation in mice treated with the anthrax toxin and protected mice against edema toxin-mediated death. In addition, the selected antibody both prevented Ca–calmodulin from binding to EF and displaced Ca–calmodulin that was already bound to EF. The anti-EF monoclonal antibody that was selected and then modified may be used either alone or in tandem with anti-PA antibodies to broaden the spectrum of protection against anthrax.

Although most strains of *E. coli* are harmless to humans, there are, nevertheless, a large number of *E. coli* strains that can cause severe food poisoning. Unfortunately, many *E. coli* strains have also become resistant to most commonly used antibiotics, so that some *E. coli* infections can become life threatening. To remedy this situation, researchers have sought additional means (other than antibiotic treatment) to combat antibiotic-resistant strains of *E. coli*. Some *E. coli* integral outer membrane proteins, including those that are constituents of the β-barrel assembly machine, are an essential component in the assembly of the *E. coli* outer membrane itself (Fig. 5.58). The β-barrel proteins are protein β-sheets that commonly form pores in membranes; they are found in a variety of Gram-negative bacteria in addition to *E. coli*. Since the depletion of the β-barrel assembly proteins

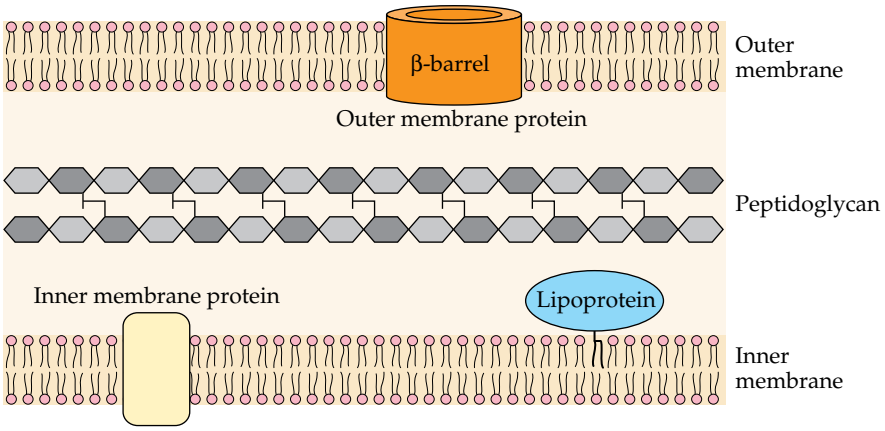


Figure 5.58 Schematic representation of a cross section of a portion of an *E. coli* cell wall and membrane.

is often detrimental to the *E. coli* cell's survival, researchers selected this protein as a target for monoclonal antibodies that disrupt bacterial cells. They isolated a monoclonal antibody that binds to the proteins that form the *E. coli* β-barrel assembly machine and ultimately results in killing of the bacterium. Notwithstanding the effectiveness of this antibody under laboratory conditions, whether this approach is an effective means of curing individuals infected with life-threatening antibiotic-resistant strains of *E. coli* remains to be determined.

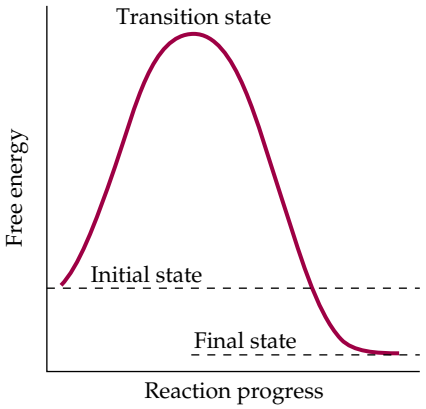
Antiobesity Antibodies

It has been estimated that as much as 15% of the world's population and more than 30% of the population of the United States are overweight or obese. In recent years, weight loss has become a major obsession, especially in many of the more affluent countries of the world, with billions of dollars spent annually on (mostly unsuccessful) attempts by individuals to lose weight. Notwithstanding the social components of carrying excess weight, overweight and obese individuals are significantly at risk for a number of potentially debilitating diseases. Consequently, there has been a considerable focus on finding a simple and painless way to deal with this epidemic.

Enzymes typically increase the rates of chemical reactions. They do this by stabilizing a high-energy transition state intermediate between the substrate (initial state) and the final product (final state), which lowers the activation energy (the difference between the energy in the initial state and the transition state) of the chemical reaction (Fig. 5.59). Chemical analogues of the transition state intermediate (i.e., transition state analogues) mimic the transition state intermediate but do not undergo any chemical reaction. As a result, transition state analogues can bind more tightly to an enzyme than either substrate or product analogues. Moreover, since transition state analogues do not undergo a chemical reaction, they act as enzyme inhibitors by blocking the active site.

In the mid-1980s, several investigators developed antibodies that had enzyme-like catalytic activity. They did this by first immunizing mice or rats with chemically synthesized small molecules that were transition state analogues of enzyme-catalyzed reactions. The transition state analogues

Figure 5.59 Idealized free energy diagram for an enzyme-catalyzed reaction. The substrate in the initial state must go through a high-energy transition state before it is converted into the product in the final state.

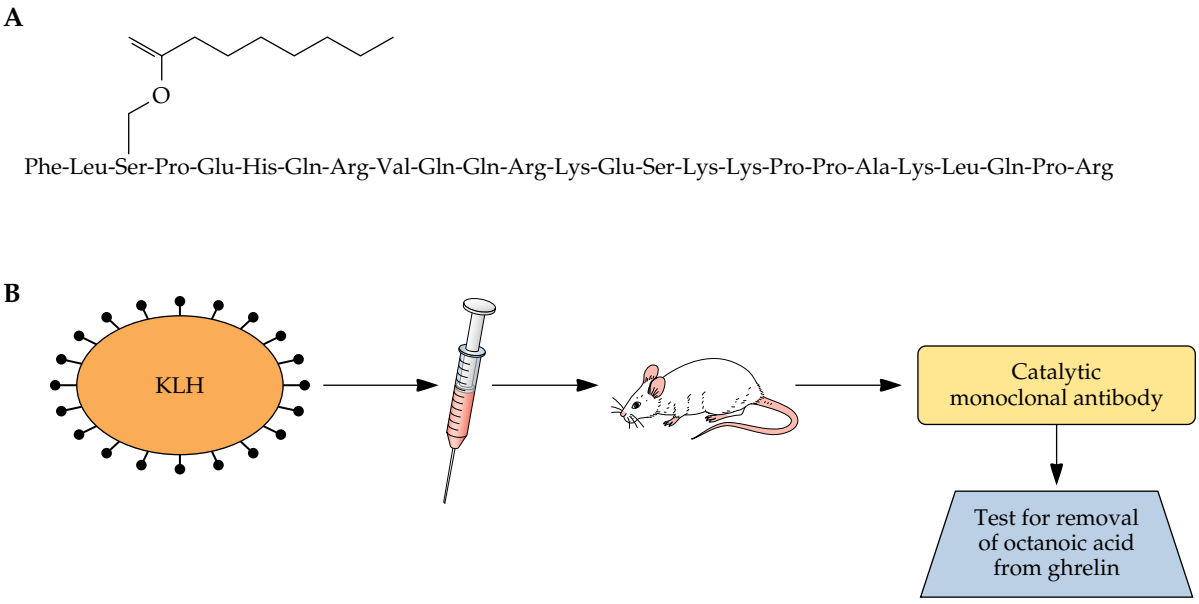


were chemically coupled to larger, generally inert, protein carrier molecules before the complex was used to immunize an animal to elicit antibody production. The transition state analogue acts as a hapten that binds but does not react with the catalytic antibodies (a hapten is a small molecule that binds to a specific antibody but generally cannot induce the formation of antibodies unless it is bound to a carrier protein). Rather, it is the enzyme substrates, which are structurally similar to the transition state analogues, that react with catalytic antibodies. After mice or rats are inoculated with a transition state analog to induce them to produce catalytic antibodies, the spleens of these animals are then excised and used to produce monoclonal antibodies with the desired catalytic activity. The genes for the monoclonal antibodies with the desired catalytic activity may then be isolated, manipulated, and overproduced in a variety of different host organisms.

Ghrelin is a small (28-amino-acid) natural peptide that stimulates food intake by regulating appetite. In humans, ghrelin levels increase when the stomach is empty, with the consequence that ghrelin facilitates weight regain and therefore impedes sustained weight loss. It is believed that lowering the level of this peptide could reduce food intake and promote permanent weight loss.

In its active form ghrelin is acylated on the hydroxyl side chain of serine 3 with *n*-octanoic acid (Fig. 5.60A). Catalytic antibodies directed against ghrelin were developed to remove the acyl group from, and therefore inactivate, the peptide (Fig. 5.60B). Several transition state analogues of ghrelin were chemically coupled to the protein keyhole limpet hemocyanin and used to inoculate mice. Keyhole limpet hemocyanin is an extremely large

Figure 5.60 Developing a catalytic antibody designed to facilitate weight loss. **(A)** Structure of the 28-amino-acid peptide ghrelin, including acylation of the serine residue at position 3 with *n*-octanoic acid. **(B)** Chemically synthesized transition state analogues of ghrelin are chemically coupled to keyhole limpet hemocyanin (KLH) and then injected into mice. Catalytic mouse monoclonal antibodies that can both bind and cleave ghrelin are selected and tested for the ability to remove the *n*-octanoic acid moiety. This cleavage renders the ghrelin inactive.



protein (its molecular mass is around 390,000 daltons) that is often used as an inert carrier protein in the production of antibodies. Its many surface lysine residues may be used to couple a large number of hapten molecules to its surface. The inoculated mice were the source of catalytic monoclonal antibodies that bind to and then specifically remove the *n*-octanoic acid (acyl) moiety from ghrelin (i.e., from the active form of the peptide molecule). Without the *n*-octanoic acid moiety, the ghrelin peptide no longer has biological activity and can no longer promote weight regain. Although this work is preliminary, it is envisioned that this catalytic antibody may effectively lower active ghrelin levels in humans and become an important adjunct to other weight loss strategies.

Enhanced Antibody Half-Life

It would be advantageous, for both patients and physicians, if the frequency of treatment with a particular therapeutic agent could be decreased as much as possible. This not only would save both patients' and physicians' time but also might decrease both the side effects and the costs of various treatments. To this end, scientists have worked to develop a variety of strategies to increase the stability of therapeutic proteins, including pegylation (Fig. 5.2) and the use of XTEN fusion proteins (Fig. 5.6). In addition, it is possible to extend the *in vivo* half-life of therapeutic antibodies. One way to achieve this goal is to engineer the Fc portion of a therapeutic antibody to bind more tightly to specific human Fc receptors. Bevacizumab (Avastin) is a humanized antivascular endothelial growth factor (anti-VEGF) IgG antibody that has been approved by the U.S. FDA for the treatment of colorectal, lung, breast, and renal cancers. Computer simulation was used to characterize the interaction of the Fc region of this antibody with its receptor, and specific variants were constructed by directed mutagenesis of the portion of the antibody DNA encoding the Fc region (Fig. 5.61). These variants were first tested *in vitro* for increased binding to the human Fc receptor. Subsequently, the constructed variants were tested in monkeys; not only did they bind more tightly to the receptor than the native form of the antibody, but also, one had a serum half-life of 31.1 days, compared to 9.7 days for the native form of the antibody. This preliminary work has yet to be tested in humans; however, if it is successful, it may be possible, for some antibody therapies, to decrease the frequency at which a patient needs to be treated, thereby providing greater convenience and lower costs to patients. Notwithstanding the early stages of this work, these researchers have already been granted several U.S. patents protecting this intellectual property.

Affibody Molecules

In an attempt to bypass the often tedious and time-consuming steps to isolate new monoclonal antibodies, several groups of researchers have constructed libraries of short peptides that can bind tightly to various target molecules. These peptides, known as affibodies, are selected for the ability to bind to selected target molecules. Currently, more than 400 literature reports of such high-affinity affibodies have been published; several of these affibody molecules and their affinity for their target molecules are summarized in Table 5.6. Importantly, the majority of these affibodies have an affinity for their targets in the nanomolar range, similar to that

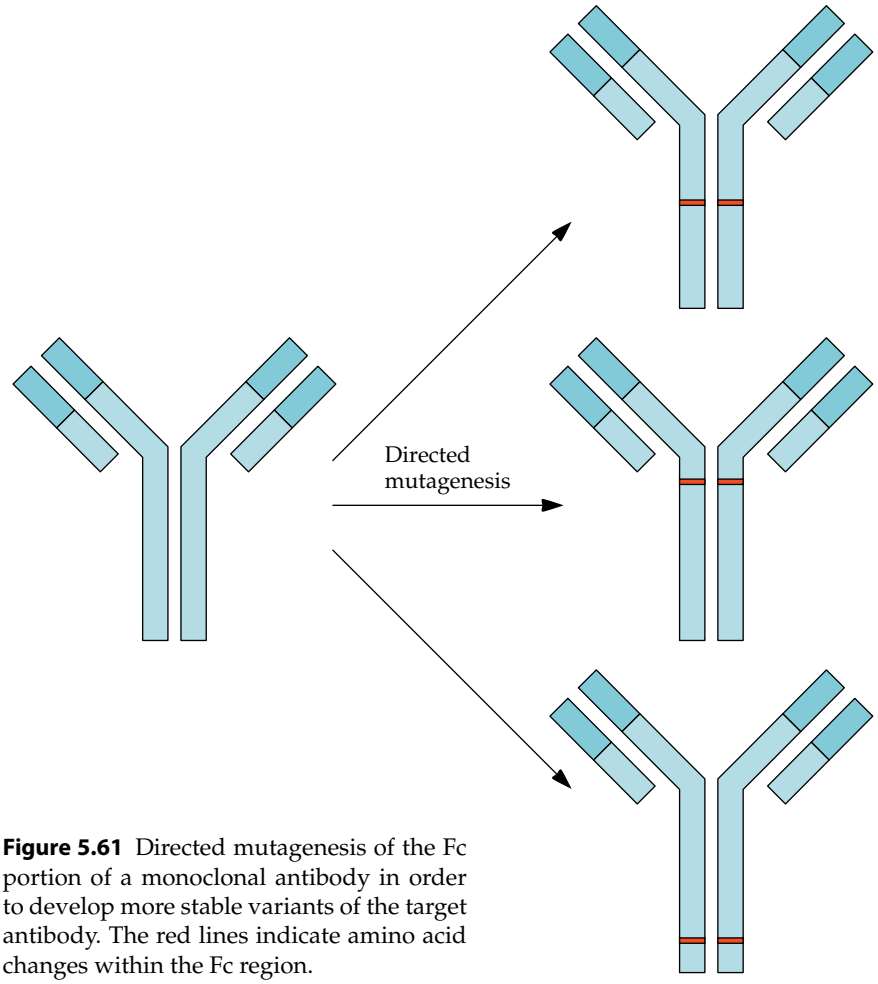


Figure 5.61 Directed mutagenesis of the Fc portion of a monoclonal antibody in order to develop more stable variants of the target antibody. The red lines indicate amino acid changes within the Fc region.

observed for many monoclonal antibodies and aptamers (nucleic acid molecules that bind with a similarly high affinity to their target molecules; see chapter 6). It is envisioned that, like monoclonal antibodies, affibodies may be used in a variety of both therapeutic and diagnostic applications. For example, affibodies' small size facilitates their penetration into tumors, even coupled to radiolabeled chemicals or toxins, often being much more efficient than similarly constructed monoclonal antibodies. In one study, it was observed that, in a mouse model, labeled affibodies were more efficient than labeled antibody fragments in penetrating the blood-brain barrier. For therapeutic applications that employ monoclonal antibody fragments (scFv, Fv, or Fab), affibodies are often equally or even more effective, and affibodies often have a low level of immunogenicity. At this stage, it appears that the efficacy of affibody molecule–drug complexes compared to antibody and antibody fragment–drug complexes will need to be tested on a case-by-case basis. This fact notwithstanding, this technology provides scientists with an increased range of options for treating various diseases.

Table 5.6 Some of the reported affibody molecules, including a few that are either commercially available or in clinical testing

Affibody target	Target size (kDa)	Affibody with the highest affinity for the target (nM)
α-Synuclein	15	240
Beta amyloid peptide	4.5	0.3
Apolipoprotein A1	28	1,000
Complement factor 5	190	Not reported
Carbonic anhydrase IX	42	1.2
IL-2 receptor alpha chain	55	130
CD28	90	8,000
c-Jun	70 (dimer)	5,000
Epidermal growth factor receptor	69	0.160
Factor VIII	170	5
Fibrinogen	340	Not reported
GP120	120	100
H-Ras	28	79
Human epidermal growth factor receptor 2	70	0.022
Human epidermal growth factor receptor 3	69	0.021
HPV16 E7	11	1,800
Human islet amyloid polypeptide	3.9	216
IgA	162	500
Insulin-like growth factor receptor	103	2.3
IgE	190	Not reported
IgM	970	Not reported
Interleukin-1	17	Not reported
Interleukin-6	22	0.5
Interleukin-8	8	Not reported
Interleukin-17	30 (dimer)	0.0003
Insulin	6	110
Light-oxygen-voltage-sensing domain	17	26
Murine IgG1	162	1.8
Platelet-derived growth factor receptor β	58	0.4
Staphylococcal protein A domain	7	0.016
Raf-1	73	90
RSV G protein	11	10,000
Taq polymerase	90	30
Tumor necrosis factor-α	54 (trimer)	0.095
Transferrin	80	400
Transthyretin	54	Not reported
Vascular endothelial growth factor	12	38
Vascular endothelial growth factor receptor 2	83	0.2
HER2-specific antibody	7	270
IgA-specific antibody	7	900
Taq-specific antibody	7	700

summary

A large number of human proteins that are useful as therapeutic agents have been synthesized from cloned cDNAs in bacteria. Because most of these proteins are from eukaryotic organisms, the original strategy for producing these proteins involved isolating mRNA enriched in the target mRNA, synthesizing a cDNA library, and subcloning the selected target cDNA into an appropriate expression vector. More recently, researchers have isolated target cDNAs based on their knowledge of the human genome. In some instances, novel and useful variants of these proteins can be constructed either by shuffling functional domains of related genes or by directed replacement of functional domains of the cloned cDNA.

Some genetically engineered enzymes may be used as therapeutic agents. For example, recombinant DNase I has been used in an aerosol form to decrease the viscosity of the mucus found in the lungs of patients with cystic fibrosis. In addition, phenylalanine ammonia lyase may help patients with phenylketonuria as a replacement for phenylalanine hydroxylase, α_1-antitrypsin may be used to limit some infections, and glycosidases may be utilized to convert blood groups A, B, and AB to type O. For diseases where a therapeutic protein is only found in bacterial or fungal organisms, its immunogenicity may be avoided either by pegylation of the nonhuman protein or genetic modification of the gene or cDNA that encodes the nonhuman protein, to prevent any adverse immune response in humans. Enzyme replacement therapy might also be an effective strategy for mitochondrial disorders. In this regard, it is possible to synthesize the therapeutic enzyme as a fusion protein with a cell-penetrating peptide that facilitates transport across the plasma and mitochondrial membranes.

Certain therapeutic agents may be delivered directly to their target cells by expressing the genes for these proteins in a bacterium that is normally associated with human tissues and has previously been shown to be safe, such as *L. lactis*. For example, the bacterium *L. lactis* was used to deliver interleukin-10 to the human intestinal tract as a means of treating individuals with Crohn's disease, leptin to obese mice to significantly reduce their food intake, and a unique single-chain insulin that has *in vitro* insulin-like biological activity. It is also possible to inhibit the enzyme L-DOPA decarboxylase from *E. faecalis* in the gut to prevent the premature conversion of L-DOPA, used to treat Parkinson's disease, to dopamine.

The development of recombinant DNA and monoclonal antibody technologies, combined with an understanding of the molecular structure and function of immunoglobulin molecules, has provided specific antibodies as therapeutic agents to treat various diseases. Antibody genes can be readily manipulated, because the various functions of an antibody molecule are confined to discrete domains.

Drugs, prodrugs, or enzymes can be coupled to monoclonal antibodies or antibody fragments that are specific for proteins found only on the surfaces of certain cells, e.g., tumor cells. These antibody–drug or antibody–enzyme combinations act as therapeutic agents. However, if the therapy requires multiple treatments, the antibody component should be from a human source to prevent immunological cross-reactivity and sensitization of the patient. To achieve this, rodent monoclonal antibodies are "humanized" by substituting into human antibodies only the CDRs of the rodent monoclonal antibodies. In addition, it has become possible to produce and select human monoclonal antibodies in *E. coli* and in transgenic mice.

It is now possible to select, as well as produce in combinatorial libraries, functional antibodies and antibody fragments in *E. coli*, thereby bypassing the tedious and time-consuming process of hybridoma production. Moreover, the CDR sequences in combinatorial libraries may be shuffled to increase the number of unique monoclonal antibodies in a particular library. In addition to monoclonal antibodies (and their derivatives) directed against various types of cancers, researchers have selected and tested a number of these molecules that are effective therapeutic agents against a wide range of pathogenic agents.

REFERENCES

Adams GP, Weiner LM. 2005. Monoclonal antibody therapy of cancer. *Nat Biotechnol* **23:**1147–1157.

Alexander LM, Oh J-H, Stapleton DS, Schueler KL, Keller MP, Attie AD, van Pijkeren J-P. 2019. Exploiting prophage-mediated lysis for biotherapeutic release by *Lactobacillus reuteri*. *Appl Environ Microbiol* **85:**e02335–e18.

Alkawash MA, Soothill JS, Schiller NL. 2006. Alginate lyase enhances antibiotic killing of mucoid *Pseudomonas aeruginosa* in biofilms. *APMIS* **114:**131–138.

Aslam B, Wang W, Arshad MI, Khurshid M, Muzammil S, Rasool MH, Nisar MA, Alvi RF, Aslam MA, Qamar MU, Salamat MKF, Baloch Z. 2018. Antibiotic resistance: a rundown of a global crisis. *Infect Drug Resist* **11:**1645–1658.

Bahey-El-Din M, Gahan CGM, Griffin BT. 2010. *Lactococcus lactis* as a cell factory for delivery of therapeutic proteins. *Curr Gene Ther* **10:**34–45.

Barbas CF III, Burton DR. 1996. Selection and evolution of high-affinity human anti-viral antibodies. *Trends Biotechnol* **14:**230–234.

Bell SM, Wendt DJ, Zhang Y, Taylor TW, Long S, Tsruda L, Zhao B, Laipis P, Fitzpatrick PA. 2017. Formulation and PEGylation optimization of the therapeutic PEGylated phenylalanine ammonia lyase for the treatment of phenylketonuria. *PLoS ONE* **12**:e0173269.

Bermúdez-Humarán LG, Nouaille S, Zilberfarb V, Corthier G, Langella P, Issad T. 2007. Effect of intranasal administration of a leptin-secreting *Lactococcus lactis* recombinant on food intake, body weight, and immune response of mice. *Appl Environ Microbiol* **73:**5300–5307.

Booth BJ, Ramakrishnan B, Narayan K, Wollacott AM, Babcock GJ, Shriver Z, Viswanathan K. 2018. Extending human IgG half-life using structure-guided design. *MAbs* **10:**1098–1110.

Braat H, Rottiers P, Hommes DW, Huyghebaert N, Remaut E, Remon J-P, van Deventer SJH, Neirynck S, Peppelenbosch MP, Steidler L. 2006. A phase I trial with transgenic bacteria expressing interleukin-10 in Crohn's disease. *Clin Gastroenterol Hepatol* **4:**754–759.

Brideau-Andersen AD, Huang X, Sun S-CC, Chen TT, Stark D, Sas IJ, Zadik L, Dawes GN, Guptill DR, McCord R, Govindarajan S, Roy A, Yang S, Gao J, Chen YH, Skartved NJØ, Pedersen AK, Lin D, Locher CP, Rebbapragada I, Jensen AD, Bass SH, Nissen TLS, Viswanathan S, Foster GR, Symons JA, Patten PA. 2007. Directed evolution of gene-shuffled IFN-α molecules with activity profiles tailored for treatment of chronic viral diseases. *Proc Natl Acad Sci USA* **104:**8269–8274.

Bron PA, Kleerebezem M. 2018. Lactic acid bacteria for delivery of endogenous or therapeutic molecules. *Front Microbiol* **9:**1821.

Buss CG, Bhatia SN. 2020. Nanoparticle delivery of immunostimulatory oligonucleotides enhances response to checkpoint inhibitor therapeutics. *Proc Natl Acad Sci USA* **117:**13428–13436.

Chen Z, Moayeri M, Zhao H, Crown D, Leppla SH, Purcell RH. 2009. Potent neutralization of anthrax edema toxin by a humanized monoclonal antibody that competes with calmodulin for edema factor binding. *Proc Natl Acad Sci USA* **106:**13487–13492.

Collet TA, Roben P, O'Kennedy R, Barbas CF III, Burton DR, Lerner RA. 1992. A binary plasmid system for shuffling combinatorial antibody libraries. *Proc Natl Acad Sci USA* **89:**10026–10030.

Cook DP, Gysemans C, Mathieu C. 2018. *Lactococcus lactis* as a versatile vehicle for tolerogenic immunotherapy. *Front Immunol* **8:**1961.

Curnis F, Sacchi A, Borgna L, Magni F, Gasparri A, Corti A. 2000. Enhancement of tumor necrosis factor α antitumor immunotherapeutic properties by targeted delivery to aminopeptidase N (CD13). *Nat Biotechnol* **18:**1185–1190.

Daniel C, Roussel Y, Kleerebezem M, Pot B. 2011. Recombinant lactic acid bacteria as mucosal biotherapeutic agents. *Trends Biotechnol* **29:**499–508.

Davis-Gardner ME, Alfant B, Weber JA, Gardner MR, Farzan M. 2020. A bispecific antibody that simultaneously recognizes the V2- and V3-glycan epitopes of the HIV-1 envelope glycoprotein is broader and more potent than its parental antibodies. *MBio* **11:**e03080–e19.

de Vos WM, Hugenholtz J. 2004. Engineering metabolic highways in lactococci and other lactic acid bacteria. *Trends Biotechnol* **22:**72–79.

Dübel S. 2007. Recombinant therapeutic antibodies. *Appl Microbiol Biotechnol* **74:**723–729.

Ferrandis E, Pradhananga SL, Touvay C, Kinoshita C, Wilkinson IR, Stafford K, Wu Z, Strasburger CJ, Sayers JR, Artymiuk PJ, Ross RJ. 2010. Immunogenicity, toxicology, pharmacokinetics and pharmacodynamics of growth hormone ligand-receptor fusions. *Clin Sci (Lond)* **119:**483–491.

Franchini M, Marano G, Pati I, Candura F, Profili S, Veropalumbo E, Masiello F, Catalano L, Piccinini V, Vaglio S, Pupella S, Liumbruno GM. 2019. Emicizumab for the treatment of haemophilia A: a narrative review. *Blood Transfus* **17:**223–228.

Frederick E. 2019. Bacterial toxin linked to severe alcoholic liver disease. *Science* **366:**784.

Friedman ND, Temkin E, Carmeli Y. 2016. The negative impact of antibiotic resistance. *Clin Microbiol Infect* **22:**416–422.

Gama L, Koup RA. 2018. New-generation high-potency and designer antibodies: role in HIV-1 treatment. *Annu Rev Med* **69:**409–419.

Gámez A, Sarkissian CN, Wang L, Kim W, Straub M, Patch MG, Chen L, Striepeke S, Fitzpatrick P, Lemontt JF, O'Neill C, Scriver CR, Stevens RC. 2005. Development of pegylated forms of recombinant *Rhodosporidium toruloides* phenylalanine ammonia-lyase for the treatment of classical phenylketonuria. *Mol Ther* **11:**986–989.

Garfall AL, June CH. 2019. Trispecific antibodies offer a third way forward for anticancer immunotherapy. *Nature* **575:**450–451.

Gopalakrishnan V, et al. 2018. Gut microbiome modulates response to anti-PD-1 immunotherapy in melanoma patients. *Science* **359:**97–103.

Grilo AL, Mantalaris A. 2019. The increasingly human and profitable monoclonal antibody market. *Trends Biotechnol* **37:**9–16.

Guggenbichler F, Büttner C, Rudolph W, Zimmermann K, Gunzer F, Pöhlmann C. 2016. Design of a covalently linked human interleukin-10 fusion protein and its secretory expression in *Escherichia coli*. *Appl Microbiol Biotechnol* **100:**10479–10493.

Hanniffy S, Wiedermann U, Repa A, Mercenier A, Daniel C, Fioramonti J, Tlaskolova H, Kozakova H, Israelsen H, Madsen S, Vrang A, Hols P, Delcour J, Bron P, Kleerebezem M, Wells J. 2004. Potential and opportunities for use of recombinant lactic acid bacteria in human health. *Adv Appl Microbiol* **56:**1–64.

Hemmerle T, Doll F, Neri D. 2014. Antibody-based delivery of IL4 to the neovasculature cures mice with arthritis. *Proc Natl Acad Sci USA* **111:**12008–12012.

Holliger P, Hudson PJ. 2005. Engineered antibody fragments and the rise of single domains. *Nat Biotechnol* **23:**1126–1136.

Huennekens FM. 1994. Tumor targeting: activation of prodrugs by enzyme-monoclonal antibody conjugates. *Trends Biotechnol* **12:**234–239.

Huse WD, Sastry L, Iverson SA, Kang AS, Alting-Mees M, Burton DR, Benkovic SJ, Lerner RA. 1989. Generation of a large combinatorial library of the immunoglobulin repertoire in phage lambda. *Science* **246:**1275–1281.

Isabella VM, Ha BN, Castillo MJ, Lubkowicz DJ, Rowe SE, Millet YA, Anderson CL, Li N, Fisher AB, West KA, Reeder PJ, Momin MM, Bergeron CG, Guilmain SE, Miller PF, Kurtz CB, Falb D. 2018. Development of a synthetic live bacterial therapeutic for the human metabolic disease phenylketonuria. *Nat Biotechnol* **36:**857–864.

Jakobovits A, Amado RG, Yang X, Roskos L, Schwab G. 2007. From XenoMouse technology to panitumumab, the first fully human antibody product from transgenic mice. *Nat Biotechnol* **25:**1134–1143.

Jean F, Thomas L, Molloy SS, Liu G, Jarvis MA, Nelson JA, Thomas G. 2000. A protein-based therapeutic for human cytomegalovirus infection. *Proc Natl Acad Sci USA* **97:**2864–2869.

John LB, Devaud C, Duong CPM, Yong CS, Beavis PA, Haynes NM, Chow MT, Smyth MJ, Kershaw MH, Darcy PK. 2013. Anti-PD-1 antibody therapy potently enhances the eradication of established tumors by gene-modified T cells. *Clin Cancer Res* **19:**5636–5646.

Jouault A, Gobet A, Simon M, Portier E, Perennou M, Corre E, Gaillard F, Vallenet D, Michel G, Fleury Y, Bazire A, Dufour A. 2020. Alterocin, an antibiofilm protein secreted by *Pseudoalteromonas* sp. strain 3J6. *Appl Environ Microbiol* **86:**e00893–e20.

Kang JC, Sun W, Khare P, Karimi M, Wang X, Shen Y, Ober RJ, Ward ES. 2019. Engineering a HER2-specific antibody-drug conjugate to increase lysosomal delivery and therapeutic efficacy. *Nat Biotechnol* **37:**523–526.

Keeffe JR, Gnanapragasam PN, Gillespie SK, Yong J, Bjorkman PJ, Mayo SL. 2011. Designed oligomers of cyanovirin-N show enhanced HIV neutralization. *Proc Natl Acad Sci USA* **108:**14079–14084.

Keyvani V, Farshchian M, Esmaeili SA, Yari H, Moghbeli M, Nezhad SK, Abbaszadegan MR. 2019. Ovarian cancer stem cells and targeted therapy. *J Ovarian Res* **12:**120.

Kadonosono T, Yimchuen W, Ota Y, See K, Furuta T, Shiozawa T, Kitazawa M, Goto Y, Patil A, Kuchimaru T, Kizaka-Kondoh S. 2020. Design strategy to create antibody mimetics harbouring immobilised complementarity determining region peptides for practical use. *Sci Rep* **10:**891.

Kaiser J. 2020. Forced into battle. *Science* **368:**930–933.

Kreitman RJ. 1999. Immunotoxins in cancer therapy. *Curr Opin Immunol* **11:**570–578.

Krawczyk A, Arndt MAE, Grosse-Hovest L, Weichert W, Giebel B, Dittmer U, Hengel H, Jäger D, Schneweis KE, Eis-Hübinger AM, Roggendorf M, Krauss J. 2013. Overcoming drug-resistant herpes simplex virus (HSV) infection by a humanized antibody. *Proc Natl Acad Sci USA* **110:**6760–6765.

Kufer P, Lutterbüse R, Baeuerle PA. 2004. A revival of bispecific antibodies. *Trends Biotechnol* **22:**238–244.

Kwan DH, Constantinescu I, Chapanian R, Higgins MA, Kötzler MP, Samain E, Boraston AB, Kizhakkedathu JN, Withers SG. 2015. Towards efficient enzymes for the generation of universal blood through structure-guided directed evolution. *J Am Chem Soc* **137:**5695–5705.

Lagassé HA, Alexaki A, Simhadri VL, Katagiri NH, Jankowski W, Sauna ZE, Kimchi-Sarfaty C. 2017. Recent advances in (therapeutic protein) drug development. *F1000 Res* **6:**113.

Lamppa JW, Ackerman ME, Lai JI, Scanlon TC, Griswold KE. 2011. Genetically engineered alginate lyase-PEG conjugates exhibit enhanced catalytic function and reduced immunoreactivity. *PLoS One* **6:**e17042.

Leader B, Baca QJ, Golan DE. 2008. Protein therapeutics: a summary and pharmacological classification. *Nat Rev Drug Discov* **7:**21–39.

Liu QP, Sulzenbacher G, Yuan H, Bennett EP, Pietz G, Saunders K, Spence J, Nudelman E, Levery SB, White T, Neveu JM, Lane WS, Bourne Y, Olsson ML, Henrissat B, Clausen H. 2007. Bacterial glycosidases for the production of universal red blood cells. *Nat Biotechnol* **25:**454–464.

Liu X, Lagenaur LA, Simpson DA, Essenmacher KP, Frazier-Parker CL, Liu Y, Tsai D, Rao SS, Hamer DH, Parks TP, Lee PP, Xu Q. 2006. Engineered vaginal *Lactobacillus* strain for mucosal delivery of the human immunodeficiency virus inhibitor cyanovirin-N. *Antimicrob Agents Chemother* **50:**3250–3259.

Lonberg N. 2005. Human antibodies from transgenic animals. *Nat Biotechnol* **23:**1117–1125.

López-Igual R, Bernal-Bayard J, Rodríguez-Patón A, Ghigo JM-M, Mazel D. 2019. Engineered toxin-intein antimicrobials can selectively target and kill antibiotic-resistant bacteria in mixed populations. *Nat Biotechnol* **37:**755–760.

Lorenzen N, Cupit PM, Einer-Jensen K, Lorenzen E, Ahrens P, Secombes CJ, Cunningham C. 2000. Immunoprophylaxis in fish by injection of mouse antibody genes. *Nat Biotechnol* **18:**1177–1180.

Mao R, Chen Y, Chi Z, Wang Y. 2019. Insulin and its single-chain analogue. *Appl Microbiol Biotechnol* **103:**8737–8751.

Marks JD, Griffiths AD, Malmqvist M, Clackson TP, Bye JM, Winter G. 1992. By-passing immunization: building high affinity human antibodies by chain shuffling. *Biotechnology (N Y)* **10:**779–783.

Matson V, Fessler J, Bao R, Chongsuwat T, Zha Y, Alegre M-L, Luke JJ, Gajewski TF. 2018. The commensal microbiome is associated with anti-PD-1 efficacy in metastatic melanoma patients. *Science* **359:**104–108.

Mayorov AV, Amara N, Chang JY, Moss JA, Hixon MS, Ruiz DI, Meijler MM, Zorrilla EP, Janda KD. 2008. Catalytic antibody degradation of ghrelin increases whole-body metabolic rate and reduces refeeding in fasting mice. *Proc Natl Acad Sci USA* **105:**17487–17492.

Mazor Y, Van Blarcom T, Mabry R, Iverson BL, Georgiou G. 2007. Isolation of engineered, full-length antibodies from libraries expressed in *Escherichia coli*. *Nat Biotechnol* **25:**563–565.

Merchant M, Ma X, Maun HR, Zheng Z, Peng J, Romero M, Huang A, Yang N-Y, Nishimura M, Greve J, Santell L, Zhang Y-W, Su Y, Kaufman DW, Billeci KL, Mai E, Moffat B, Lim A, Duenas ET, Phillips HS, Xiang H, Young JC, Vande Woude GF, Dennis MS, Reilly DE, Schwall RH, Starovasnik MA, Lazarus RA, Yansura DG. 2013. Monovalent antibody design and mechanism of action of onartuzumab, a MET antagonist with anti-tumor activity as a therapeutic agent. *Proc Natl Acad Sci USA* **110:**E2987–E2996.

Molina A. 2008. A decade of rituximab: improving survival outcomes in non-Hodgkin's lymphoma. *Annu Rev Med* **59:**237–250.

Murata K, Inose T, Hisano T, Abe S, Yonemoto Y, Yamashita T, Takagi M, Sakaguchi K, Kimura A, Imanaka T. 1993. Bacterial alginate lyase: enzymology, genetics and application. *J Ferment Bioeng* **76:**427–437.

Ng DTW, Sarkar CA. 2011. Nisin-inducible secretion of a biologically active single-chain insulin analog by *Lactococcus lactis* NZ9000. *Biotechnol Bioeng* **108:**1987–1996.

O'Neill C. 2019. Gut microbes metabolize Parkinson's disease drug. *Science* **364:**1030–1031.

Papadopoulou LC, Tsiftsoglou AS. 2011. Transduction of human recombinant proteins into mitochondria as a protein therapeutic approach for mitochondrial disorders. *Pharm Res* **28:**2639–2656.

Pardoll DM. 2012. The blockade of immune checkpoints in cancer immunotherapy. *Nat Rev Cancer* **12:**252–264.

Plavec TV, Berlec A. 2019. Engineering of lactic acid bacteria for delivery of therapeutic proteins and peptides. *Appl Microbiol Biotechnol* **103:**2053–2066.

Podust VN, Balan S, Sim B-C, Coyle MP, Ernst U, Peters RT, Schellenberger V. 2016. Extension of in vivo half-life of biologically active molecules by XTEN protein polymers. *J Control Release* **240:**52–66.

Qiu X-Q, Wang H, Cai B, Wang L-L, Yue S-T. 2007. Small antibody mimetics comprising two complementarity-determining regions and a framework region for tumor targeting. *Nat Biotechnol* **25:**921–929.

Rahfeld P, Sim L, Moon H, Constantinescu I, Morgan-Lang C, Hallam SJ, Kizhakkedathu JN, Withers SG. 2019. An enzymatic pathway in the human gut microbiome that converts A to universal O type blood. *Nat Microbiol* **4:**1475–1485.

Rapoport M, Saada A, Elpeleg O, Lorberboum-Galski H. 2008. TAT-mediated delivery of LAD restores pyruvate dehydrogenase complex activity in the mitochondria of patients with LAD deficiency. *Mol Ther* **16:**691–697.

Rapoport M, Salman L, Sabag O, Patel MS, Lorberboum-Galski H. 2011. Successful TAT-mediated enzyme replacement therapy in a mouse model of mitochondrial E3 deficiency. *J Mol Med (Berl)* **89:**161–170.

Ratner M. 2014. Genentech's glyco-engineered antibody to succeed Rituxan. *Nat Biotechnol* **32:**6–7.

Maini Rekdal V, Bess EN, Bisanz JE, Turnbaugh PJ, Balskus EP. 2019. Discovery and inhibition of an interspecies gut bacterial pathway for Levodopa metabolism. *Science* **364:**eaau6323.

Reichert JM. 2006. Trends in US approvals: new biopharmaceuticals and vaccines. *Trends Biotechnol* **24:**293–298.

Reichert JM, Rosensweig CJ, Faden LB, Dewitz MC. 2005. Monoclonal antibody successes in the clinic. *Nat Biotechnol* **23:**1073–1078.

Roberts NJ, Zhou S, Diaz LA Jr, Holdhoff M. 2011. Systemic use of tumor necrosis factor alpha as an anticancer agent. *Oncotarget* **2:**739–751.

Routy B, Le Chatelier E, Derosa L, Duong CPM, Alou MT, Daillère R, Fluckiger A, Messaoudene M, Rauber C, Roberti MP, Fidelle M, Flament C, Poirier-Colame V, Opolon P, Klein C, Iribarren K, Mondragón L, Jacquelot N, Qu B, Ferrere G, Clémenson C, Mezquita L, Masip JR, Naltet C, Brosseau S, Kaderbhai C, Richard C, Rizvi H, Levenez F, Galleron N, Quinquis B, Pons N, Ryffel B, Minard-Colin V, Gonin P, Soria J-C, Deutsch E, Loriot Y, Ghiringhelli F, Zalcman G, Goldwasser F, Escudier B, Hellmann MD, Eggermont A, Raoult D, Albiges L, Kroemer G, Zitvogel L. 2018. Gut microbiome influences efficacy of PD-1-based immunotherapy against epithelial tumors. *Science* **359:**91–97.

Rubinfeld B, Upadhyay A, Clark SL, Fong SE, Smith V, Koeppen H, Ross S, Polakis P. 2006. Identification and immunotherapeutic targeting of antigens induced by chemotherapy. *Nat Biotechnol* **24:**205–209.

Salvat RS, Verma D, Parker AS, Kirsch JR, Brooks SA, Bailey-Kellogg C, Griswold KE. 2017. Computationally optimized deimmunization libraries yield highly mutated enzymes with low immunogenicity and enhanced activity. *Proc Natl Acad Sci USA* **114:**E5085–E5093.

Sampson JH, Crotty LE, Lee S, Archer GE, Ashley DM, Wikstrand CJ, Hale LP, Small C, Dranoff G, Friedman AH, Friedman HS, Bigner DD. 2000. Unarmed, tumor-specific monoclonal antibody effectively treats brain tumors. *Proc Natl Acad Sci USA* **97:**7503–7508.

Sarkissian CN, Gámez A. 2005. Phenylalanine ammonia lyase, enzyme substitution therapy for phenylketonuria, where are we now? *Mol Genet Metab* **86**(Suppl 1)**:**S22–S26.

Sarkissian CN, Shao Z, Blain F, Peevers R, Su H, Heft R, Chang TMS, Scriver CR. 1999. A different approach to treatment of phenylketonuria: phenylalanine degradation with recombinant phenylalanine ammonia lyase. *Proc Natl Acad Sci USA* **96:**2339–2344.

Schellenberger V, Wang C-W, Geething NC, Spink BJ, Campbell A, To W, Scholle MD, Yin Y, Yao Y, Bogin O, Cleland JL, Silverman J, Stemmer WPC. 2009. A recombinant polypeptide extends the in vivo half-life of peptides and proteins in a tunable manner. *Nat Biotechnol* **27:**1186–1190.

Söderlind E, Strandberg L, Jirholt P, Kobayashi N, Alexeiva V, Åberg A-M, Nilsson A, Jansson B, Ohlin M, Wingren C, Danielsson L, Carlsson R, Borrebaeck CAK. 2000. Recombining germline-derived CDR sequences for creating diverse single-framework antibody libraries. *Nat Biotechnol* **18:**852–856.

Ståhl S, Gräslund T, Eriksson Karlström A, Frejd FY, Nygren P-Å, Löfblom J. 2017. Affibody molecules in biotechnological and medical applications. *Trends Biotechnol* **35:**691–712.

Steidler L, Hans W, Schotte L, Neirynck S, Obermeier F, Falk W, Fiers W, Remaut E. 2000. Treatment of murine colitis by *Lactococcus lactis* secreting interleukin-10. *Science* **289:**1352–1355.

Steidler L, Neirynck S, Huyghebaert N, Snoeck V, Vermeire A, Goddeeris B, Cox E, Remon JP, Remaut E. 2003. Biological containment of genetically modified *Lactococcus lactis* for intestinal delivery of human interleukin 10. *Nat Biotechnol* **21:**785–789.

Storek KM, Auerbach MR, Shi H, Garcia NK, Sun D, Nickerson NN, Vij R, Lin Z, Chiang N, Schneider K, Wecksler AT, Skippington E, Nakamura G, Seshasayee D, Koerber JT, Payandeh J, Smith PA, Rutherford ST. 2018. Monoclonal antibody targeting the β-barrel assembly machine of *Escherichia coli* is bactericidal. *Proc Natl Acad Sci USA* **115:**3692–3697.

Strisciuglio P, Concolino D. 2014. New strategies for the treatment of phenylketonuria (PKU). *Metabolites* **4:**1007–1017.

Subramanian GM, Fiscella M, Lamousé-Smith A, Zeuzem S, McHutchison JG. 2007. Albinterferon α-2b: a genetic fusion protein for the treatment of chronic hepatitis C. *Nat Biotechnol* **25:**1411–1419.

Ta AN, McNaughton BR. 2017. Antibody and antibody mimetic immunotherapeutics. *Future Med Chem* **9:**1301–1304.

Teteneva NA, Mart'yanov SV, Esteban-López M, Kahnt J, Glatter T, Netrusov AI, Plakunov VK, Sourjik V. 2020. Multiple drug-induced stress responses inhibit formation of *Escherichia coli* biofilms. *Appl Environ Microbiol* **86:**e01113–e01120.

Trang VH, Zhang X, Yumul RC, Zeng W, Stone IJ, Wo SW, Dominguez MM, Cochran JH, Simmons JK, Ryan MC, Lyon RP, Senter PD, Levengood MR. 2019. A coiled-coil masking domain for selective activation of therapeutic antibodies. *Nat Biotechnol* **37:**761–765.

Ulmer JS, Herzka A, Toy KJ, Baker DL, Dodge AH, Sinicropi D, Shak S, Lazarus RA. 1996. Engineering actin-resistant human DNase I for treatment of cystic fibrosis. *Proc Natl Acad Sci USA* **93:**8225–8229.

Vaughan TJ, Williams AJ, Pritchard K, Osbourn JK, Pope AR, Earnshaw JC, McCafferty J, Hodits RA, Wilton J, Johnson KS. 1996. Human antibodies with sub-nanomolar affinities isolated from a large non-immunized phage display library. *Nat Biotechnol* **14:**309–314.

Veronese FM, Pasut G. 2005. PEGylation, successful approach to drug delivery. *Drug Discov Today* **10:**1451–1458.

Villatoro-Hernandez J, Montes-de-Oca-Luna R, Kuipers OP. 2011. Targeting diseases with genetically engineered *Lactococcus lactis* and its course towards medical translation. *Expert Opin Biol Ther* **11:**261–267.

Walsh G. 2018. Biopharmaceutical benchmarks 2018. *Nat Biotechnol* **36:**1136–1145.

Wang M, Gao Z, Zhang Y, Pan L. 2016. Lactic acid bacteria as mucosal delivery vehicles: a realistic therapeutic option. *Appl Microbiol Biotechnol* **100:**5691–5701.

Wenthur CJ, Gautam R, Zhou B, Vendruscolo LF, Leggio L, Janda KD. 2019. Ghrelin receptor influence on cocaine reward is not directly dependent on peripheral acyl-ghrelin. *Sci Rep* **9:**1841.

Weisser NE, Hall JC. 2009. Applications of single-chain variable fragment antibodies in therapeutics and diagnostics. *Biotechnol Adv* **27:**502–520.

Wilkinson IR, Ferrandis E, Artymiuk PJ, Teillot M, Soulard C, Touvay C, Pradhananga SL, Justice S, Wu Z, Leung KC, Strasburger CJ, Sayers JR, Ross RJ. 2007. A ligand-receptor fusion of growth hormone forms a dimer and is a potent long-acting agonist. *Nat Med* **13:**1108–1113.

Wu C, Ying H, Grinnell C, Bryant S, Miller R, Clabbers A, Bose S, McCarthy D, Zhu R-R, Santora L, Davis-Taber R, Kunes Y, Fung E, Schwartz A, Sakorafas P, Gu J, Tarcsa E, Murtaza A, Ghayur T. 2007. Simultaneous targeting of multiple disease mediators by a dual-variable-domain immunoglobulin. *Nat Biotechnol* **25:**1290–1297.

Wu L, Seung E, Xu L, Rao E, Lord DM, Wei RR, Cortez-Retamozo V, Ospina B, Posternak V, Olinski G, Piepenhagen P, Francesconi E, El-Murr N, Beil C, Kirby P, Li A, Fretland J, Vicente R, Deng G, Dabdoubi T, Cameron B, Bertrand T, Ferrarei P, Pouzieux S, Lemoine C, Prades C, Park A, Qiu H, Song Z, Zhang B, Sun F, Chiron M, Rao S, Radosevic K, Yang Z-Y, Nabel GJ. 2019. Trispecific antibodies enhance the therapeutic efficiency of tumor-directed T cells through T cell receptor co-stimulation. *Nat Can* **1:**86–98.

Yi L, Jin M, Li J, Grenier D, Wang Y. 2020. Antibiotic resistance related to biofilm formation in *Streptococcus suis*. *Appl Microbiol Biotechnol* **104:**8649–8660.

Zahm J-M, Debordeaux C, Maurer C, Hubert D, Dusser D, Bonnet N, Lazarus RA, Puchelle E. 2001. Improved activity of an actin-resistant DNase I variant on the cystic fibrosis airway secretions. *Am J Respir Crit Care Med* **163:**1153–1157.

Zalevsky J, Chamberlain AK, Horton HM, Karki S, Leung IWL, Sproule TJ, Lazar GA, Roopenian DC, Desjarlais JR. 2010. Enhanced antibody half-life improves *in vivo* activity. *Nat Biotechnol* **28:**157–159.

Zeng Y, Wang J, Gu Z, Gu Z. 2019. Engineering glucose-responsive insulin. *Med Drug Discover* **3:**100010.

Zhang S, Cui B, Lai H, Liu G, Ghia EM, Widhopf GF II, Zhang Z, Wu CCN, Chen L, Wu R, Schwab R, Carson DA, Kipps TJ. 2014. Ovarian cancer stem cells express ROR1, which can be targeted for anti-cancer-stem-cell therapy. *Proc Natl Acad Sci USA* **111:**17266–17271.

Zhang S, Zhang H, Ghia EM, Huang J, Wu L, Zhang J, Lam S, Lei Y, He J, Cui B, Widhopf GF II, Yu J, Schwab R, Messer K, Jiang W, Parker BA, Carson DA, Kipps TJ. 2019. Inhibition of chemotherapy resistant breast cancer stem cells by a ROR1 specific antibody. *Proc Natl Acad Sci USA* **116:**1370–1377.

Zhu S, Liu Y, Wang PC, Gu X, Shan L. 2017. Recombinant immunotoxin therapy of glioblastoma: smart design, key findings, and specific challenges. *BioMed Res Int* 2017**:**7929286.

Zivin JA. 2000. Understanding clinical trials. *Sci Am* **282:** 69–75.

review questions

1. How are antibody light and heavy chains coordinately synthesized in *E. coli?*
2. How would you modify growth hormone to make it longer acting?
3. Why would DNase I and alginate lyase be useful for treating cystic fibrosis?
4. How is the production of alginate lyase from a cloned gene detected in *E. coli* transformants?
5. What is a combinatorial cDNA library?
6. How is bacteriophage M13 used to select Fv fragments that bind to specific target antigens?
7. What are disulfide-stabilized and scFv molecules?
8. How are enzymes coupled to monoclonal antibodies or Fv fragments used as therapeutic agents?
9. How are mouse monoclonal antibodies "humanized"? Discuss the reasons for creating humanized monoclonal antibodies.

10. Describe a protocol for producing a therapeutic agent that targets and kills a specific cell type.

11. How would you express foreign genes in mitochondria?

12. Besides pegylation, how would you extend the *in vivo* half-life of therapeutic antibodies?

13. How would you engineer TNF-α to be a more specific and effective anticancer agent?

14. What would you do to make interleukin-10 more effective for treating inflammatory bowel disease?

15. How can the gene for DNase I be manipulated so that the enzyme becomes more effective for treating cystic fibrosis patients?

16. What are inteins, and how can they be used to kill *Vibrio cholerae* cells?

17. How would you develop a strategy to protect at-risk women from HIV infection?

18. How might low levels of phenylalanine be attained, other than with a phenylalanine-free diet, in patients with the human genetic disease phenylketonuria?

19. How might particular gut bacteria affect patients being treated for Parkinson's disease?

20. What types of genetic manipulations can be used to generate a very large bacterial library of highly specific single-chain human monoclonal antibodies?

21. How would you engineer a mouse so that it produces only human antibodies?

22. What is a bispecific diabody?

23. How would you design a short peptide so that it retains the antigen-binding specificity of an entire immunoglobulin molecule?

24. How would you select antibodies against specific antigens in *E. coli?*

25. What are dual-variable-domain immunoglobulins?

26. How can you use a chemotherapy agent to facilitate the targeting of tumor cells with monoclonal antibodies?

27. How would you develop antibodies that protect people against anthrax?

28. How would you develop antibodies that facilitate weight loss?

29. How can donated blood be treated so that it all becomes type O?

30. How can the immunogenicity of an enzyme from a nonhuman source be altered so that the enzyme can be used as a human therapeutic agent?

31. How can a monoclonal antibody be used in place of blood coagulation factor VIII?

32. What is a bispecific antibody? How could it be used as a therapeutic agent directed against HIV?

33. What are immune checkpoints? How can antibodies directed against particular immune checkpoints be used to inhibit the proliferation of certain types of cancers?

34. How can antibody fragments be used to inhibit the progression of rheumatoid arthritis?

35. How can monoclonal antibodies prevent tumor regrowth and metastasis?

36. How can monoclonal antibodies be used to prevent the proliferation of antibiotic-resistant pathogenic bacteria?

Targeting Specific mRNA and DNA Sequences
Antisense RNA
Aptamers
Ribozymes and DNAzymes
Interfering RNA
Zinc Finger Nucleases
CRISPR-Cas System
Nanozymes
Nanoparticles
Engineering Bacteriophages
Viral Delivery Systems
Nonviral Delivery Systems
Direct Injection
Lipids
Bacteria
Dendrimers
Antibodies
Aptamers
Transposons
Gene Therapy
Mitochondrial Diseases
Prodrug Activation Therapy
Promoterless Gene Targeting
SUMMARY
REFERENCES
REVIEW QUESTIONS

6 Nucleic Acids as Therapeutic Agents

In many instances prior to the 1990s, the human gene that caused a particular disorder was unknown, and consequently, the root cause of the disorder was often a mystery. Moreover, many genetic diseases have multiple effects that are difficult to treat in a straightforward manner and may require diverse therapies. Because of the physiological complexity of genetic diseases, curative measures have consisted mainly of treating the symptoms by administering drugs, performing surgery, restricting dietary intake, transplanting organs or bone marrow, or transfusing blood.

To treat many genetic disorders, the use of genes as therapeutic agents (gene therapy) was a logical extension of other therapies. In its simplest form, gene therapy was envisioned as a way to correct a defect at its biological source, with a single treatment that would alleviate all symptoms of the disorder. Since 1990, there has been a proliferation of strategies designed to treat a variety of human genetic disorders. From 1989 to 2019, more than 3,000 human gene therapy clinical trials have been conducted, with tens of thousands of individuals taking part. Moreover, 126 of these clinical trials are already in phase III. Treatments for various cancers, hemophilia, AIDS, cystic fibrosis, hypercholesterolemia, amyotrophic lateral sclerosis, and many other disorders have been tested for safety and efficacy. At present, all research on human gene therapy is directed toward correcting genetic defects of somatic cells. For ethical, safety, and technical reasons, human germ line gene therapy, the introduction of DNA into egg cells, sperm, or early embryonic cells that can form part of the human germ line, is not being examined experimentally at this time. However, considerable research on genetic modification of germ line cells in nonhuman organisms such as rodents is being carried out.

Molecular Biotechnology: Principles and Applications of Recombinant DNA,
Sixth Edition. Bernard R. Glick and Cheryl L. Patten.
© 2022 American Society for Microbiology.

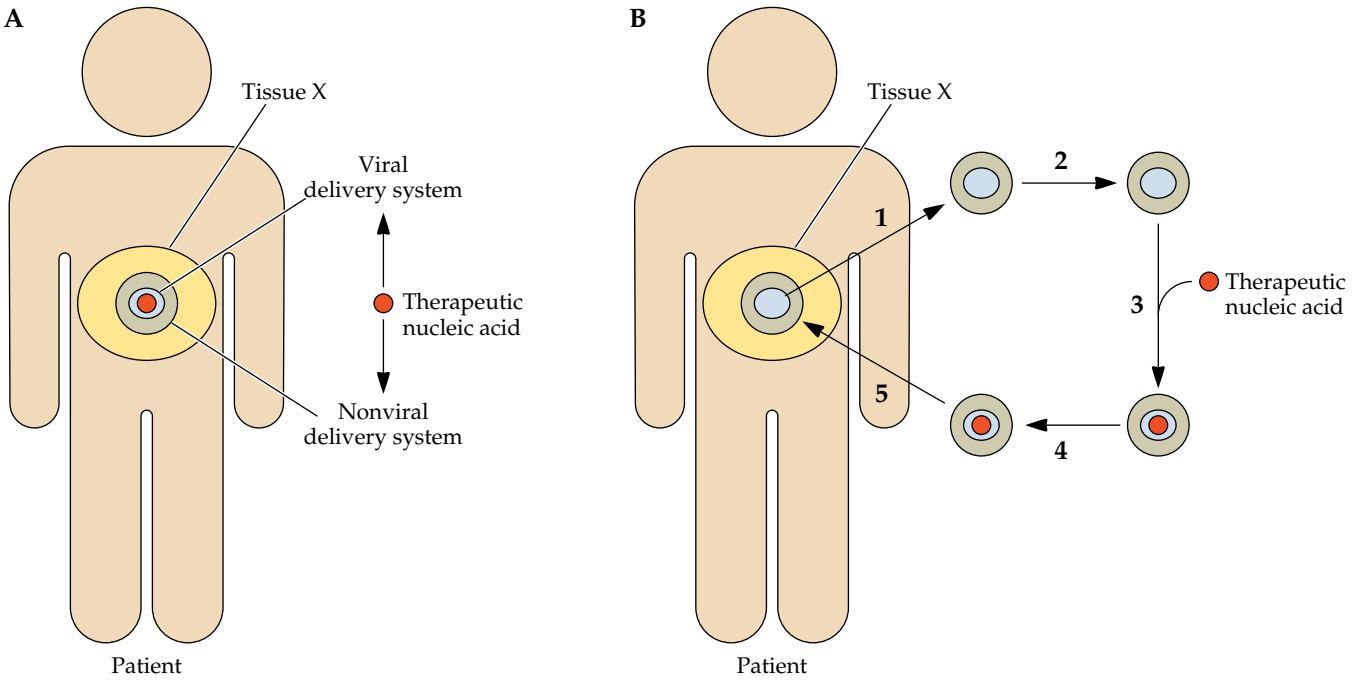


Figure 6.1 Schematic representation of *in vivo* gene therapy. **(A)** The therapeutic nucleic acid is introduced into a specific tissue by any one of various viral or nonviral delivery systems. **(B)** Schematic representation of *ex vivo* gene therapy. A cell sample containing stem cells is taken from an individual with a genetic disorder **(1)** and grown in culture **(2)**. The isolated cells are transfected with a therapeutic nucleic acid **(3)**. Modified cells are grown in culture **(4)** before transplantation or transfusion back into the patient **(5)**.

Somatic cell gene therapy is implemented in one of two ways: *in vivo* or *ex vivo*. For *in vivo* gene therapy, a therapeutic gene is introduced by either a viral or nonviral delivery system into a targeted tissue of the patient (Fig. 6.1A). On the other hand, *ex vivo* gene therapy entails collecting and culturing stem cells from an affected individual, introducing the therapeutic gene into these cultured cells, growing the cells with the therapeutic gene, and then either infusing or transplanting these cells back into the patient (Fig. 6.1B). The use of a patient's own cells ensures that after their reintroduction, no adverse immunological responses will occur. In addition, the use of a particular type of stem cell ensures that a differentiated progeny cell that otherwise might be inaccessible to genetic modification carries the therapeutic gene. Under these conditions, the genetically modified stem cells provide a continual source of "corrected" cells that function properly.

Human disorders such as cancer, inflammatory conditions, and both viral and parasitic infections often result from the overproduction of a normal protein. Therapeutic systems using nucleotide sequences are being developed to treat these types of conditions. Thus, small antisense oligonucleotides that can hybridize to a specific gene or mRNA and diminish transcription or translation, respectively, can decrease the amount of protein that is synthesized. Some synthetic nucleic acid molecules called aptamers can bind with a high affinity to specific regions of a wide range of proteins and prevent the targeted protein from functioning. Ribozymes (and deoxyribozymes) are nucleic acid sequences that bind and cleave specific RNA molecules and can

be engineered to target an mRNA and subsequently decrease the amount of a particular protein that is synthesized. Small RNA molecules, interfering RNA, can direct the sequence-specific degradation of targeted mRNAs. Zinc finger nucleases and specific single-stranded RNAs (CRISPR) in concert with an enzyme called Cas9 can cleave targeted defective genes which are subsequently repaired by the cell's DNA repair machinery.

Targeting Specific mRNA and DNA Sequences

Antisense RNA

An antisense RNA is an RNA sequence that is complementary to all or part of a functional RNA (usually an mRNA). To be a useful therapeutic agent, an antisense RNA or an RNA oligonucleotide must bind to a specified mRNA and prevent translation of the encoded protein, with the RNA duplex eventually being degraded (Fig. 6.2).

To be effective, a chemically synthesized antisense oligonucleotide (Fig. 6.2B) must be readily taken up into cells, resist degradation by cellular nucleases, and hybridize to an accessible nucleotide sequence on the target mRNA. Oligonucleotides of about 15 to 24 nucleotides in length have sufficient specificity to hybridize to a unique mRNA. Unfortunately, there are no general rules for predicting the best target sites in various RNA transcripts. Antisense oligonucleotides that are directed to the 5′ and 3′ ends of mRNAs, intron–exon boundaries, and regions that are naturally double stranded have all been shown to be effective.

Since oligonucleotides are susceptible to degradation by intracellular nucleases, it is important to synthesize molecules that are resistant to attack by nucleases without affecting the ability of the antisense oligonucleotide to hybridize to a target sequence. With this in mind, both pyrimidines (cytosine, C; thymine, T; and uracil, U) and sugar moieties of the nucleotides have been modified. Second-generation antisense drugs typically contain 2′ methoxyethyl substituted nucleotides at the 5′ and 3′ ends of the oligonucleotides, with a stretch of deoxy nucleotides in the center designed to support RNase H activity and phosphorothioates at each internucleotide linkage. The inclusion of methoxyethyl nucleotides results in a large increase in affinity toward its target nucleic acid, as well as a significant increase in nuclease resistance, resulting in drugs that typically have elimination half-lives of 2 to 4 weeks. Thus, second-generation antisense oligonucleotides are generally more specific, longer lasting, and less toxic than first-generation oligonucleotides. Third-generation antisense oligonucleotides contain a variety of modifications within the ribose ring and/or the phosphate backbone and are less toxic than either first- or second-generation antisense oligonucleotides. In recent years, a number of antisense RNA-based drugs have undergone clinical trials, with a few now approved for human use (Table 6.1; Box 6.1).

The use of an expression vector to deliver an antisense RNA into cells to suppress a pathogenic condition has been tested. For example, plasmid-based expression vectors that carry the cDNA sequence for either insulin-like growth factor 1 or insulin-like growth factor 1 receptor were constructed with the cloned sequences oriented so that the transcripts were antisense rather than mRNA (sense) sequences. Excessive levels of insulin-like growth factor 1 have been implicated in malignant glioma, which is the most common form of human brain tumor. In addition, excess production of insulin-like growth factor 1 receptor occurs in prostate carcinoma, which

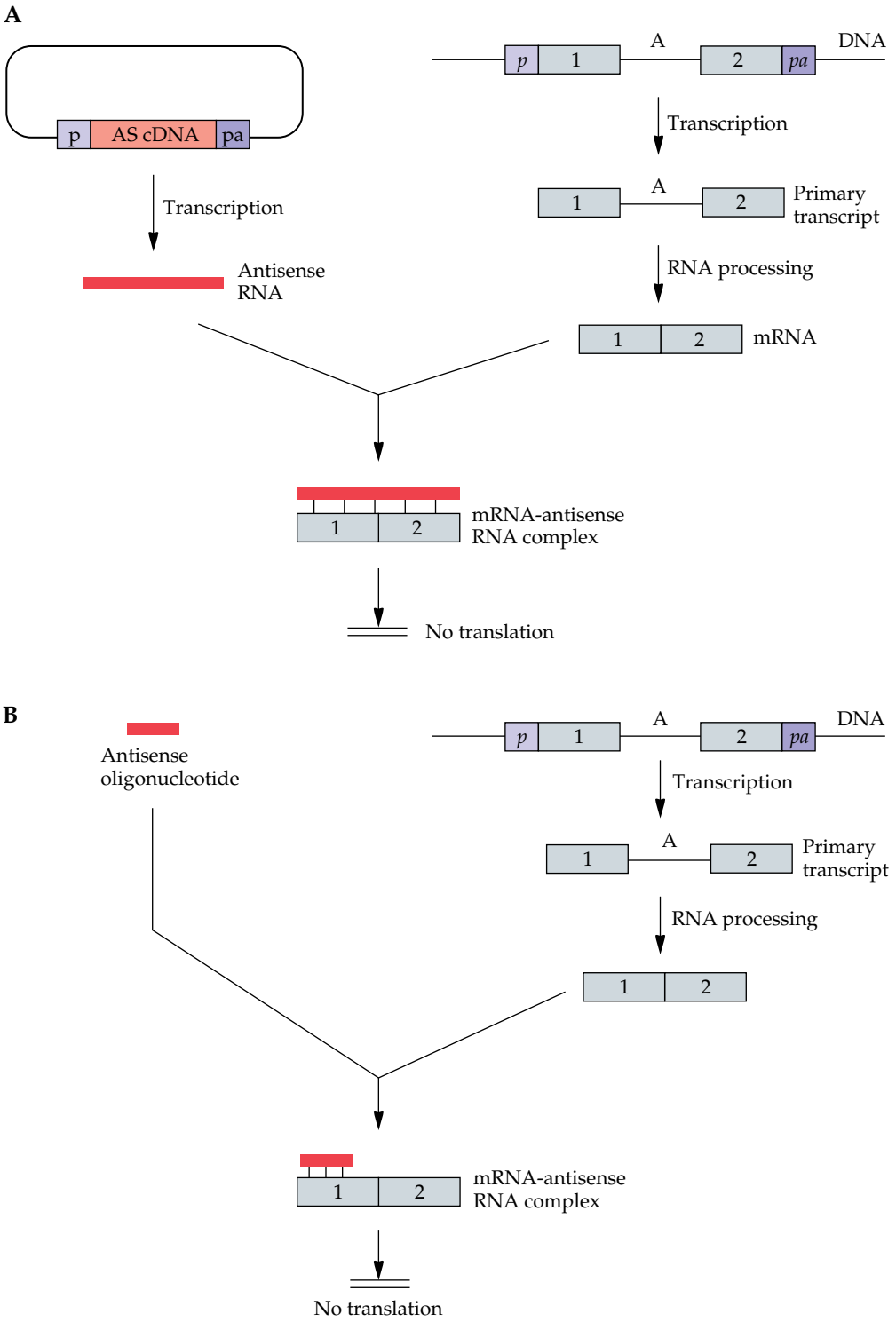


Figure 6.2 Inhibition of translation of specific mRNAs by antisense (AS) nucleic acid molecules. The promoter and polyadenylation regions are marked by *p* and *pa*, respectively, the intron is indicated by the letter A, and the exons are indicated by numbers (1 and 2). **(A)** A cDNA is cloned into an expression vector in reverse (antisense) orientation (AS cDNA), and the construct is transfected into a cell, where the antisense RNA is synthesized. The antisense RNA hybridizes to the target mRNA, and translation is blocked. **(B)** An antisense oligonucleotide is introduced into a cell, and after it hybridizes with the target mRNA, translation is blocked.

Table 6.1 Antisense RNAs that have undergone or are in the midst of clinical trials

Drug	Target	Date
Fomivirsen	Cytomegalovirus retinitis	Approved 1998; removed from market 2006
Mipomersen	Homozygous familial hypercholesterolemia	Approved 2013
Eteplirsen	Duchenne muscular dystrophy	Approved 2016
Nusinersen	Spinal muscular atrophy	Approved 2016
Inotersen	Hereditary transthyretin-mediated amyloidosis	Approved 2018
QR-010	Cystic fibrosis	Completed phase IIb, 2019
QR-421a	Retinitis pigmentosa	Phase I/II, 2019–2020
QR-110	Leber's congenital amaurosis	Recruiting for phase I/II, 2020

is a common type of cancer in males. The cDNAs encoding these proteins were inserted into a vector in the reverse-oriented (antisense) orientation under the control of the metallothionein promoter, which is induced by low levels of $ZnSO_4$. Cultured glioma cells were transfected with the vector. In the absence of $ZnSO_4$, the tumor properties were retained; however, when $ZnSO_4$ was added to the culture medium, these properties were lost (Fig. 6.3). In another experiment, nontransfected glioma cells caused tumors after they were injected into rats, whereas glioma cells that had been transfected with the vector expressing antisense insulin-like growth factor 1 cDNA did not develop tumors. Moreover, when mice were injected with rat prostate carcinoma cells that were transfected with the insulin-like growth factor 1 receptor cDNA in the antisense orientation, they developed either small or no tumors, whereas large tumors were formed when mice were treated with either nontransfected or control-transfected (i.e., transfected with an empty vector without the insulin-like growth factor 1

Spinal Muscular Atrophy and Antisense RNA

box 6.1

Spinal muscular atrophy is a relatively rare neuromuscular disease (it affects ~1 in 10,000 people at birth) that is caused by the loss of motor neurons and subsequent progressive muscle wasting. This disease is caused by an autosomal recessive genetic defect of the *SMN1* gene and is the leading genetic cause of death in infants. In 2019, the U.S. FDA approved a (one-time injection of the *SMN1* gene) gene therapy treatment, given together with a course of corticosteroids, for this disease for children under 2 years of age. The treatment cost $2.125 million U.S., making it the most expensive medication in the world at the time. This therapy was subsequently approved in Europe in 2020. Whether this treatment, which is considerably more expensive than the antisense RNA drug treatment discussed below, is utilized to any great extent remains to be seen.

Prior to the development and approval of the gene therapy treatment for spinal muscular atrophy, in December 2016, the U.S. FDA approved nusinersen (marketed as Spinraza) as a medication for spinal muscular atrophy. This antisense oligonucleotide is injected directly into the fluid-filled space at the base of the spinal column, thereby allowing for lower doses and fewer side effects. Trials of the RNA drug were halted early, as the results were overwhelmingly successful. All of the trial's initial 13 participants gained the ability to kick, sit, crawl, and stand, and following treatment for 1 to 13 months, none of them required the use of a ventilator. This success has prompted calls for all newborns to be routinely screened for spinal muscular atrophy at a very early age. Notwithstanding its very high cost (~$750,000 U.S. for 1 year's treatment), as of May 2019, nusinersen was available through public health care in more than 40 countries worldwide. Notably, health officials in both Norway and Ireland have refused to fund this drug because of what they call its "unethically high cost."

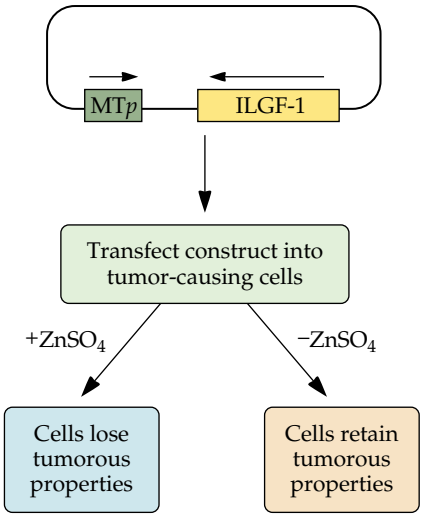


Figure 6.3 A cDNA for human insulin-like growth factor 1 (ILGF-1) cloned on a vector in the antisense orientation under the transcriptional control of a metallothionein promoter (MT*p*). Following transfection into tumor-causing cells, when low levels of $ZnSO_4$ are added to induce promoter activity, the cells have decreased tumorigenicity. The arrows above the gene and promoter indicate the normal direction of transcription. The origin of replication and other plasmid sequences have been omitted for clarity.

receptor cDNA in the antisense orientation) rat prostate carcinoma cells. In both cases, it was assumed that the antisense RNA hybridized with its complementary mRNA sequence and blocked translation of insulin-like growth factor 1 or insulin-like growth factor 1 receptor, thus preventing the proliferation of the cancer cells.

One phosphorothioate antisense oligonucleotide was approved (in 1998) by the U.S. FDA to treat cytomegalovirus infections of the retina in patients with AIDS. This particular antisense oligonucleotide, called fomivirsen and sold as Vitravene, was administered by injection directly into an affected eye after the application of a topical or local anesthetic. Fomivirsen was typically administered once every 2 weeks for 4 weeks, followed by once every 4 weeks. Before treatment with fomivirsen was started, it was deemed essential that the presence of cytomegalovirus be absolutely confirmed, since several other infective agents produce similar symptoms. In addition, this drug was somewhat problematic in that it could increase intraocular pressure so that extensive monitoring of patients was necessary. Moreover, fomivirsen treatment is not a cure for cytomegalovirus infection since there are other regions of the body that can be affected by this virus and require additional treatment. In this regard, what is required here is a treatment that more specifically targets cytomegalovirus infections. Fomivirsen was withdrawn from the marketplace in the United States in 2006 because the extensive use of antiviral compounds as a treatment for AIDS dramatically reduced the number of people with cytomegalovirus infections of the retina.

Mipomersen is an antisense therapeutic agent that targets the mRNA for apolipoprotein B. It is a second-generation antisense oligonucleotide that is 20 nucleotides long with a molecular mass of approximately 7,000 and is considered to be a cholesterol-reducing drug. Mipomersen binds to the mRNA for apolipoprotein B-100, the main component of low-density lipoprotein, causing the mRNA to be degraded by RNase H. Mipomersen is rapidly absorbed following subcutaneous (under the skin) administration and has an elimination half-life of approximately 30 days. In January 2013, following phase III clinical trials, the FDA approved the use of mipomersen as an addition to lipid-lowering medications and diet to treat patients with a rare type of high cholesterol called homozygous familial hypercholesterolemia. It is typically administered once a week by injection into the abdomen, thigh region, or outer area of the upper arm. Unfortunately, the use of mipomersen is associated with liver damage in nearly 20% of the people being treated with this therapeutic agent.

It has recently been shown that it is possible to protect mice against some retroviruses by injecting them, intravenously or intraperitoneally (into the peritoneum or body cavity), with antisense oligonucleotides that bind tightly to the viral RNA genome and thereby prevent the conversion of the viral RNA into double-stranded DNA that would be incorporated into and propagated with the host genome. In this way, the introduced antisense oligonucleotide effectively blocks replication of the retrovirus. For example, when an antisense oligonucleotide binds to the junction of a polypurine tract and a specific nucleotide sequence known as a U3 element, which is present in many retroviral genomes, a structure is formed that mimics the normal substrate for the virus-encoded enzyme RNase H (Fig. 6.4). This causes premature cleavage of the viral RNA, resulting in the destruction of the virus before reverse transcription occurs.

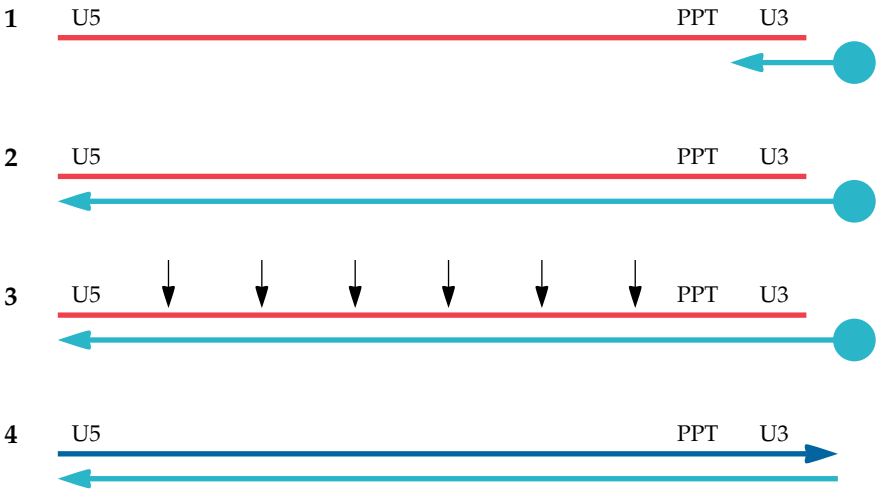


Figure 6.4 Schematic representation of the reverse transcription of retroviral RNA (red) to produce double-stranded DNA (the minus strand is light blue, and the plus strand is dark blue). The steps shown depict the stages involved in the replication of a retrovirus. In step 1, a tRNA (shown as a light blue circle) primes synthesis of the minus strand. The RNA includes U3 and U5 elements and a polypurine tract (PPT). In step 2, a complete RNA–DNA duplex is formed. In step 3, the enzyme RNase H digests the viral RNA in the RNA–DNA duplex into small pieces (the arrows indicate digestion sites). In step 4, a double-stranded DNA copy of the viral RNA is produced, which can exist as a provirus integrated into a cell's DNA. To block the formation of the minus strand and hence the synthesis of a double-stranded DNA version of the virus (and ultimately the replication of the virus), an antisense oligonucleotide that hybridizes to the PPT region is added.

This strategy is effective in protecting mice from retroviral infections. In principle, it should also work in humans and on a range of different retroviruses. However, a number of technical obstacles must be overcome before this approach is ready for testing in humans.

While there are currently no available treatments to slow or prevent the progression of Huntington's disease (an autosomal dominant neurodegenerative disease), two separate clinical trials using antisense RNA have recently shown promising results. In addition, clinical trials of antisense RNA therapy to treat several other neurodegenerative diseases, including amyotrophic lateral sclerosis, Parkinson's disease, and Alzheimer's disease, have recently (early 2020) begun. One of the major advantages of using antisense RNA, compared to interfering RNA (discussed later in this chapter), for the treatment of neurodegenerative diseases is that the antisense RNAs are readily taken up by neurons. However, repeated treatment with antisense RNAs is generally required to maintain the therapeutic effect.

Aptamers

Aptamers are short nucleic acid sequences, either RNA or DNA, that fold into a unique tertiary structure and bind very tightly to proteins, amino acids, drugs, or other molecules. They may be used either to bind directly and inhibit the functioning of a target molecule or to deliver inhibitors

Table 6.2 Comparison of some of the properties of aptamers with monoclonal antibodies

Characteristic	Monoclonal antibody	Aptamer
Size (nm)	10–15	3–4
Target	Must be immunogenic	Need not be immunogenic
Specificity	High	High
Affinity	High	High
Tissue penetration	Low due to large size	High due to small size
pH and temperature stability	Low to medium	Medium to high
Shelf life	Several months at low temperature	Several months to years at low temperature
Nuclease susceptibility	None	Moderate
Protease susceptibility	Moderate	None
Immune response	High unless it is humanized	Low to none
Cost of synthesis	High	Low to medium

to specific cells. Aptamers are typically 15 to 50 nucleotides long, have highly organized secondary and tertiary structures, and bind with very high affinity to their target molecules (i.e., $10^{-12} < K_d < 10^{-7}$, where K_d is the dissociation constant of the aptamer–target molecule complex). Thus, aptamers achieve approximately the same affinity and specificity toward their targets as therapeutic monoclonal antibodies (Table 6.2). However, because nucleic acids are not especially immunogenic, aptamers avoid the immunogenicity concerns of protein drugs and are more easily generated than therapeutic antibodies and antibody fragments. As a result, aptamers are attractive as potential therapeutic agents because of their high specificity, relative ease of production, low level or absence of immunogenicity, and long-term stability. In fact, as a consequence of these properties, the use of aptamers may be preferred over the use of antibodies in some applications.

Aptamers that are directed against specific target molecules are typically selected by a procedure known as SELEX (systematic evolution of ligands by exponential enrichment), in which RNA ligands that bind to the target molecule are highly enriched during multiple rounds of selection (Fig. 6.5). In this procedure, a random DNA sequence is cloned in a vector between two known DNA sequences. The 3′ region of the cloned sequence contains a binding (annealing) site for reverse transcriptase primers, and the 5′ region contains an annealing site for a PCR primer that will later be used to amplify the aptamer. Using T7 RNA polymerase, the double-stranded DNA is converted to RNA, which are the aptamers. In the SELEX procedure, a large library of potential aptamers, each with a different nucleotide sequence, is incubated with the target molecule, with the expectation that several different members of this collection of nucleotides will bind to various sites on the surface of the target molecule. The bound aptamer molecules are separated from the unbound aptamers before the bound RNA aptamers are dissociated from the target molecule, converted to cDNA using reverse transcriptase, and amplified by PCR. The aptamers are then retested for binding to the target molecule. The entire procedure requires several rounds of aptamer binding (to the surface of a target molecule), partitioning, and amplification of selected nucleotide sequences from an initial pool of up to 10^{16} nucleotide sequence variants. At each step, only

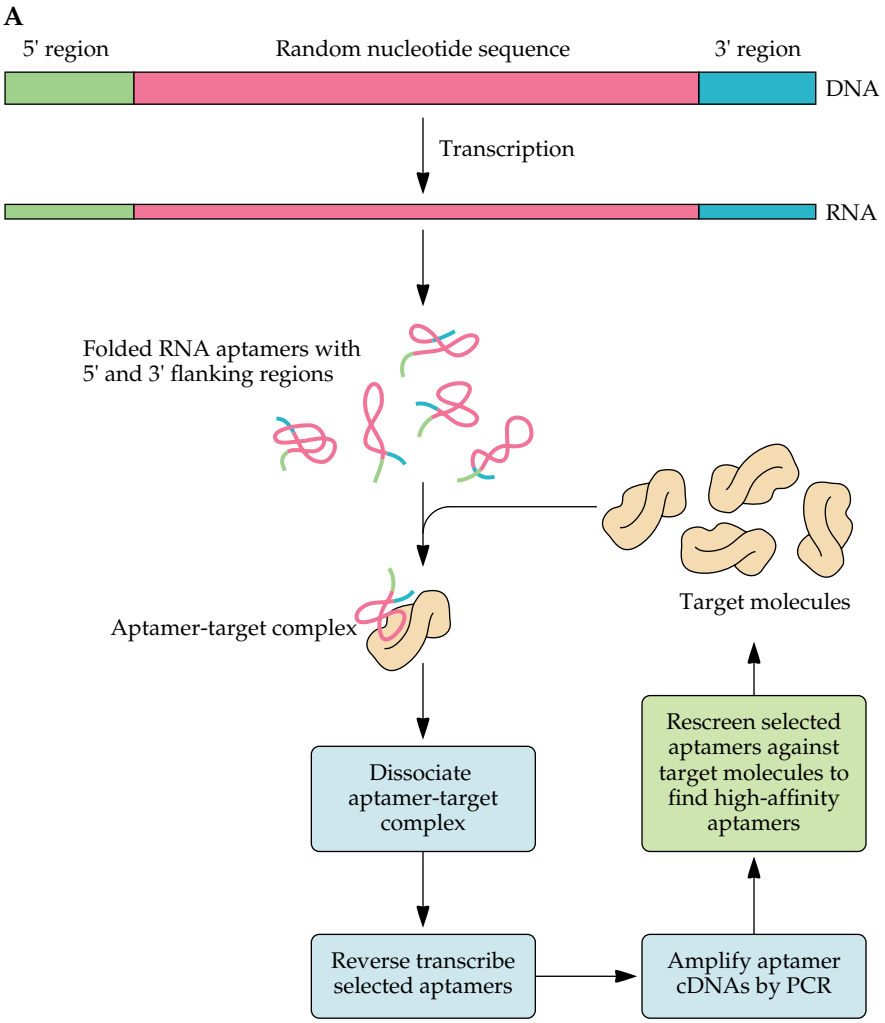


Figure 6.5 Overview of the SELEX procedure for selecting aptamers with high affinity to a target molecule (usually a protein). **(A)** The selected aptamers are typically cycled through this procedure 5 to 15 times. **(B)** A view of the surface of the target molecule binding some aptamers while not binding others.

the aptamers that bind most tightly to the target molecule are retained. The end result of this procedure is the selection of aptamers that bind to the target molecule with a relatively high affinity. Ultimately, the SELEX procedure yields one (or just a few) unique nucleic acid sequence(s) from the original mixture with a high affinity for the target molecule. To make aptamers more stable and less sensitive to nuclease digestion, the OH residues at the 2′ positions of purine nucleotides may be replaced with 2′-*O*-methyl residues, and their 3′ ends may be capped with a deoxythymidine residue. Table 6.3 lists some of the proteins against which aptamers have been generated, as well as the range of affinities of the aptamer for the target protein.

The bioavailability (i.e., delivery to the target site) of aptamers is a key issue that affects their application as *in vivo* diagnostic and therapeutic tools. The most common approaches to improve aptamer bioavailability include surrounding the aptamer with lipoproteins (e.g., liposomes) or the attachment of bulky groups, such as polyethylene glycol (pegylation), to the 5′ or 3′ ends, resulting in a decreased rate of renal (kidney) clearance and therefore an increased plasma half-life.

Table 6.3 Some proteins against which aptamers have been generated and the affinities for the proteins

Protein	K_d (nM)	Protein	K_d (nM)
Keratinocyte growth factor	0.0003	HIV gp120	5.0
HIV-1 reverse transcriptase	0.02	Tenascin C	5.0
Transforming growth factor β1	0.03	Bovine prion protein	6.8
P-selectin	0.04	Integrin	8.0
VEGF receptor	0.05	Hepatitis C virus NS3 protease	10.0
Platelet-derived growth factor	0.09	HIV-1 integrase	10.0
Immunoglobulin E	0.1	Cytotoxic T-lymphocyte-associated protein 4	10.0
Mucin 1	0.135		
Extracellular signal-regulated kinase	0.2	NS3 protease	10.0
Fibroblast growth factor 2, basic	0.35	Factor VIIa	11.0
CD4 antigen	0.5	E2F transcription factor	15.0
HIV-1 RNase H	0.5	*Yersinia pestis* tyrosine phosphatase	18.0
Factor IXa	0.58	Protein tyrosine phosphatase	18.0
Angiogenin	0.7	Anti-insulin receptor antibody MA20	30.0
Complement factor 5	1.0	Epidermal growth factor receptor variant III	33.0
Transforming growth factor β2	1.0	Ghrelin	35.0
Nuclear factor κβ	1.0	Receptor tyrosine kinase RETC634Y	35.0
Neurotensin 1	1.5	Substance P	40.0
Secretory phospholipase A2	2.0	Respiratory syncytial virus	40.0
Thrombin	2.0	HER3	45.0
Acetylcholine receptor	2.0	Gonadotropin-releasing hormone 1	50.0
αvβ3 integrin	2.0		
Angiopoietin 2	2.2	Acetylcholine-specific auto-antibodies	60.0
γ-Interferon	2.7	Nociceptin	110.0
L-selectin	3.0	Activated protein C	110.0
Calcitonin gene-related peptide	3.0	Cytohesin 2	115.0
Amylin	3.0	*Trypanosoma cruzi* cell adhesion receptor	172
U1A	4.5	*Staphylococcus enterotoxin B*	420.0
Human neutrophil elastase	5.0	Lymphocyte function-associated antigen 1	500.0
Cytohesin 1	5.0		

Approximately a dozen different aptamers for the treatment of several types of cancer have successfully undergone, or are in the midst of undergoing, clinical trials. In addition, a number of aptamers are currently being developed for the treatment of neurological diseases. Diseases of the nervous system are often problematic to treat because it is generally quite difficult to access the nervous system. However, as a consequence of their small size, aptamers can penetrate deep into many tissues and can even cross the blood-brain barrier. Moreover, the high affinity of aptamers for their target reduces both the amount of aptamer that is needed for treatment as well as the likelihood of unwanted side effects.

An aptamer known as pegaptanib received approval from the FDA in December 2004 for use as a human therapeutic agent. Pegaptanib is a pegylated (to improve its nuclease resistance) 27-nucleotide-long RNA aptamer that targets vascular endothelial growth factor (VEGF) and binds to the protein with an extremely high affinity (K_d = 0.05 nM). This secreted protein promotes the growth of new blood vessels by stimulating the endothelial cells that not only form the walls of the blood vessels but also transport nutrients and oxygen to the tissues. When retinal pigment epithelial cells begin to senesce from lack of nutrition (ischemia), VEGF acts to stimulate the synthesis

of new blood vessels (neovascularization). However, this process is imperfect, and often, the blood vessels do not form properly, so leakage results, causing scarring in the macular region of the retina, with the eventual loss of central vision. These physiological changes contribute to the condition known as age-related macular degeneration, a leading cause of blindness in older adults. Pegaptanib was selected to bind to one of the four isoforms of the vascular endothelial growth factor protein (i.e., VEGF165) and prevents the excessive vascularization that is responsible for age-related macular degeneration. The aptamer drug is injected (0.3 mg per injection) directly into the vitreous humor of the eye every 6 weeks, or about 9 times a year. All four forms of VEGF have a receptor-binding domain (Fig. 6.6), while only VEGF165 has a heparin-binding domain, which is the specific target for pegaptanib. Pegaptanib was a new type of therapeutic agent with an unusual specificity that effectively suppressed age-related macular degeneration; to date, approximately 95% of the patients receiving pegaptanib are at least 65 years old. Patients treated with pegaptanib often continue to experience vision loss; however, their rate of vision decline is significantly lower than patients who do not receive this treatment.

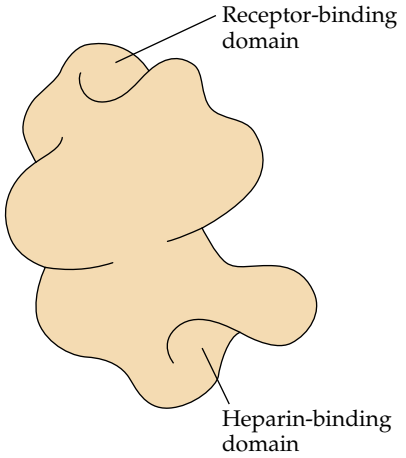


Figure 6.6 Schematic representation of protein VEGF, which contains both a receptor-binding domain and a heparin-binding domain.

The aptamer NOX-E36 is a pegylated RNA aptamer that is 40 nucleotides long and binds with a high affinity (K_d = 1.32 nM) to human chemokine CCL2. Human chemokines are a family of small (8 to 10 kilodaltons) cytokines that are secreted by cells and induce chemotaxis in nearby cells. CCL2 induces monocytes, memory T cells, and dendritic cells to the sites of inflammation. Importantly, when CCL2 is induced, it becomes involved in several inflammatory diseases and is also involved in cancer spread. Chemokine CCL2 is involved in the recruitment of tumor-associated macrophages, so that removing CCL2 and hence the tumor-associated macrophages from the tumor environment is expected to improve the immune response against the tumor. As of mid-2019, NOX-E36 had undergone successful phase I and phase II clinical trials with more than 100 patients.

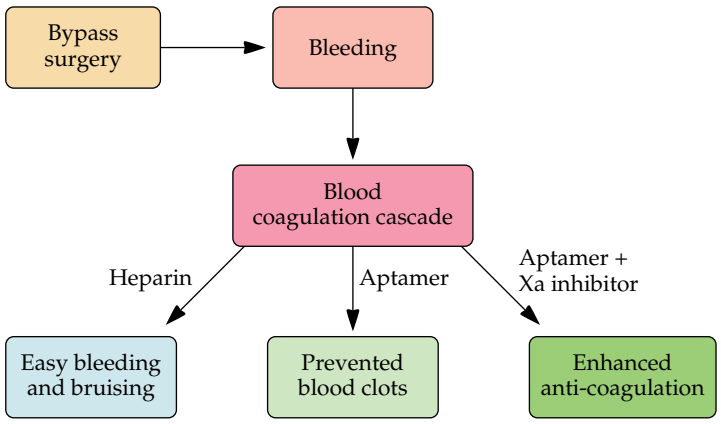
Prostate cancer is the second most common cause of cancer-related death for men in the United States. To treat this cancer, researchers sought a strategy to target anticancer drugs to local and metastatic tumor lesions. One recent approach included identifying an aptamer that bound specifically to tumor cells by first selecting aptamers that bound tightly to prostate cancer cells in culture and then from that pool of aptamers selecting those that did not bind to normal prostate cells in culture. The result of this positive/negative selection yielded an aptamer that bound to three different types of prostate cancer cells with a K_d of ~150 nM. Subsequently, the selected aptamer was conjugated to the drug monomethyl auristatin F, an antimitotic agent that inhibits cell division. This drug cannot be used on its own, as it is extremely toxic to both cancerous and normal cells, but has previously been coupled to a tumor-specific monoclonal antibody for delivery to tumorous cells. Coupling the drug to the aptamer specifically targets prostate cancer cells but not normal cells. To date, this aptamer–drug complex has been found to be highly effective in targeting prostate tumors *in vivo* and in inhibiting prostate tumor growth in mice.

In some instances, the safety of aptamers used in clinical trials is a concern, especially when the optimal dose of a particular aptamer is not known. One way to overcome this problem is through the use of aptamer

"antidotes." These molecules consist of short oligonucleotides whose sequences are complementary to the aptamers being tested. When antidotes are added to aptamers, they hybridize to the aptamers and inhibit their binding to the clinical target, thereby obviating concerns about too high a dose of the aptamer being used. One example is the application of antidotes during cardiopulmonary bypass surgery. This surgery presents a significant risk of postoperative bleeding and the subsequent formation of blood clots. To avoid blood clots, patients are typically administered heparin, a naturally occurring glycosaminoglycan that acts as an anticoagulant or blood thinner. Unfortunately, heparin can cause excessive bleeding and bruising and is contraindicated in persons with high blood pressure, liver disease, and hypertension. Protamine sulfate may be given to reverse the effects of too much heparin; however, its use may also be problematic. In one recent experiment (Fig. 6.7), researchers selected an aptamer that bound to blood coagulation factor Xa and found that this aptamer could also prevent blood clots from forming. In addition, by combining the factor Xa aptamer with a chemical inhibitor of factor Xa, the anticoagulation effect was enhanced. In both cases, the use of the aptamer directed against factor Xa avoided the unwanted side effects of heparin treatment. However, should it be necessary, an "antidote" to this aptamer could be applied. The results of this experiment suggest that drugs and aptamers with a shared target may be efficiently combined to exert more specific and effective results than when either of these agents is used by itself.

Until recently, it was not possible to select RNA aptamers that were able to distinguish a mutant protein from its wild-type counterpart that differed in a single amino acid. This is quite important since point mutations in various tumor suppressor proteins often cause cancer. In particular, protein p53 is a tumor suppressor that functions as a transcription factor that

Figure 6.7 Flow chart showing the prevention of blood clots and excessive bleeding following bypass surgery. Heparin, commonly used following bypass surgery to prevent blood clots, has a number of undesirable side effects. An aptamer that binds to blood coagulation factor Xa also prevents blood clots but without heparin's side effects. A complex of the factor Xa-binding aptamer and a chemical inhibitor of factor Xa is even more effective than the aptamer by itself at preventing blood clots without problematic side effects.



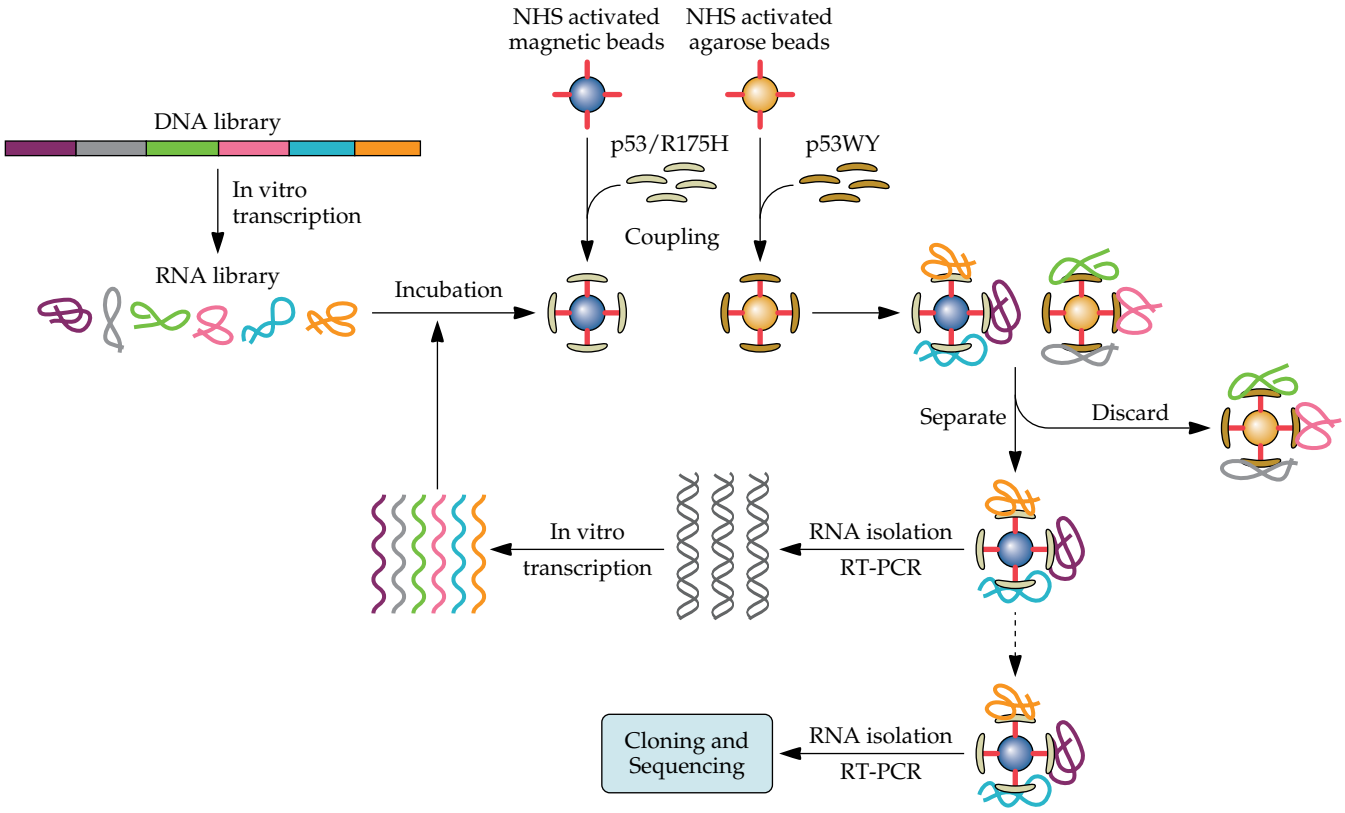


Figure 6.8 Overview of the screening procedure used to isolate a specific RNA aptamer that has activity specifically against a mutant protein (p53/R175H) with a single amino acid substitution. Proteins p53/R175H and native p53 (p53WT) were immobilized to NHS Mag Sepharose (magnetic beads designed for the coupling of aptamers and proteins through primary amino groups on these molecules) and NHS-Sephinose (agarose beads), respectively. A mixture of the two types of beads bound to the proteins was then incubated with an RNA oligomer library. Following this competitive binding of RNA aptamers to either bound p53/R175H or bound p53, the magnetic beads were separated from the agarose beads. The RNAs bound to the agarose beads were discarded, while the RNAs bound to the magnetic beads were released from the beads, converted into double-stranded DNA, and PCR amplified. Following *in vitro* transcription of the amplified DNA, the entire cycle was repeated several times. Following the last cycle, the DNA was cloned and sequenced.

regulates the expression of genes involved in DNA repair, cell cycle control, and apoptosis. In fact, it has been suggested that more than half of human cancers relate to a mutation in protein p53, with many of those mutations being single amino acid changes. With this in mind, researchers set about to develop an aptamer that could specifically target the mutant protein p53/R175H (in which an arginine residue at position 175 was replaced by a histidine residue) while not binding to the native p53 protein. This was done by utilizing a unique procedure that selectively enriched a mixture of aptamers until several aptamers that bound more tightly to the mutant than to the native form of protein p53 were isolated (Fig. 6.8). Importantly, one of the aptamers that was selected using this procedure effectively inhibited tumor formation in mice when the aptamer was complexed with nanoparticles and delivered intravenously.

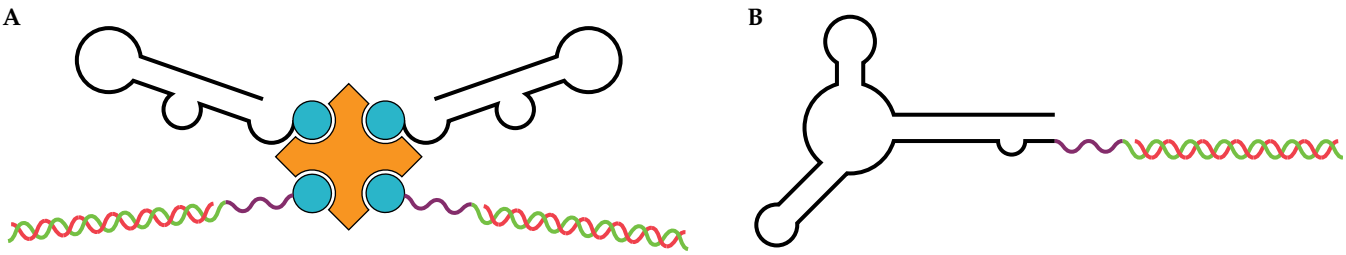


Figure 6.9 Aptamer-mediated targeted delivery of RNAi. **(A)** Example of a noncovalent aptamer-siRNA complex. Both the aptamer (upper molecule) and the siRNA (lower double-stranded molecule) are complexed with biotin (blue circles), with the biotin portion of these molecules binding tightly to streptavidin (large orange shape). **(B)** Example of a covalent aptamer-siRNA complex.

In addition to their therapeutic value, in which they bind to and inhibit a specific target, aptamers may also provide a means of delivering oligonucleotides and other small molecules in a cell-specific manner. In this case the aptamers are joined with other therapeutic nucleic acids or small molecules, typically by covalent linkage, although noncovalent linkage is also possible (Fig. 6.9). To date, researchers have reported linking aptamers with interfering RNAs (siRNAs; discussed later in this chapter), toxins, enzymes, and radionuclides. This strategy allows for the combination of two therapeutic approaches in a single chimeric molecule, where the aptamer directs the complex to the targeted cell type (typically a cancerous cell or a pathogen). In this regard, aptamers are considered to have some advantages over antibodies and antibody fragments as therapeutic reagents. Aptamers generally lack immunogenicity, whereas antibodies, unless they are completely human, are immunogenic. Several recent studies have linked various toxins and other molecules to an aptamer that targets prostate-specific membrane antigens (PSMA; a prostate cancer marker protein whose expression is increased in the membranes of prostate cells). For example, one group of researchers noncovalently linked an aptamer against PSMA with a therapeutic siRNA directed against lamin A/C (lamin A/C proteins, encoded by the LMNA gene, are major, ubiquitous components of nuclear laminae). The aptamer directs the complex to the membranes of prostate tumor cells while the therapeutic siRNA inhibits the activity of lamin A/C, where mutants of this protein have been associated with a variety of different diseases. In another recent study, an aptamer against PSMA was used to target a DNA-activated protein kinase to prostate cancer cells; this specifically increased the sensitivity of prostate cells to ionizing radiation by making those cells incapable of repairing the damage induced by radiation.

Ribozymes and DNAzymes

Ribozymes are naturally occurring catalytic RNA molecules (RNA metalloenzymes) that bind to and cleave RNA molecules. They are typically 40 to 50 nucleotides in length and have separate catalytic and substrate-binding domains. The substrate-binding sequence combines with its target RNA by nucleotide complementarity and, possibly, nonhydrogen-bond interactions. The catalytic portion cleaves the target

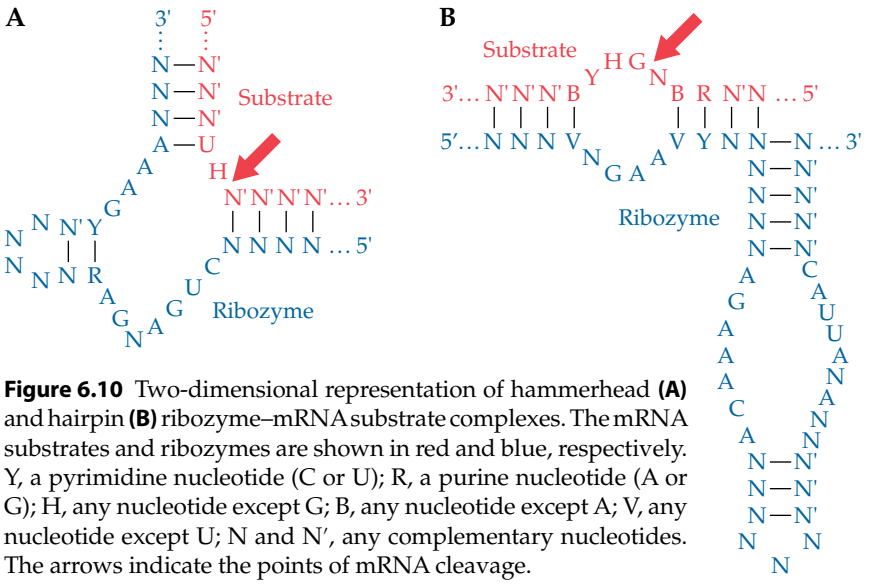


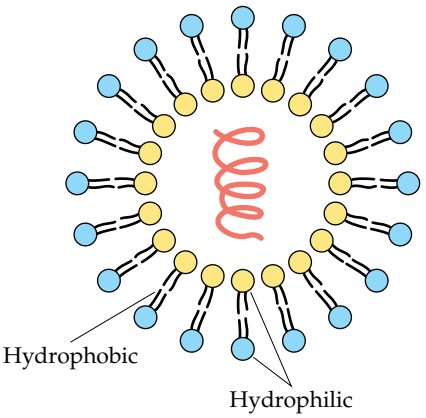
Figure 6.10 Two-dimensional representation of hammerhead **(A)** and hairpin **(B)** ribozyme–mRNA substrate complexes. The mRNA substrates and ribozymes are shown in red and blue, respectively. Y, a pyrimidine nucleotide (C or U); R, a purine nucleotide (A or G); H, any nucleotide except G; B, any nucleotide except A; V, any nucleotide except U; N and N′, any complementary nucleotides. The arrows indicate the points of mRNA cleavage.

RNA at a specific site. By altering the nucleotide sequences within the substrate-binding domain, a ribozyme can be engineered to specifically cleave any mRNA sequence (Fig. 6.10). For therapeutic purposes, either hammerhead or hairpin ribozymes—named after the appearance of their secondary structure that results from intrastrand base-pairing—may be used (Fig. 6.10). As is the case for most nucleic acid therapeutic agents, compared with protein therapeutics, an important advantage of ribozymes is that they are unlikely to evoke an immune response in a treated animal or human.

In practice, an indirect strategy is often used for creating a therapeutic ribozyme, since the large-scale production of synthetic RNA molecules is difficult and RNA molecules are susceptible to degradation after delivery to a target cell. One approach to overcome these drawbacks entails chemically synthesizing a double-stranded oligodeoxyribonucleotide encoding a ribozyme catalytic domain (~20 nucleotides) flanked by sequences that hybridize to the target mRNA after it is transcribed. The double-stranded form of the ribozyme oligodeoxyribonucleotide is cloned into a eukaryotic expression vector (usually a retrovirus). Cells are transfected with the construct, and the transcribed ribozyme cleaves the target mRNA, thereby suppressing the translation of the protein that is responsible for a disorder. To do this, target cells may be removed from a patient and then grown and transfected in culture (preventing the functioning of the problematic protein) before they are returned to the original tissue.

As an alternative to intracellular ribozyme expression, ribozymes may be delivered directly to cells by injection or in the interior of liposomes, spherical particles comprising a lipid bilayer surrounding a therapeutic agent such as nucleic acids or drugs (Fig. 6.11). Directly delivered ribozymes may be chemically modified to protect them from rapid breakdown by nucleases—the 2′ hydroxyl groups may be modified by alkylation or by substitution with either an amino group or a fluorine atom. These modifications typically increase the half-life of ribozymes in serum from minutes to days.

Figure 6.11 Schematic representation of a liposome containing a nucleic acid.



Under laboratory conditions, ribozymes can inhibit the expression of a variety of viral genes and, as a result, significantly inhibit the proliferation of numerous viruses. For example, in cell culture, ribozymes have been observed to inhibit the expression of (i) human cytomegalovirus (a common herpesvirus) transcriptional regulatory proteins, resulting in a 150-fold decrease in viral growth, (ii) human herpes simplex virus 1 (HSV-1) transcriptional activator, resulting in a reduction of around 1,000-fold in viral growth, and (iii) a reovirus mRNA encoding a protein required for viral proliferation. Thus, ribozymes that specifically target and cleave viral RNA could potentially be used to treat viral infections. They are also promising treatments for noninfectious diseases. For example, a hammerhead ribozyme was designed to treat collagen-induced arthritis in mice. A plasmid-encoded ribozyme directed against the mRNA for tumor necrosis factor alpha, which is involved in rheumatoid arthritis, was intravenously injected into affected mice. Following this treatment, there was a significant reduction in the level of the tumor necrosis factor alpha mRNA, as well as a decrease in collagen-induced arthritis. This type of ribozyme, used so far only in mice, has the potential to become a therapeutic agent for humans.

Sometimes, mutations in a viral gene that is a specific therapeutic target may render the particular therapy in question ineffectual because the ribozyme no longer binds tightly to this target sequence. However, a ribozyme can be designed to simultaneously bind to two or more different sites on a single viral mRNA (Fig. 6.12). In this way, a target mRNA may be cleaved at multiple sites, making it much less likely that any single viral mutation will confer resistance to the added therapeutic ribozyme.

Following the discovery of ribozymes, scientists speculated whether DNA might carry out some catalytic functions similar to what had been observed for RNA. Notwithstanding the fact that DNA does not contain a 2′ hydroxyl group on the sugar moiety (seemingly a key functional group), in the mid-1990s it was demonstrated that some short single-stranded DNAs could cleave RNA strands at specific sites. In fact, the kinetic parameters of the DNAzymes on RNA targets were superior to those of their RNA counterparts, and some DNAzymes are as efficient as protein enzymes in terms of catalytic efficiency (despite the fact that DNAzymes

Figure 6.12 Schematic representation of two different ribozymes bound to two different sites on the same mRNA with the subsequent cleavage of the mRNA at both sites. If the RNA sequence of one of the two ribozyme-binding sites is mutated so that one of the ribozymes no longer binds to that site and cleaves the mRNA, the mRNA will still be cleaved at the other site.

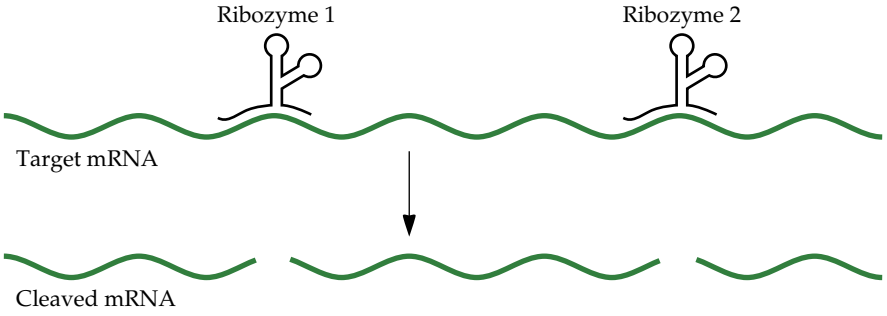


Table 6.4 Some examples of DNAzymes being tested in preclinical trials in mice

Target mRNA	Tumor type
c-*jun*	Basal cell carcinoma, squamous cell carcinoma, melanoma, breast cancer, prostate cancer, liposarcoma
Bcl-xL	Prostate cancer
Bcl-2	Prostate cancer
Egr-1	Breast carcinoma
VEGFR1	Nasopharyngeal carcinoma, melanoma
VEGFR2	Breast carcinoma
β1 integrin	Colon adenocarcinoma, prostate cancer
Aurora kinase A	Prostate cancer
EBV-LMP1	Nasopharyngeal carcinoma
Akt1	Nasopharyngeal carcinoma
HSP70	Breast carcinoma
MMP9	Breast carcinoma

require metal ions as cofactors). Other advantages of DNAzymes versus ribozymes are the greater ease of chemical synthesis, broad target recognition properties, and high catalytic turnover. Since their initial development, numerous applications of these catalytic molecules have been tested, in particular as anticancer therapeutics. Despite some encouraging results, with DNAzymes, like many other nucleic acid molecules being developed as therapeutic agents, delivery to target sites has been problematic. Building on the observed catalytic activity of DNAzymes, scientists have synthesized a number of different synthetic genetic (DNA-like) polymers that are able to both cleave and direct the synthesis of RNA molecules (Table 6.4). Of these unnatural genetic polymers, the most active compared to DNAzymes are 2′-fluoroarabino nucleic acids (in which fluorine atoms occupy the 2′ positions of the deoxyribose molecule). This provides scientists with yet another means of developing stable and highly effective nucleic acid enzymes that may eventually be used in a variety of therapeutic applications.

Recently, a DNAzyme that was complexed with a lipid carrier was injected directly into a mouse tumor that acts as a model of skin cancer. The DNAzyme targeted c-*jun* mRNA (where the c-jun protein is a component of a transcription factor whose overproduction is associated with several human cancers), significantly reducing levels of both c-*jun* mRNA and protein and profoundly inhibiting skin cancer cell proliferation and metastases (Fig. 6.13). Moreover, the injection of the DNAzyme–lipid complex triggered additional antitumor immune responses. These desired effects occurred in the absence of any nonspecific toxicities or side effects, a major advantage compared to some other nucleic acid-based therapeutic agents. Whether DNAzymes will turn out to be as effective as these early indications suggest, and be free of side effects in the treatment of human tumors, remains to be determined. Nevertheless, this DNAzyme is currently in phase I clinical trials.

Interfering RNA

The introduction of double-stranded RNA to animal or plant cells can reduce the expression of the protein encoded by a target RNA sequence which is complementary to the introduced double-stranded RNA. This "gene

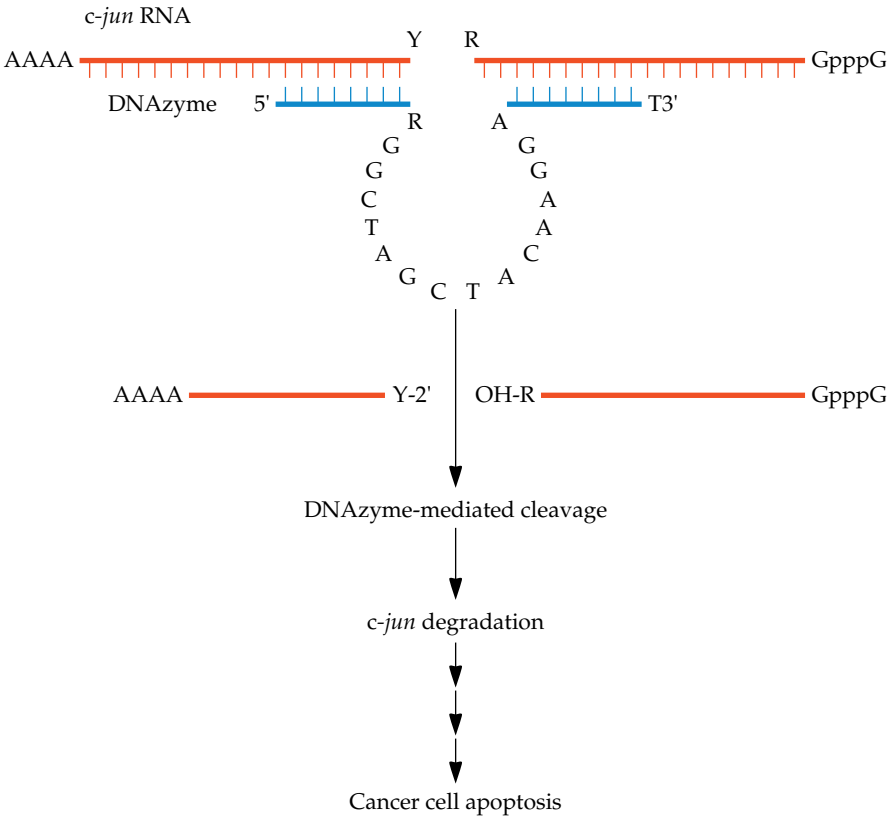


Figure 6.13 Schematic representation of a DNAzyme having just cleaved the c-*jun* mRNA. A purine (R = G or A) at the beginning of the DNAzyme catalytic core base-pairs with a pyrimidine (Y = C or U) in the target c-*jun* mRNA. Cleavage takes place between the pyrimidine and purine in the target sequence. Once cleaved, the mRNA is degraded by cellular ribonucleases. Inhibition of c-*jun* expression triggers a cascade of events eventually leading to apoptosis of the cancer cell.

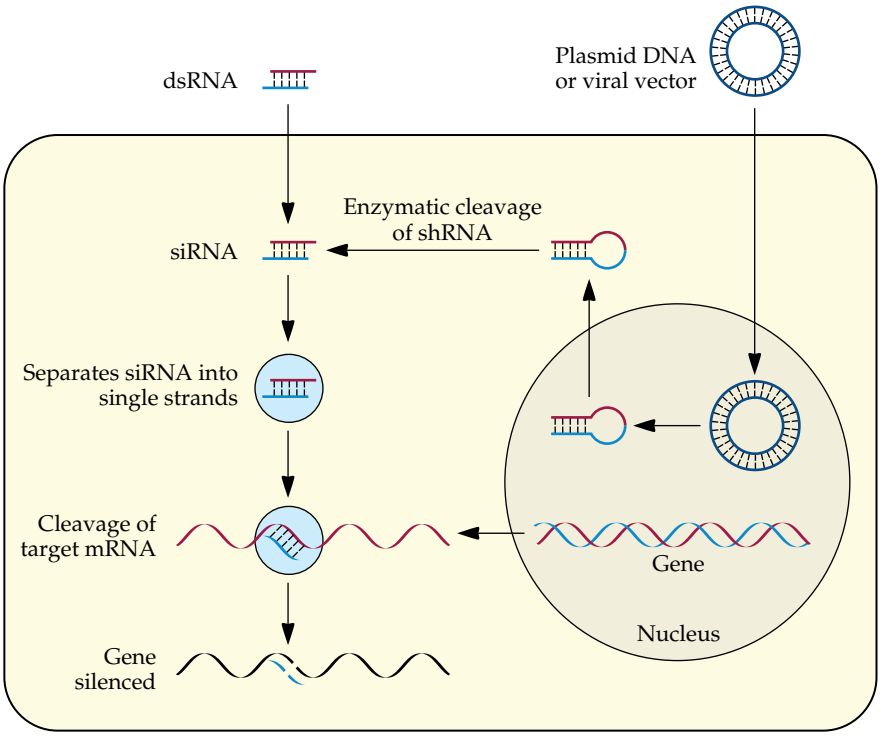
silencing," which can specifically reduce the concentration of a target mRNA by up to 90%, is reversible, since there is no change in the target cells' DNA content or composition. This phenomenon has been termed RNA interference (RNAi) and occurs naturally in virtually all eukaryotic organisms. Although its various biological roles remain to be completely elaborated, the available evidence indicates that RNAi acts to protect both animals and plants from viruses and from the accumulation of transposons (a transposon is a DNA sequence capable of inserting itself into a new position within the genome). A working model for RNAi has been formulated based on experimental analyses (Fig. 6.14). Following the introduction of a double-stranded RNA molecule into a cell, the double-stranded RNA is cleaved by the RNase III-like enzyme Dicer into single-stranded pieces of RNA, approximately 21 to 23 nucleotides in length. The antisense strand of an siRNA is incorporated into an RNA-induced silencing complex (RISC) that includes several other proteins, including the enzyme Argonaute, and directs it to a target mRNA. The specific binding of the siRNA to the mRNA is based on the complementarity of the two RNA sequences. RISC cleaves the targeted mRNA between nucleotides 10 and 11 relative to the 5′ end of the siRNA (antisense) guide

strand. Consistent with this model, the transfection of mammalian cells in culture with duplexes of 21-nucleotide RNA can also mediate RNAi activity. As shown in Fig. 6.14, the interfering RNA may also be encoded on a plasmid or viral vector and expressed as a single-stranded short hairpin RNA (shRNA) that folds back on itself into a hairpin-like structure. Viral vectors that integrate into the chromosomal DNA are generally used when persistent long-term knockdown of gene expression is desired. The introduced RNA duplexes and expressed shRNAs are processed by Dicer to form siRNAs. To improve the efficiency of specific mRNA cleavage, it is possible to introduce both the siRNA and the Argonaute protein at the same time. One way of doing this is for both the siRNA and the Argonaute protein to be encoded in the plasmid DNA as shown in Fig. 6.14.

Despite the fact that many aspects of RNAi are still not completely understood, this technique could form the basis for new therapeutic agents. The use of short RNA duplexes or plasmids encoding short hairpin RNAs, or shRNAs, that may be processed to form siRNAs, may eventually provide an alternative approach to gene silencing through the use of antisense oligonucleotides, aptamers, or ribozymes.

The phenomenon of RNAi is expected to facilitate the development of a wide range of therapies utilizing specific siRNAs delivered to the appropriate target cell. The expression of endogenous eukaryotic genes or viral RNA

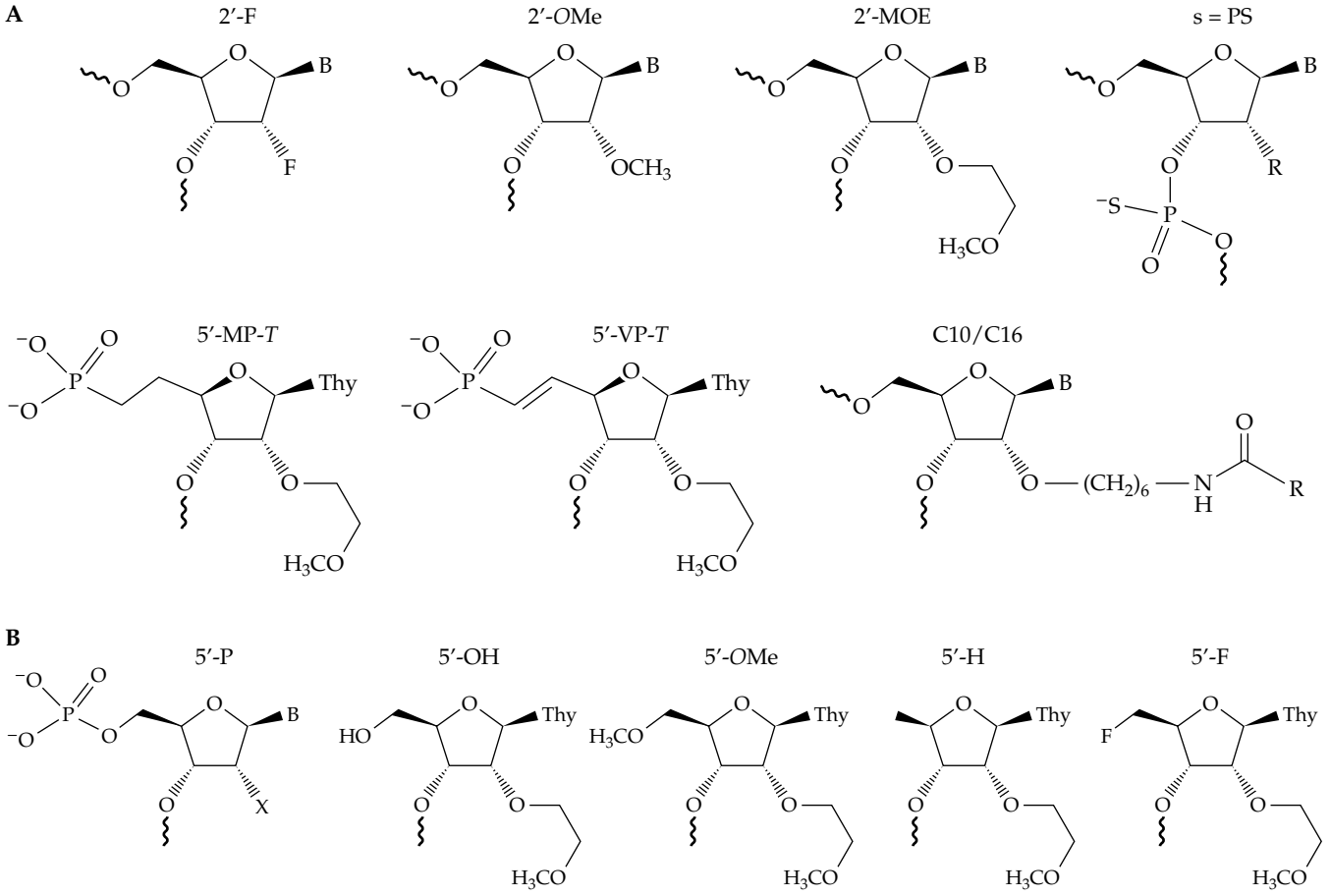
Figure 6.14 Schematic representation of the silencing of the expression of a specific gene by either siRNA (small interfering RNA) or shRNA (short hairpin RNA) which is cleaved to siRNA by the enzyme Dicer. The target gene in the nucleus is labeled "Gene." The target mRNA is bound by a single complementary strand, following the separation of the siRNA into single strands, and is then cleaved by RISC. The small RNAs in RISC guide the Argonaute protein to the sequences complementary to the mRNA targeted for cleavage. dsRNA, double-stranded RNA.



in infected cells may be inhibited by the siRNAs. In fact, there are a number of reports of the use of either duplex RNA or shRNA to suppress virus replication in tissue culture. Moreover, RNAi is effective *in vivo* (in mice), suggesting that most if not all viral infections may be inactivated by RNAi.

The introduced siRNA or shRNA may have nonspecific effects. For example, siRNA or shRNA may also be complementary to nontarget mRNAs. Fortunately, several experimental approaches may be utilized to lessen or avoid these problems. (i) By lowering the concentration of siRNA or shRNA as much as possible (often to 20 nM or less), off-target effects are generally avoided. (ii) Since the interferon response can be induced by double-stranded RNAs with as few as 11 base pairs (bp) that are perfectly complementary, siRNA or shRNA is designed to contain at least a 1-nucleotide bulge (where the bases on opposing strands are noncomplementary) near the center of the molecule (typically 21 bp). (iii) Since siRNA or shRNA can

Figure 6.15 Chemical structures of some of the modified nucleotides used in the synthesis of single-stranded antisense RNA that may be used instead of double-stranded siRNAs. **(A)** Phosphorothioate (PS), 2′-fluororibose (2′-F), 2′-methoxyribose (2′-OMe), 2′-methoxyethylribose (2′-MOE), 5′-methylenephosphonate (5′-MP), 5′-(E)-vinyl-phosphonate (5′-VP), C10 (TC10), and C16 (TC16). **(B)** 5′-Terminal modifications include 5′ phosphate (5′-P), 5′ hydroxyl (5′-OH), 5′ methoxy (5′-OMe), 5′ deoxy (5′-H), and 5′ fluoro (5′-F).



exert a toxic effect when it contains the nucleotide sequence 5′-UGGC-3′, this sequence is specifically avoided. (iv) Blunt-ended 27-bp RNA duplexes or 29-bp shRNAs with 2 nucleotides overhanging at the 3′ end are much more potent inducers of RNAi than 21-mer siRNAs. Using these slightly longer RNAs at low concentrations avoids side reactions associated with 21-mer siRNAs.

Currently, the therapeutic usefulness of siRNA is limited by the complex mechanisms used to deliver these molecules to the tissues where they are required. It would be advantageous if single-stranded RNA molecules could be developed that could activate the RNAi pathway. This would appear to be feasible since (i) single-stranded antisense oligonucleotides do not require special formulations to distribute to peripheral tissues and (ii) human Ago2 (the Argonaute protein that is part of the RISC complex) binds only to short single-stranded RNAs. Since the RNAi mechanism requires the removal of the sense strand in order for the antisense strand of the siRNA to bind the mRNA target, scientists tested a number of modified nucleotides (Fig. 6.15) in an effort to enhance the chemical stability of single-stranded RNA that might be used in place of siRNAs. Eventually, scientists were able to produce single-stranded antisense RNAs with enhanced *in vitro* potencies of approximately 20-fold and the metabolic stability to achieve a tissue half-life of 10 days in mice. These data indicate that single-stranded antisense RNAs are advantageous compared to double-stranded siRNAs. Thus, the use of a single strand removes the risk that the other strand or its metabolites might cause undesirable off-target effects.

Huntington's disease, an incurable neurodegenerative disorder that affects muscle coordination and eventually leads to mental decline and behavioral problems, is caused by mutant huntingtin protein (normally 347 kDa in molecular mass). The disease symptoms are typically manifest in middle age, with death occurring around 20 years following disease diagnosis. Huntington's disease is caused by a dominant heterozygous expansion of CAG trinucleotide repeats within the protein-encoding region of the huntingtin gene, where CAG is a codon for the amino acid glutamine, and normal individuals typically contain no more than 35 consecutive CAG trinucleotides. The average mutant huntingtin gene in Huntington's disease patients contains around 45 consecutive CAG trinucleotides (Fig. 6.16); the additional glutamines cause the protein to aggregate in the brain. By synthesizing and testing a number of different single-stranded siRNAs with modified nucleotides, researchers found that it was possible to inhibit production of mutant huntingtin protein (in patient-derived fibroblasts) with an IC_{50} value (i.e., how much of the drug is needed to inhibit a given biological process by half) of 0.9 nM and an allele selectivity of >100-fold, demonstrating the potential to achieve allele selectivity in cell lines within the median range of CAG repeat copy number. In subsequent experiments in mice, single-stranded siRNAs were continuously infused into the right lateral ventricle of the brain for 28 days. This long treatment time was required because of the long *in vivo* half-life of the huntingtin protein and the need to monitor protein levels (the disease causes a small but measurable increase in the protein molecular mass), rather than RNA, to ensure that reduced levels of huntingtin synthesis could be detected. This is the first experiment in which it was successfully demonstrated that single-stranded siRNAs are able to inhibit mutant huntingtin protein expression *in vivo*, demonstrating allele-specific gene knockdown. Notwithstanding

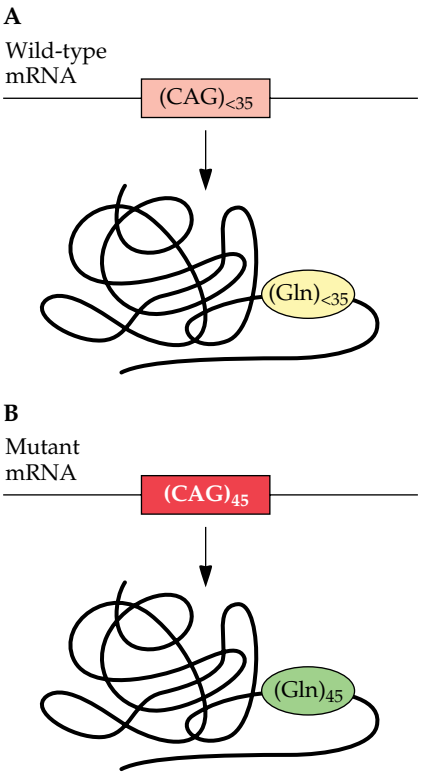


Figure 6.16 Schematic representation of **(A)** wild-type huntingtin mRNA with <35 tandem copies of the CAG trinucleotide codon being translated into native protein and **(B)** mutant huntingtin mRNA with 45 tandem copies of CAG trinucleotide codon being translated into a nonfunctional protein containing 45 tandem glutamine residues. The individuals with the mutant gene that produces this mRNA have Huntington's disease.

the fact that this strategy is still at an early stage of development so that it appears to be rather complex, this result suggests that single-stranded siRNAs may represent a whole new class of nucleic acid therapeutic agents for treating some diseases for which no effective treatment currently exists. Unfortunately, for siRNAs to be effective in the long term, repeated treatments are required. Moreover, this approach often results in off-target effects that can cause liver and kidney toxicity. On the other hand, a single

Table 6.5 Some clinical trials of siRNA-based drugs[a]

Drug	Target
AGN211745	Age-related macular degeneration
Atu027	Solid tumors
APN401	Melanoma, pancreatic cancer, renal cell cancer
ALN-AS1	Acute intermittent porphyria
ALN-CC5	Atypical hemolytic uremic syndrome
ALN-PCS02	Elevated LDL cholesterol
ALN-PCSSC	Renal impairment, hypercholesterolemia, atherosclerotic cardiovascular disease
ALN-GO1	Primary hyperoxaluria type 1
ALN-RSV01	Respiratory syncytial virus
ALN-TTR02	Transthyretin-mediated amyloidosis
ALN-TTRSC	Transthyretin-mediated amyloidosis
ALN-VSP02	Solid tumors
ARC-520	Hepatitis B
AT3SC	Hemophilia
BCL-11a	Sickle cell disease
Bevasiranib	Macular degeneration
BMT101	Hypertrophic scar
CALAA-01	Solid tumors
DCR-MYC	Solid tumors, multiple myeloma, non-Hodgkin's lymphoma
DCR-PH1	Primary hyperoxaluria type 1
DCR-PHXC	Primary hyperoxaluria
DCR-HBVS	Hepatitis B
KRASG12D	Pancreatic cancer
ND-L02-s0201	Hepatic fibrosis
OLX10010	Hypertrophic cicatrix
PF-04523655	Age-related macular degeneration, diabetic retinopathy
QPI-1002	Acute renal failure
siG12D-LODER	Pancreatic ductal adenocarcinoma
EphA2-DOPC	Advanced cancers
SYL1001	Dry eye disease
SYL040012	Glaucoma
STP705	Hypertrophic scar
TKM-080301	Adrenocortical carcinoma
TKM-100201	Ebola virus
TD101	Pachyonychia congenita
VIR-2218	Hepatitis B

[a]Note that not all of these trials have been successful. The drug designation is in part an abbreviation of the name of the company that developed the drug.

dose of an shRNA that is delivered using an adeno-associated virus vector (discussed later in this chapter) appears to be much longer lasting with fewer side effects.

Interfering RNAs have already found widespread use as tools in research that is directed toward understanding how gene expression is regulated in natural systems. One company that specializes in producing RNAi directed against human mRNAs advertises, "For each target (human) gene, we provide four plasmids each with a different short hairpin RNAi sequence (shRNA). Our experimentally verified design algorithm minimizes the risk of off target effects and ensures the maximum knock-down. At least one of the four shRNA plasmids will reduce the target mRNA levels in the transfected cells by >70%." Thus, with the ready availability of human shRNA libraries, new insights and understanding of many fundamental and disease processes should be rapidly forthcoming. It is hoped that this, in turn, will lead to a variety of new therapeutic agents and approaches.

There are an increasing number of reports of the *in vivo* effectiveness of RNAi. shRNAs against a wide range of proteins, viruses, and diseases have been successfully expressed in mice, including siRNAs directed against HSV-2, hepatitis B virus, hepatitis C virus, Huntington's disease, metastatic Ewing sarcoma (a form of cancer), respiratory syncytial virus, hepatic cancer, transforming growth factor receptor 2, severe acute respiratory syndrome, heme oxygenase 1, keratinocyte-derived chemokine, tumor necrosis factor alpha, and human epidermal growth factor receptor 2. In addition, by mid-2019, 34 different RNAi therapeutics were being tested in clinical trials, and in August 2018, the first RNAi-based product was approved by the U.S. FDA (Table 6.5; Box 6.2). Fifteen of these compounds were in phase I trials, ten in phase II trials, and nine in phase III trials.

Approval of the First RNAi Drug

box 6.2

Hereditary transthyretin amyloidosis (also called familial amyloid polyneuropathy or Corino de Andrade's disease) is an autosomal dominant neurodegenerative disease (where the disease-causing allele is dominant over the normal recessive allele). This relatively rare disease is characterized by a buildup of abnormal amyloid protein in peripheral nerves, causing pain, paresthesia (abnormal skin sensation), muscular weakness, and autonomic nervous system dysfunction, all of which are typically observed at ~20 to 40 years of age. This disease is caused by a mutation of the transthyretin protein that transports thyroxine and retinol in the serum and cerebrospinal fluid. It is endemic to a few towns in northern Portugal and elsewhere and is thought to affect around 50,000 people worldwide. Following diagnosis of this disease, in the absence of a liver transplant, it is invariably fatal within around 10 years.

In August 2018, the U.S. FDA approved the use of patisiran, an siRNA for the treatment of hereditary transthyretin amyloidosis. This followed successful clinical trials that included 225 patients, 148 of whom received the drug and 77 of whom received a placebo. The drug, which works by shutting down the production of transthyretin, was administered once every 3 weeks for 1.5 years. Upon its approval, patisiran received an Orphan Drug designation; this provides financial incentives to encourage the development of drugs for rare diseases. It was developed by Alnylam, a Cambridge, Massachusetts-based company, and treatment cost approximately $450,000 U.S. per year. The drug is not without a number of side effects that largely stem from the drug infusion protocol, which includes steroid pretreatment. Notwithstanding the side effects of this drug as well as its very high cost, it works and, as the first RNAi drug to be approved, it is expected that this success will make it easier for other RNAi drugs to gain U.S. FDA approval.

Zinc Finger Nucleases

Instead of treating a genetic disease with a therapeutic agent or providing an additional "corrective" gene, it might be advantageous to directly repair a cell's defective genes. This has become possible using either of two recently developed techniques, zinc finger nucleases or the CRISPR-Cas system. With zinc finger nucleases, the nucleases bind to a specific gene and cleave its DNA. The cell then repairs the broken DNA strands using copies of all or part of a replacement gene that is supplied. In the case of gene therapy, the copies lack the disease-causing mutation that is found in the original gene. Repair of double-strand breaks in the DNA of animal cells can occur by either nonhomologous end-joining or homologous recombination. With nonhomologous end-joining, insertions or deletions may result, leading to disruption of the target gene. On the other hand, homologous recombination between the endogenous target locus and an exogenously introduced homologous DNA fragment (the donor DNA) allows for the possibility of targeted gene alteration. This is because the presence of a double-strand break within the target locus stimulates recombination with the exogenous donor DNA by several orders of magnitude. In this instance, the sequence of the donor DNA determines the outcome of the gene targeting event. Thus, the donor DNA can be designed to either create or correct a mutation in a specific gene locus. In addition, the donor DNA can be synthesized to contain either an entire expression cassette or a complementary DNA fragment of the gene to be corrected.

Zinc finger nucleases are engineered proteins that contain two domains—a zinc finger domain that binds to a target DNA sequence and a nuclease domain that cleaves the DNA. Zinc fingers are small protein structural domains that each bind a single molecule of Zn^{2+} that stabilizes its structure. They are often found on proteins that bind DNA, RNA, proteins, or small molecules. They may bind to DNA in a sequence-specific manner by inserting a protein α-helical region into the major groove of the DNA double helix, with each zinc finger interacting with a specific DNA triplet sequence. Moreover, since the zinc fingers bind to the DNA independently of one another, they can be linked together in a peptide by genetic engineering, so that they will bind to a predetermined sequence on a DNA fragment. To create a protein that will cut DNA at a target site, the zinc fingers are expressed as a fusion protein with the nonspecific nuclease FokI from the bacterium *Flavobacterium okeanokoites* (Fig. 6.17). The DNA-binding zinc finger domain directs the nonspecific FokI cleavage domain to a specific DNA target site. The FokI cleavage domain is enzymatically active only as a dimer, so the creation of a specific nuclease that makes double-stranded breaks is dependent upon dimerization of zinc finger nucleases. Therefore, two zinc finger nuclease subunits are typically designed to recognize the target sequence in a tail-to-tail conformation, with each monomer binding to half-sites that are separated by a spacer sequence (Fig. 6.17).

To date, the possibility of using zinc fingers to correct a wide range of genetic mutations has been limited by the lack of simple procedures for producing and selecting specific zinc fingers. Nevertheless, scientists are optimistic that these procedures will soon be developed, and the promise of this approach will be realized. This technology has been used to disrupt genes in cultured Chinese hamster ovary cells and in fruit flies, zebra fish, and rats; to correct genes in fruit flies and cultured human, mouse, and plant cells; and to add DNA to genes in cultured human, plant, and

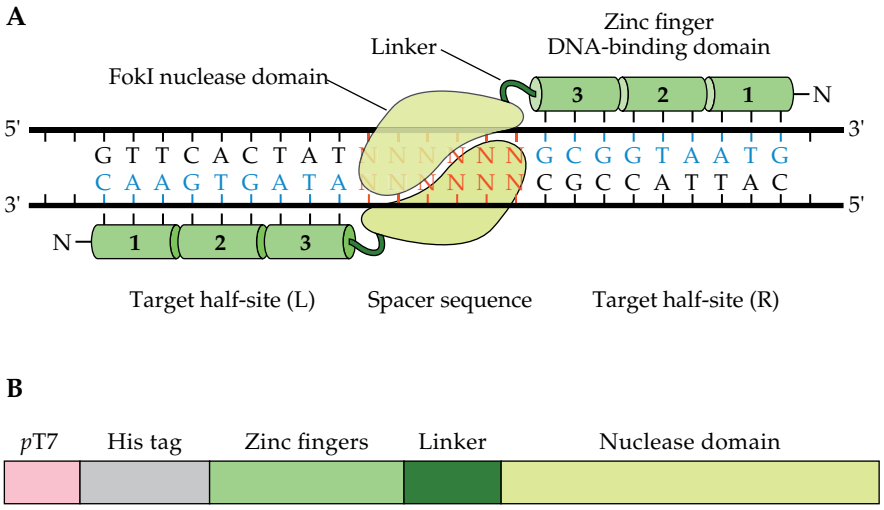


Figure 6.17 **(A)** Schematic representation of the binding of a zinc finger nuclease to DNA targeted to be cleaved. **(B)** Genetic construct used to produce a zinc finger nuclease in *E. coli* under the control of the T7 promoter. The histidine tag facilitates the purification of this protein.

mouse cells. By mid-2020 more than a dozen U.S. patents describing the development and use of zinc fingers had been granted.

The first clinical study (phase I/II) is using zinc finger nucleases to correct a gene involved in Hunter's syndrome (mucopolysaccharidosis II), which is a rare genetic disorder in which the glycosaminoglycans accumulate in body tissues. This study began in 2018 and is expected to last through 2022, when it will be decided whether phase III trials are warranted. In this case, a DNA construct encoding the zinc finger nuclease was delivered to patients using adeno-associated virus (discussed later in this chapter).

CRISPR-Cas System

In recent years scientists have developed a number of different methodologies that may be used to repair a cell's defective genes. Arguably, the most promising of the existing targeted genome editing methodologies (and also the newest) exploits the bacterial clustered regularly interspaced short palindromic repeats (CRISPR)-Cas system (see chapter 2). The simplest form of this technology consists of a single guide RNA molecule (sgRNA), where part of the molecule matches the target DNA sequence, and an endonuclease from the bacterium *Streptococcocus pyogenes* (Cas9). The sgRNA and Cas9 form a complex where the sgRNA guides the complex to its target DNA, and the nuclease activity of Cas9 then cleaves the target DNA on both strands before the cell repairs the cleaved DNA by either nonhomologous end-joining or homologous recombination, which are described in a preceding section.

Notwithstanding the enormous benefit to human health of antibiotics, the growing antibiotic resistance of many bacterial strains threatens the effective treatment of many bacterial infections. This problem is worldwide, and it is getting worse as new antibiotic resistance mechanisms and strains continue to develop. Thus, there is an acute need to develop

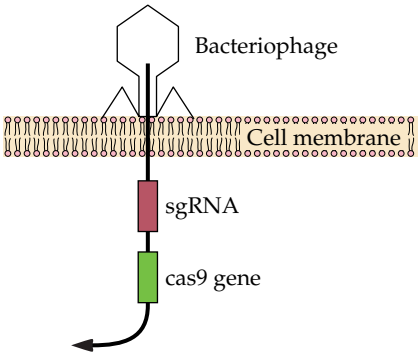


Figure 6.18 Schematic representation of a bacteriophage injecting its phagemid DNA, which contains an sgRNA-*cas* construct, into a target bacterial cell. The sgRNA-*cas* DNA construct is transcribed, and the *cas* mRNA is translated and assembled and then used to cleave the bacterial chromosomal DNA.

both new antibiotics as well as approaches that overcome or prevent the development of antibiotic resistance. In this regard, one group of researchers recently reported using sgRNA-Cas nucleases to produce several sequence-specific antimicrobials. In these experiments, sgRNA-*cas* nucleases were programmed to kill virulent, but not avirulent, *Staphylococcus aureus* (a cause of skin, intestinal, and other infections) and/or to target and destroy staphylococcal plasmids carrying antibiotic resistance genes. In order to do this effectively, the sgRNA-Cas system has to be delivered to all of the target bacterial cells in a population. One way to do this is to package the DNA encoding the sgRNA-Cas system into a bacteriophage (Fig. 6.18). The engineered bacteriophage can subsequently inject its DNA into the host bacteria. The vector that was used to deliver the *cas9* gene and its RNA guide sequences was a phagemid, which is a plasmid that contains a bacteriophage origin of replication; it can be replicated as a plasmid and also be packaged into viral particles. One advantage of using this system is that in order to kill all of the target bacteria (including the small fraction that might otherwise escape this treatment), it is possible to program the nuclease with two or more sgRNA guides to target different bacterial pathogen chromosomal locations. In addition, this strategy is also effective at cleaving plasmid-encoded antibiotic resistance genes, thereby preventing the proliferation of antibiotic resistance. Finally, this particular sgRNA-Cas nuclease system was found to be effective at killing *S. aureus in vivo* in mice. Although this is clearly a preliminary report, the data suggest that this approach may be useful in the treatment of a wide range of human bacterial infections.

In late 2018, a biophysicist in China announced that using a modified CRISPR-Cas system, he had engineered mutations into human embryos. Moreover, these embryos were subsequently implanted into human mothers. This human germline gene therapy was done to disable the human *CCR5* gene, which encodes a protein that allows HIV to enter cells. The idea behind this unprecedented and unsanctioned experiment was to mimic a mutation that exists in around 10% of persons of European heritage and appears to protect them from HIV infection. In addition to the many ethical concerns that this work engendered, it has been suggested that the procedure also could have made "off-target" changes to the genomes of the two twin girls who were born following this treatment. In December 2019, the researcher who conducted this experiment was found guilty by a court in China of conducting "illegal medical practices" and subsequently sentenced to 3 years in jail. Two collaborating scientists also received jail sentences.

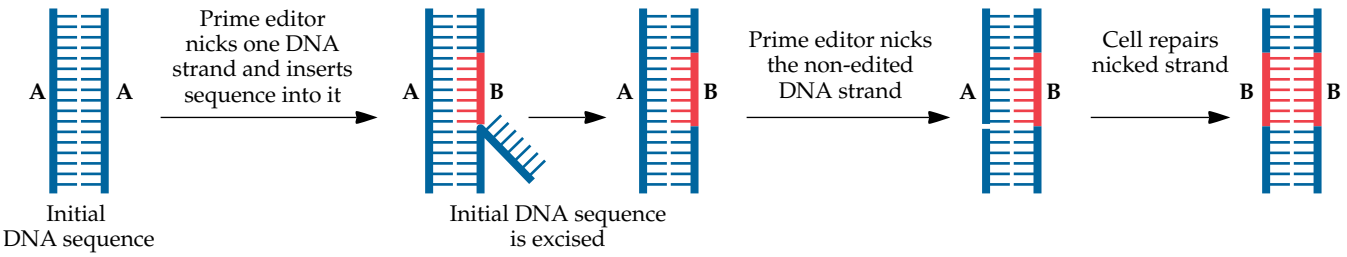
The CRISPR-Cas9 system breaks both strands of the target DNA at the same time, and then the cell's repair system patches the cut DNA and makes the edits. As a consequence of nonhomologous end-joining, several nucleotides may be inserted or deleted at the site in the genome where cleavage occurred. The uncertain behavior of the cell's repair system makes it difficult for scientists to depend upon the CRISPR-Cas9 system to overwrite an aberrant DNA sequence with a corrective DNA sequence, especially when attempting to correct an aberrant human gene. Several groups of scientists have sought to develop variants of the CRISPR-Cas9 system that would yield higher-fidelity changes, for example, by utilizing unique sources of Cas9 such as *Francisella novicida* instead of the more commonly used enzyme from *S. pyogenes*. On the other hand, in late 2019, one group

of researchers reported a unique solution to this problem. They developed a system called "prime editing," which uses a modified version of the Cas9 enzyme that initially nicks only one strand of the double-stranded DNA at the target sequence (Fig. 6.19). Then, the enzyme reverse transcriptase, guided by an sgRNA, is used to excise the single strand of incorrect DNA. The prime editor then nicks the nonedited DNA strand, and the cell repair system replaces the sequence on the nicked strand with the correct DNA. It has been suggested that in its present form, this technique is likely to be limited to changing only one or two nucleotides at a time. Nevertheless, if this technique is effective with a large number of different types of cells, it may be used to correct a number of (human) mutations where only a single base change has occurred.

Nanozymes

Instead of depending on cellular enzymes to bind to and cleave mRNA in a RISC complex, one group of scientists has synthesized a nanoparticle that can cleave specific cellular mRNAs. This synthetic, RISC-like molecule has been called a nanozyme (Fig. 6.20). This nanozyme consists of gold nanoparticles that are approximately 13 nm in diameter complexed to RNase A and a DNA oligonucleotide complementary to the target mRNA sequence. RNase A was used because it readily degrades the target mRNA in a non-sequence-specific manner but does not degrade the DNA component of the nanozyme. The DNA oligonucleotide includes a 9-nucleotide (poly dA) spacer linked on one end to the gold nanoparticle and on the other end to an 18-deoxynucleotide-long recognition sequence that is complementary to a portion of the target mRNA. Following binding of the target mRNA to the nanozyme, the target mRNA is cleaved by the bound RNase A. When this system was tested, it was found to be highly active in both tissue culture with human liver cells and *in vivo* in mice as a treatment for hepatitis C virus infection. In mice, the nanozymes were delivered by subcutaneous injection. The DNA component of the nanozyme was synthesized to target the 5′ nontranslated region in the hepatitis virus RNA genome, which is conserved among the six major genotypes of the virus. This region of the genome controls the initiation of viral mRNA translation, and cleaving this region prevents the mRNA from being translated and the virus from being replicated. Subject to further testing, and eventually to

Figure 6.19 Schematic representation of the technique of "prime editing," which avoids unintended changes to a genome by inserting the desired changes to the genome into the DNA itself instead of depending upon the cell's repair system. X indicates the nucleotide(s) that require changing. Y indicates the corrected nucleotide(s).



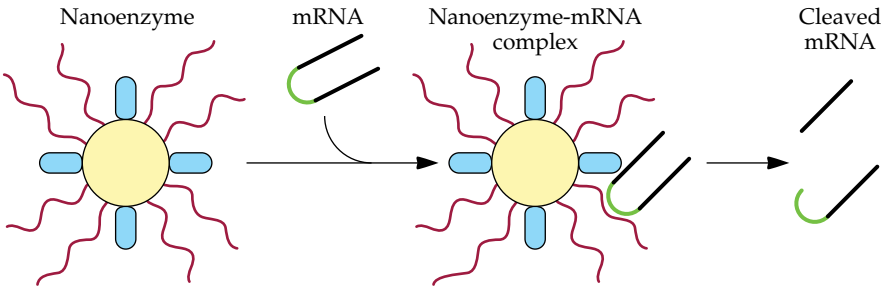


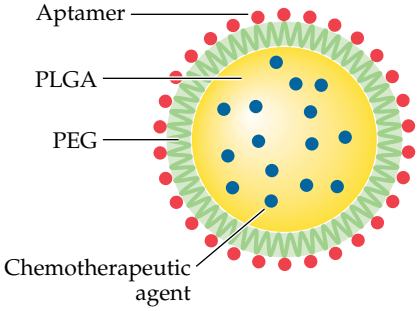
Figure 6.20 Schematic representation illustrating nanozyme cleavage of a targeted mRNA. The nanozyme consists of a gold nanoparticle (shown in yellow) with bound molecules of RNase A (shown in blue) and DNA oligomers (shown in red). Following the binding of the target mRNA to complementary sequences on the DNA oligomers, the mRNA is cleaved by the RNase A. The portion of the mRNA that is complementary to the bound DNA oligomers is shown in green.

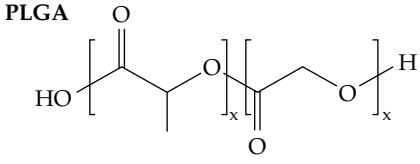
human clinical trials, it has been suggested that this approach might be therapeutically simpler and more effective than the use of siRNA.

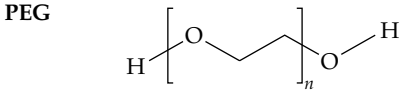
Nanoparticles

While some nanoparticles provide an effective means of delivering various chemotherapeutic agents to cells, they do not always target a specific cell type. To overcome this drawback, scientists have derivatized the surfaces of some nanoparticles to contain aptamers that are very cell-specific in their binding (Fig. 6.21). For example, for testing in rat models, researchers devised a nanoparticle that contained paclitaxel (an anticancer chemical agent) and had a surface aptamer directed against the protein nucleolin to treat an intracranial (inside the skull) tumor. The aptamer efficiently facilitated the internalization of the complex into the rat glioma (brain tumor) cells. Antinucleolin aptamers have also been conjugated to pegylated lipid-PLGA [poly(D,L lactic-co-glycolic acid)] hybrid nanoparticles containing paclitaxel to target tumor cells that overexpressed nucleolin receptors. This promising approach is still at a very early stage of development.

Figure 6.21 Schematic representation of a nanoparticle that consists of PEG (polyethylene glycol) and PLGA [poly(D,L lactic-co-glycolic acid)] carrying a chemotherapeutic agent (shown in blue) and having a cell-targeting aptamer (shown in red) on its surface. The structures of PLGA and polyethylene glycol are shown.







Engineering Bacteriophages

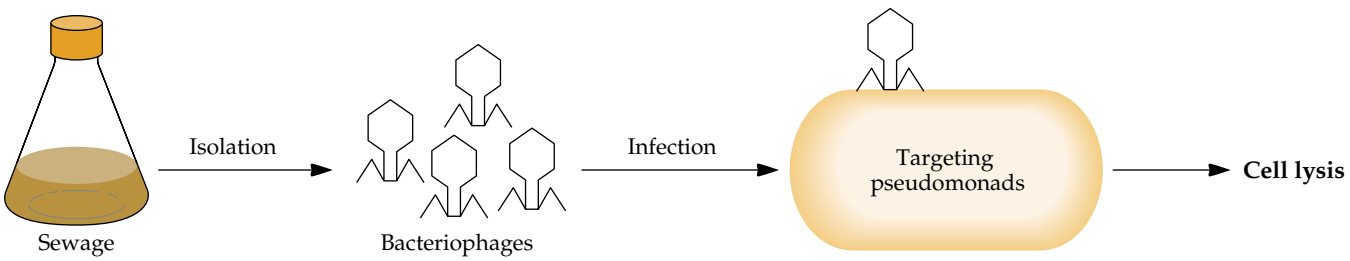
Bacteriophages were discovered independently by Frederick Twort in 1915 and by Félix d'Hérelle in 1917. Bacteriophages were initially used as antimicrobial agents, with the first published report of bacteriophages being used to treat human infections occurring in 1921. However, with the discovery of antibiotics such as penicillin in 1928 and streptomycin in 1943 and their subsequent widespread commercialization beginning after the Second World War, a paradigm shift in public health began worldwide. With the wide-scale availability of antibiotics, many previously devastating and deadly bacterial diseases were soon reduced to relatively minor inconveniences, and millions of deaths due to bacterial infections have been prevented. Unfortunately, this situation did not last for long. Within the past 10 to 20 years, an increasing number of antibiotic-resistant infections have developed and have created real problems for the global

medical community. There is some indication that antibiotic-resistant bacterial strains are now emerging at a faster pace than new antibiotics are being discovered. The economic impact of multidrug-resistant infections is difficult to quantify, but it was estimated in 2018 to be between $20 billion and $35 billion per year in the United States. Should this trend continue, it is possible that the health of a large number of individuals may be put at risk. For example, a World Health Organization report that was released in April 2014 stated, "This serious threat is no longer a prediction for the future, it is happening right now in every region of the world and has the potential to affect anyone, of any age, in any country. Antibiotic resistance—when bacteria change so antibiotics no longer work in people who need them to treat infections—is now a major threat to public health."

The widespread clinical and agricultural use of antibiotics has played a significant role in the emergence of antibiotic-resistant strains of bacteria. Although the development of vaccines against different pathogenic bacteria would avoid antibiotic use, this is not a simple or inexpensive task. Therefore, scientists have sought to develop a variety of innovative approaches to deal with this problem. One approach includes utilizing lytic bacteriophages directed against various multidrug-resistant bacteria. With this strategy, the bacteriophage inject their DNA into and eventually lyse bacterial cells.

Cystic fibrosis is an inherited genetic disorder that chronically affects the lungs of individuals worldwide. The lungs of cystic fibrosis patients become filled with mucus and are subsequently colonized with *Pseudomonas aeruginosa*, a situation that damages the epithelial surface and results in a decline in lung function. While treatment with DNaseI temporarily relieves some of the blockage in the lungs (see chapter 5), it is essential to develop longer-term and more effective means of removing *P. aeruginosa* from the lungs of cystic fibrosis patients. One possible approach to this problem includes selecting bacteriophage (viruses that infect bacteria) that specifically target and lyse various strains of *Pseudomonas* spp. (Fig. 6.22). Previous studies with bacteriophage showed that they could efficiently remove from 50 to 99% of *P. aeruginosa* from the surfaces of catheters and microtiter plates. Initially, two different previously unknown bacteriophages were isolated from sewage and then characterized; bacteriophages that target *P. aeruginosa* have been isolated all over the world, with about half of them being isolated from sewage. These two bacteriophages infect and lyse *Pseudomonas* spp., and this bacteriophage mixture cleared the bacteria from the lungs of treated mice. The

Figure 6.22 Schematic representation of the isolation and use of bacteriophages that bind to and lyse pathogenic strains of *Pseudomonas* spp. This approach may eventually replace or augment the extensive use of antibiotics to kill these pathogens.



pathogen, *P. aeruginosa*, was labeled with *lux* (a bacterial gene that encodes a bioluminescence-producing protein) so that, by monitoring the bioluminescent light produced by the pathogen, researchers were able to sensitively measure the lysis of the pathogen *in vivo*. The bacteriophage mixture was found to be highly effective in killing *P. aeruginosa* growing either *in vitro* in cell culture in the lab or in a mouse lung, even when it was contained within a biofilm. Following treatment with the bacteriophage mixture, the pathogen numbers were reduced by 3 to 4 orders of magnitude. In addition to lysing a large fraction of the pathogenic bacteria, treatment with bacteriophages also made the remaining bacteria more susceptible to antibiotics. In this type of treatment, the use of a mixture of at least two different bacteriophages is advantageous over the use of a single bacteriophage because it significantly reduces the probability of the emergence of a population of *P. aeruginosa* that is resistant to lysis by a particular bacteriophage (which might occur if the bacterial surface receptor targeted by that bacteriophage was mutated). With the success of this experiment and the demonstrated absence of sequences encoding proteins that are virulent in humans in the bacteriophage genomes, it remains to be determined whether this approach is as effective in humans as in mice. To date, bacteriophage therapy has been found to be effective with mice, dogs, catfish, and wax moth larvae. In one recent study, researchers used bacteriophage therapy to treat *P. aeruginosa* infections in a cystic fibrosis zebra fish model. Zebra fish were chosen for these experiments because the structure of the membrane protein chloride channel that is encoded by the cystic fibrosis transmembrane conductance regulator (CFTR) gene responsible for cystic fibrosis is similar in zebra fish and humans. In the experiments with zebra fish, phage therapy decreased lethality, bacterial burden, and the inflammatory response caused by the bacterial infection. Importantly, this experiment represented the first report of bacteriophage therapy being used to cure *P. aeruginosa* infections in a cystic fibrosis animal model. Recent results of human clinical trials using bacteriophage directed against *P. aeruginosa* have been encouraging.

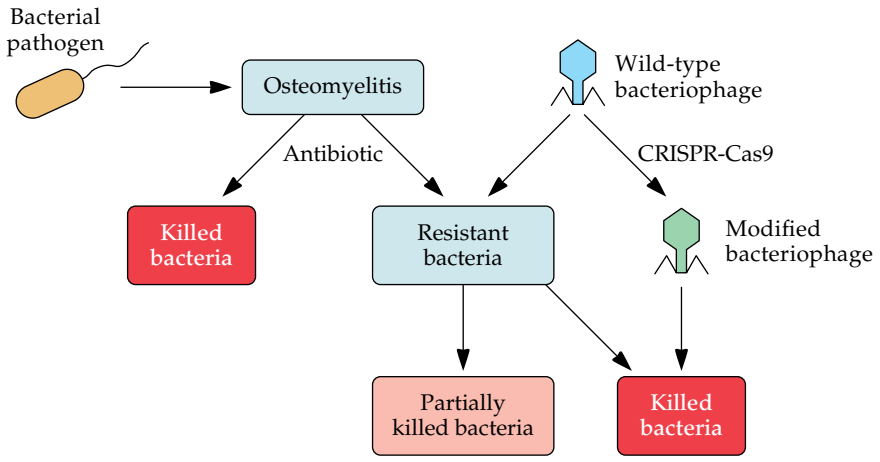
Mycobacterium spp. are Gram-positive non-spore-forming bacteria that include around 140 different species and are considered to be environmental opportunistic pathogens. *Mycobacterium abscessus* includes a group of rapidly growing, multidrug-resistant strains that are the cause of a number of skin and soft tissue diseases, central nervous system infections, blood poisoning, and other infections. Patients with cystic fibrosis often have *M. abscessus* infections. These infections are not treatable with antibiotics. In one particular case in 2019, a 15-year-old cystic fibrosis patient infected with *M. abscessus* and not responsive to antibiotic treatment was treated with a three-bacteriophage cocktail following a lung transplant. The bacteriophages that were used were developed by genetic engineering of naturally occurring bacteriophage. The bacteriophages, which were administered intravenously, were well tolerated by the patient and resulted in improved liver function (a by-product of the disease) and a decreased level of skin infections. Not only was this the first reported use of bacteriophage for mycobacterial infection, but it was also the first reported clinical use of engineered bacteriophage. However, whether this treatment is an effective long-term solution for this patient and others with similar infections remains to be determined.

Osteomyelitis is a rare (affects ~70,000 people per year in the United States) but serious bone infection that is often caused by the pathogenic

bacterium *S. aureus* (a common Gram-positive organism). The infection typically affects the bone either by traveling through the bloodstream or from infections of the surrounding tissue and is generally treated by long-term administration of antibiotics, although many strains of *S. aureus* are resistant to antibiotics. Prior to the discovery of antibiotics, this type of infection often resulted in the loss of a limb and/or death. Although antibiotics are currently quite effective at clearing this infection, more and more bacterial strains are becoming antibiotic resistant, including many strains of *S. aureus*. Given that diabetic foot ulcers (the source of many *S. aureus* infections) occur in ~25% of diabetic patients and that the incidence of diabetes is expected to reach 55 million people in the United States by 2030, osteomyelitis is expected to be a significant health care concern in the not too distant future. In early 2019, clinical trials of bacteriophages (delivered intravenously) began in the United States for the treatment of drug-resistant bacterial infections. Although bacteriophage effectively kill *S. aureus*, one group of researchers modified the bacteriophage host range and its bactericidal activity (by altering the gene for the external bacteriophage tail fiber protein). This modified bacteriophage was found to be considerably more effective (bactericidal) than the native unmodified bacteriophage (Fig. 6.23). The modified bacteriophage was also highly effective at mitigating the infection of human soft tissues by a strain of biofilm-forming *S. aureus*. Finally, these researchers developed an *in vivo* rat model of osteomyelitis and demonstrated that the modified bacteriophage could readily be delivered to rat tissues and could subsequently reduce *in vivo S. aureus* infections of soft tissues. If this modified bacteriophage is eventually used as a human therapeutic, it is likely that it will be administered together with another bacteriophage that also targets *S. aureus* to avoid resistance.

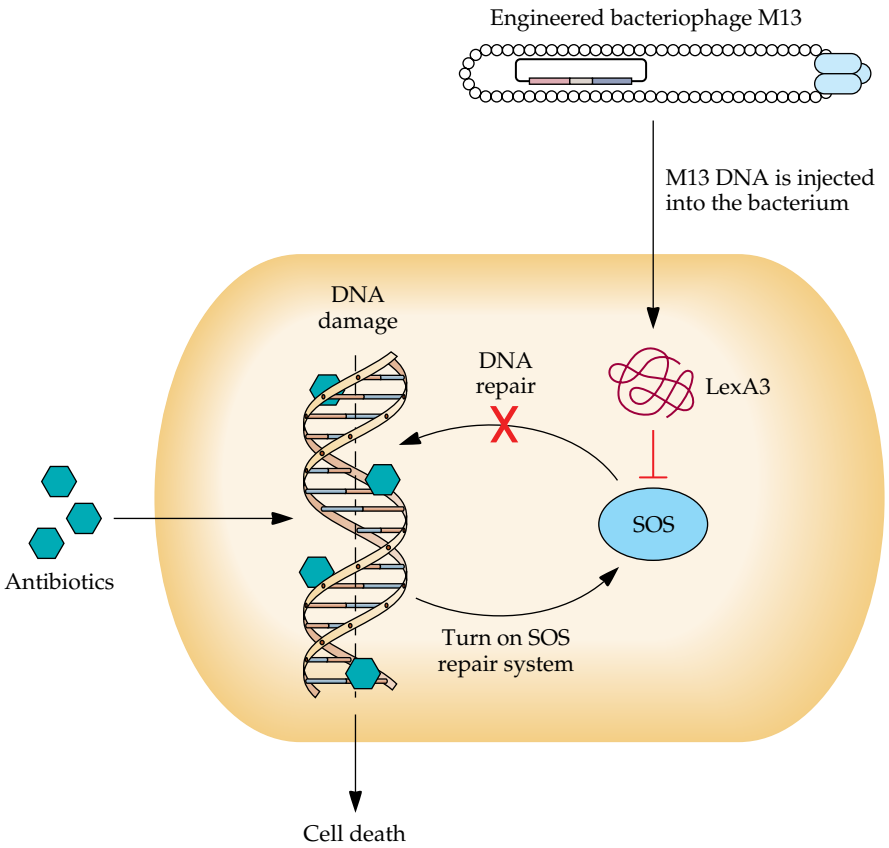
Just as bacteria can develop resistance to antibiotics, they can also develop resistance to infection and lysis by bacteriophages. To avoid the development of bacteriophage-resistant bacteria, a recent and innovative approach, still in an early stage of development, utilizes a nonlytic bacteriophage to deliver a protein that increases bacterial susceptibility to

Figure 6.23 Development of a modified bacteriophage that is more effective at killing pathogenic bacteria (that cause osteomyelitis) than the starting wild-type bacteriophage.



specific antibiotics (Fig. 6.24). The LexA3 protein represses the bacterial SOS response, a global response to DNA damage in which DNA repair is induced. The *Escherichia coli lexA3* gene was placed under the control of a strong inducible promoter and spliced into double-stranded bacteriophage M13. Following synthesis of LexA3 inside a phage-transformed bacterium, the SOS system is constitutively repressed and therefore no longer able to repair DNA that is damaged by added quinolone antibiotics. In fact, the LexA3 protein enhances the bactericidal activity of certain antibiotics by several orders of magnitude. This approach works well for a number of different antibiotics (including gentamicin and ampicillin) and not just quinolones. Moreover, this approach can enhance the killing of bacteria that have already acquired antibiotic resistance as well as the killing of wild-type bacteria. When the engineered bacteriophage system was tested in mice, it was found to significantly increase their survival after they were infected with a pathogenic strain of *E. coli*. Researchers are now developing engineered bacteriophage strains that have bacterial targets other than the SOS repair system, so one might eventually have engineered bacteriophages with multiple mechanisms to debilitate pathogenic bacteria.

Figure 6.24 Use of an engineered bacteriophage as an adjunct to antibiotic therapy. Bacteriophage M13 was engineered to introduce the *lexA3* gene, as part of the phage genome, into bacteria that infect an animal host. Upon induction, large amounts of the LexA3 protein are produced in the bacteria, which prevents the bacterial SOS system from repairing DNA that has been damaged by a quinolone antibiotic. The antibiotic is therefore more effective at killing the bacteria.



In addition to the engineered bacteriophage described here, researchers are currently developing a large number of other genetically engineered bacteriophage.

To make bacteriophage more effective for use in antibacterial therapy, bacteriophage previously designed to be targeted to specific bacteria (by engineering them to express a protein on the bacteriophage surface that bound tightly to the target pathogen) were conjugated with gold nanorods (Fig. 6.25). The complex was added to animal cells infected with the target pathogen as well as nontarget bacteria. The mixture was then irradiated with infrared light, which heated the gold nanorods and caused the bacteriophage–nanorod complex to heat and thereby kill the target bacteria. The localized heating also destroyed the bacteriophages. In testing this system, bacteriophage were engineered to target several different human pathogens, including *E. coli*, *P. aeruginosa*, and *Vibrio cholerae*. Moreover, in addition to killing pathogen cells in suspension culture, this procedure also effectively killed pathogenic bacteria that were part of a biofilm (which might have protected the majority of those cells from being killed by antibiotics). This technique still needs to be tested in animals and then, if successful, in humans. Nevertheless, this is an exciting new approach that may facilitate the use of bacteriophage to kill antibiotic-resistant bacterial pathogens.

Interestingly, in one recent study it was observed that in some cases when mutations arise in a bacterial surface structure so that the bacteria become resistant to bacteriophages to which they were previously sensitive, they simultaneously become more sensitive to different antibiotics. The authors of this study suggest that in these instances there is "an evolutionary trade-off between phage resistance and antibiotic resistance." It is not clear whether this relationship holds for a large number of pathogenic bacteria or is limited to only certain bacterial strains. Nevertheless, this observation suggests that in some instances a combination of bacteriophage and antibiotic treatment may prove to be a highly effective treatment for pathogenic bacteria.

The long-term effectiveness of modified bacteriophage remains to be determined. However, a note of caution is in order here, as bacteriophages have not yet gained wide acceptance in clinical practice. Thus, many of the first applications of this technology are likely to be in the treatment of diseased animals.

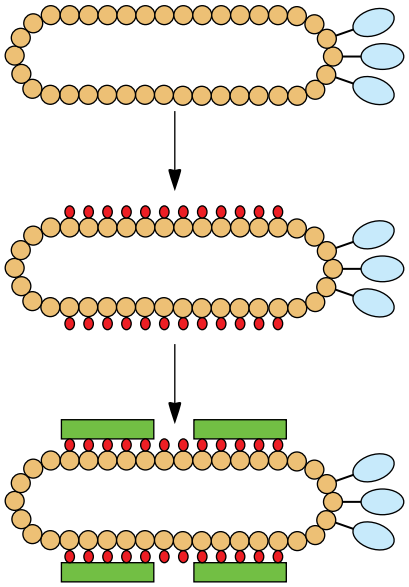


Figure 6.25 The filamentous bacteriophage M13 was engineered to express a targeting protein (shown in blue) that binds tightly to the outer surface of a specific pathogenic bacterium and not to nontarget bacteria. Thiol groups (shown in red) were introduced by chemical modification along the surface of the bacteriophage coat protein (shown in tan). Gold nanorods (shown in green) were then conjugated to the thiol groups. The modified bacteriophage–nanorod complex was then added to the pathogenic bacteria, where the targeting protein bound the bacteriophage to the pathogen. Finally, the complex was exposed to near infrared light, which caused the gold nanorods to exhibit a specific resonance and become heated, with the localized elevated temperature killing both the targeted cells and the modified bacteriophage.

Viral Delivery Systems

The major limitation in introducing corrective human genes as well as different types of nucleic acids to prevent the functioning of various pathogenic agents and tumor cells is the requirement for a simple, reliable, and safe delivery system for those therapeutic molecules. Systemic introduction of a therapeutic agent often leads to the nonspecific accumulation of very high levels in tissues where the agent is not required and may result in serious side effects. Consequently, viral vectors that deliver small nucleic acids to specific cellular targets have been developed. Several different viral vectors have been used to introduce exogenous nucleic acids into target cells. These include (but are not limited to) gammaretroviruses, lentiviruses, adenoviruses, herpes simplex virus I, and adeno-associated viruses (Table 6.6). Although virus-based gene delivery systems have been

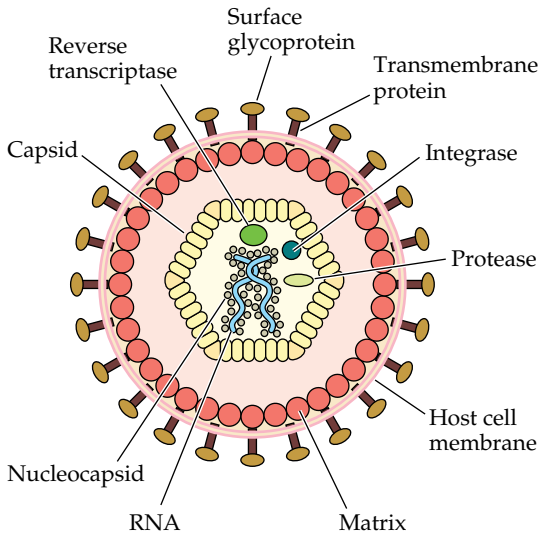
Table 6.6 Some of the viruses that are used to deliver foreign DNA into human cells

Virus	Genome	Genome size (kb)	Capsid	Coat	Virion diameter (nm)	Packing capacity (kb)
Gammaretrovirus	ssRNA (+)	8.3	Icosahedral	Enveloped	100	8
Lentivirus	ssRNA (+)	3–9	Icosahedral	Enveloped	120–130	8
Adenovirus	dsDNA	36	Icosahedral	Naked	90–100	7.5
Herpes simplex virus	dsDNA	152	Icosahedral	Enveloped	150	>30
Adeno-associated virus	ssDNA	4.6	Icosahedral	Naked	20	5

somewhat successful, a number of safety concerns still remain with regard to the use of these vectors. As a result, several nonviral methods are being developed as an alternative to virus-based systems.

The first viral delivery system to be developed was based on a mouse RNA virus, Moloney murine leukemia virus, which is a gammaretrovirus of the family *Retroviridae*. Some of the features that favored a gammaretrovirus as a therapeutic gene delivery system included the small size of the genome (8,332 nucleotides), which could be manipulated readily with recombinant DNA techniques, and the relative ease of growing the virus in cell culture. In addition, as part of its life cycle, it integrates into host chromosomal DNA, which suggested that long-term gene expression of an introduced therapeutic gene was possible. The mature gammaretrovirus virion is a complex structure about 100 nm in diameter (Fig. 6.26). The outermost lipid bilayer that envelops the virion is derived from the host cell. All other components are viral in origin. Embedded within the cell membrane are protrusions (spikes) composed of two proteins (surface glycoprotein and transmembrane protein). The inner surface of the lipid bilayer is coated with the matrix protein. Separated from the matrix layer, there is a protein shell made up of the capsid protein that encases the core. The core contains capsid proteins (nucleocapsid) bound to the viral

Figure 6.26 Schematic representation of a mature gammaretrovirus virion.



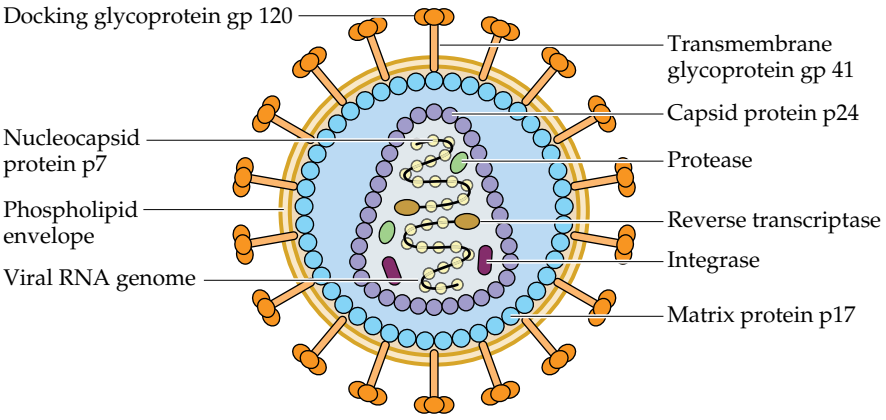
genome and the enzymes reverse transcriptase, integrase, and protease. To convert this and other animal viruses into effective gene delivery systems, the viral genome is genetically engineered to include a therapeutic gene, which upon infection is released into the cell for expression of the therapeutic protein.

Another commonly employed viral delivery system includes lentiviruses, also a member of the *Retroviridae* family. Unlike gammaretroviruses, they infect and integrate into nondividing cells. In addition, lentiviruses have long incubation times, suggesting that stable therapeutic gene expression over a long period is likely. Moreover, lentiviruses rarely induce an immunological response in a host organism. Despite the overall similarities between gammaretroviruses and lentiviruses, there are differences. Lentiviruses are about 20% larger, their capsid is a rounded trapezoid, and they have a number of genes that are not present in gammaretroviruses that encode accessory and regulatory proteins that participate in various aspects of the life cycle (Fig. 6.27).

The most thoroughly studied lentivirus at the molecular level is human immunodeficiency virus type 1 (HIV-1), which is responsible for acquired immunodeficiency syndrome (AIDS). This disease has caused millions of deaths, and currently, millions of people worldwide are HIV-1 positive. Therefore, why would HIV-1, which is a dangerous human virus, be considered as a therapeutic gene viral delivery system? The primary reason is that nondividing cells such as neurons and differentiated cells can be targeted for gene therapy using lentivirus vectors. Development of AIDS when using lentiviruses as a therapeutic gene delivery system is not a major concern because all of the viral genes are removed from the therapeutic gene vector construct and do not produce any active HIV-1 virions.

A major problem with the use of lentivirus vectors is that they lack sufficient specificity, with the lentivirus envelope proteins binding ubiquitously to proteins present on the surface of most human cells, leading to significant off-target effects. To increase the specificity of lentivirus vectors, in one study, researchers created bispecific antibodies (see chapter 5) that could bind to both the surface of a lentivirus envelope and the targeted human cells. In this case, the normal binding domain on the lentivirus was mutated so as to minimize off-target binding. These manipulations

Figure 6.27 Schematic representation of a lentivirus virion.

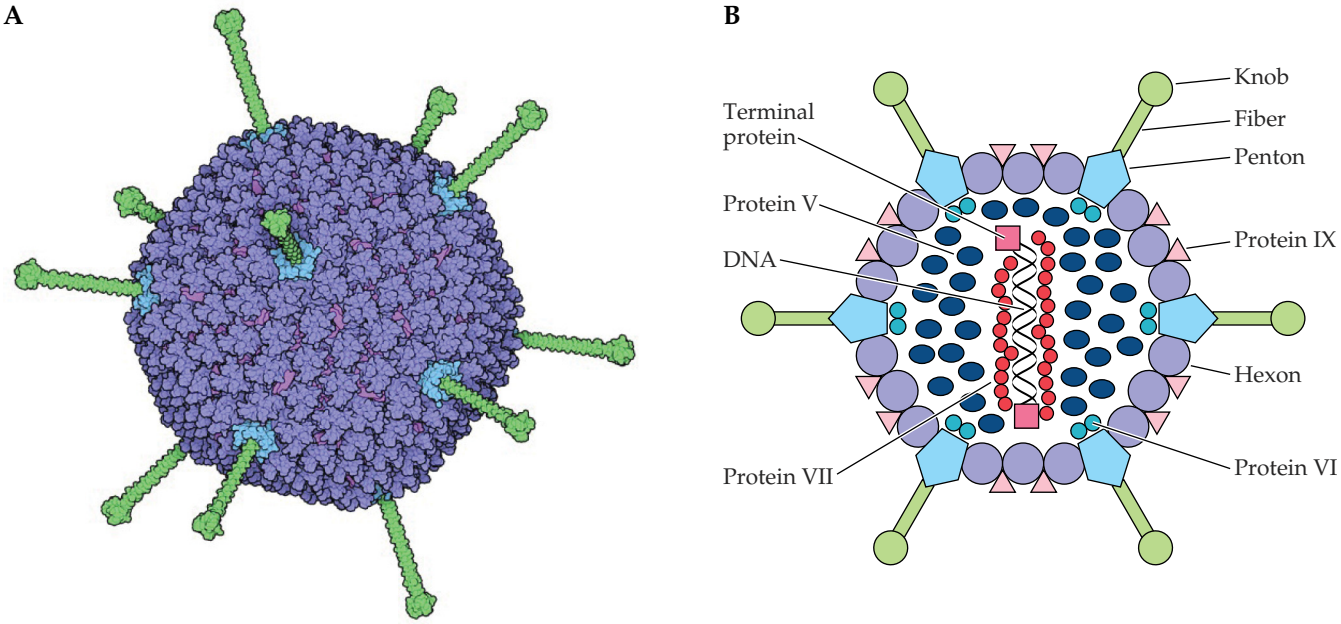


significantly increased the selectivity of the lentivirus vector. This work is still at a very early stage of development and remains to be tested in animal models.

Adenoviruses are nonenveloped viruses (~100-nm-diameter particles) that infect the upper respiratory tract, with usually mild consequences (i.e., common cold-like symptoms). The capsid has 20-sided symmetry with protruding knobbed fibers anchored to a penton protein at the junction of the individual units of the capsid (Fig. 6.28). The most abundant protein of the virion capsid has been designated hexon (Fig. 6.28). The core carries proteins that bind to the genome, which is a linear, double-stranded DNA molecule that is about 36 kb in length.

Adenoviruses (and adeno-associated viruses) enter a cell and reach the nuclear membrane in similar ways. The knob of the fiber protein of adenovirus makes contact with a cell surface receptor, and then the penton protein binds to surface proteins. Importantly, the knob component can be genetically engineered to target an adenovirus vector to specific cell surface receptors. After binding, the virion is enclosed in a clathrin-coated vesicle that fuses with an endosome, and the virus is routed to the nucleus. At a nuclear pore, the released virus is disassembled, and adenoviral DNA (or a genetically engineered variant of this DNA) enters the nucleus. Adenovirus is well suited as a therapeutic gene vector because it infects a large number of different dividing and nondividing cells. Pathogenicity from this virus is mild, and there is no concern about insertional mutagenesis because adenovirus does not insert into host chromosomes.

Figure 6.28 Structure and proteins of adenovirus. **(A)** Three-dimensional representation of adenovirus. Image source: David S. Goodsell/RCSB Protein Data Bank [http://pdb101.rcsb.org/motm/132]), under license CC BY 4.0. **(B)** Schematic representation of adenovirus. Labels on the right indicate the major capsid proteins, and those on the left mark some of the core proteins.

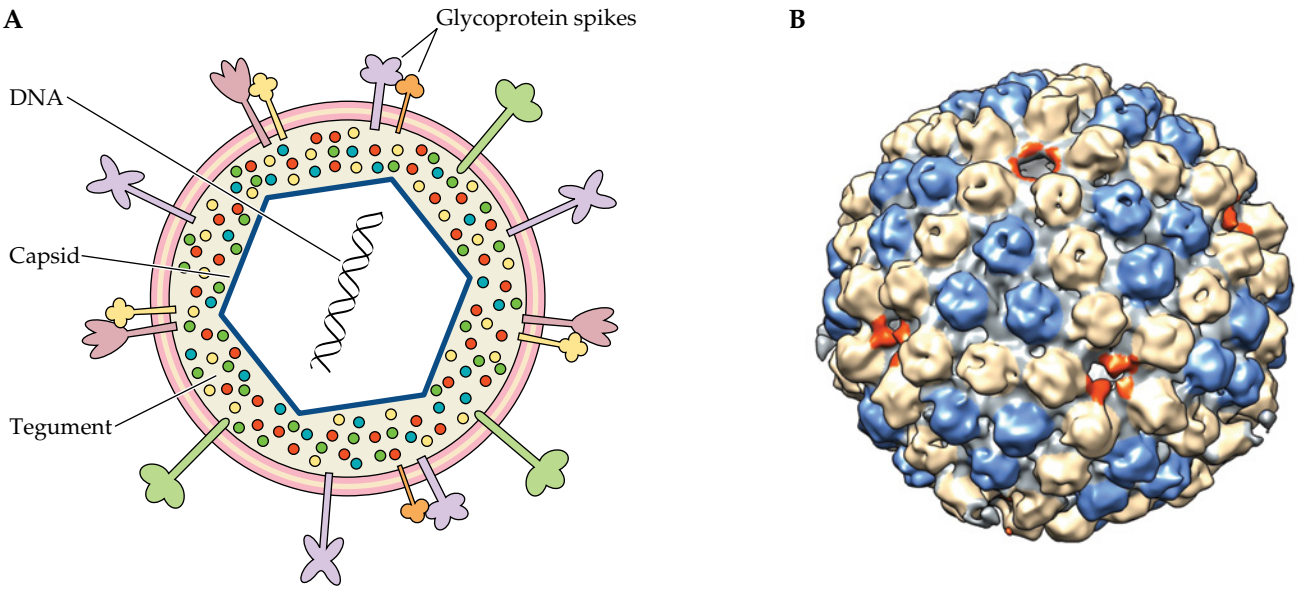


Herpes simplex virus 1 (HSV-1) is a common human pathogen that causes periodic cold sores; it infects and persists within nondividing neurons. The virus usually remains quiescent in the neurons, although the viral production cycle can be initiated by stress or hormonal changes. There are many different disorders that affect the central and peripheral nervous systems, including tumors, metabolic and immunological defects, and neurodegenerative syndromes. Because of its preference for neurons, HSV-1 is a suitable candidate as a gene therapy vector for treating these kinds of illnesses.

HSV-1 is an enveloped virus with a diameter of about 150 nm (Fig. 6.29). There are at least 13 different glycoproteins that protrude from the envelope, forming ~700 spikes of various lengths. Beneath the envelope, there is an amorphous matrix of about 20 proteins (tegument) that surrounds a capsid (Fig. 6.29). The HSV-1 genome is a 152-kb, double-stranded DNA molecule that contains about 90 genes. Entry of the virus into a cell is facilitated by the glycoproteins that bind to cell surface receptors. These glycoproteins trigger fusion of the virus with either the host cell plasma membrane or directly to endocytic vesicle membranes, bypassing the formation of a clathrin-coated vesicle. The capsid–tegument complex is routed to the nucleus, where the protein components are removed, and the viral DNA enters the nucleus. Once inside the nucleus, the viral DNA circularizes. HSV-1 DNA does not integrate into host cell chromosomes.

HSV-1-based vectors have a number of advantages when used as therapeutic gene delivery systems. These vectors can carry up to 150 kb of input DNA with either multiple copies of one or more therapeutic genes or very large multigene constructs. Moreover, HSV-1-based vectors are

Figure 6.29 Organization and structure of herpes simplex virus 1 (HSV-1). **(A)** Schematic representation of HSV-1; **(B)** three-dimensional structure of HSV-1 capsid. Panel B is reproduced with permission from the Chimera Image Gallery at https://www.rbvi.ucsf.edu/chimera. Molecular graphics and analyses performed with UCSF Chimera, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from NIH P41-GM103311.



neither toxic nor pathogenic, because they carry no viral coding genes. Once packaged, these vectors infect neurons and other cell types with the therapeutic gene(s) and become localized in the nucleus of the targeted cell. HSV-1-based vectors have been tested in a number of preclinical trials with different therapeutic gene constructs.

Adeno-associated virus (AAV) is a small (~20-nm-diameter), nonpathogenic, nonenveloped, single-stranded DNA virus that infects humans. The AAV genome is about 4.6 kb. The production of AAV particles requires not only AAV gene products but also both host cell proteins and the gene products of another virus (helper virus), which may be adenovirus, herpesvirus, or another virus. In other words, the AAV genome does not contain all the genes that are necessary for the propagation of its own virions. Thus, AAV is assigned to the genus *Dependoparvovirus*. The name adeno-associated virus was coined initially because AAV was found in an adenovirus sample.

Currently, there are 12 known AAV strains. From the perspective of human gene therapy, the various AAV strains provide vectors that are specific for certain cell types and many kinds of tumors. For example, AAV1 is specific for skeletal muscle and cardiac tissue, while AAV8 is specific for liver cells, and in the brain AAV2 is neuron-specific. In addition, it is possible to combine viral capsid protein sequences from different strains to create viral particles that bind to novel cell surface receptors. AAV has the ability to infect both dividing and quiescent cells. The AAV infection cycle is shown in Fig. 6.30. In the absence of a helper virus, integration of the viral genome into a specific site on the short arm of human chromosome 19 occurs. Alternatively, the presence of a helper virus provides proteins necessary for the production of AAV particles. Assembly of virions occurs in the nucleus, and the virus particles pass through the cytoplasm and eventually escape from the cell.

It has been suggested by David Liu, a chemical biologist at Harvard University, that "there are thousands of diseases in humans associated with specific genetic changes," and a significant fraction of those diseases "have the potential to be addressed using genome editing." To make such therapeutic changes, many scientists have recently turned their attention to the use of the CRISPR-Cas system. However, using the AAV to deliver a CRISPR-Cas construct to defective or diseased cells is problematic because AAV virions can only accommodate a limited amount of DNA (~4.5 kb). To get around this roadblock, scientists analyzed the genes for more than 600 different Cas9 enzymes from a wide range of bacteria and found a smaller version of the *cas9* gene that could be packaged in AAV and subsequently be delivered to human cells. The ~3.2-kb *cas9* gene from the bacterium *S. aureus* is substantially smaller than ~4.2 kb for the gene from the common source of this enzyme, *S. pyogenes*. This reduced-size *cas9* gene was small enough to package into AAV virions (Fig. 6.31). Subsequently, scientists packaged the *S. aureus cas9* gene and a gene encoding a single guide RNA (directed against the cholesterol regulatory gene Pcsk9-Proprotein convertase subtilisin/kexin type 9) into an AAV vector (Fig. 6.31). One week after injecting this vector into mice, the researchers observed a significant reduction in serum Pcsk9 levels and in total cholesterol levels, indicating that this approach was successful in mice.

Currently, AAV vectors have been used in ~145 clinical trials worldwide (mostly phase I and II). As can be seen in Table 6.7, it is common for AAV

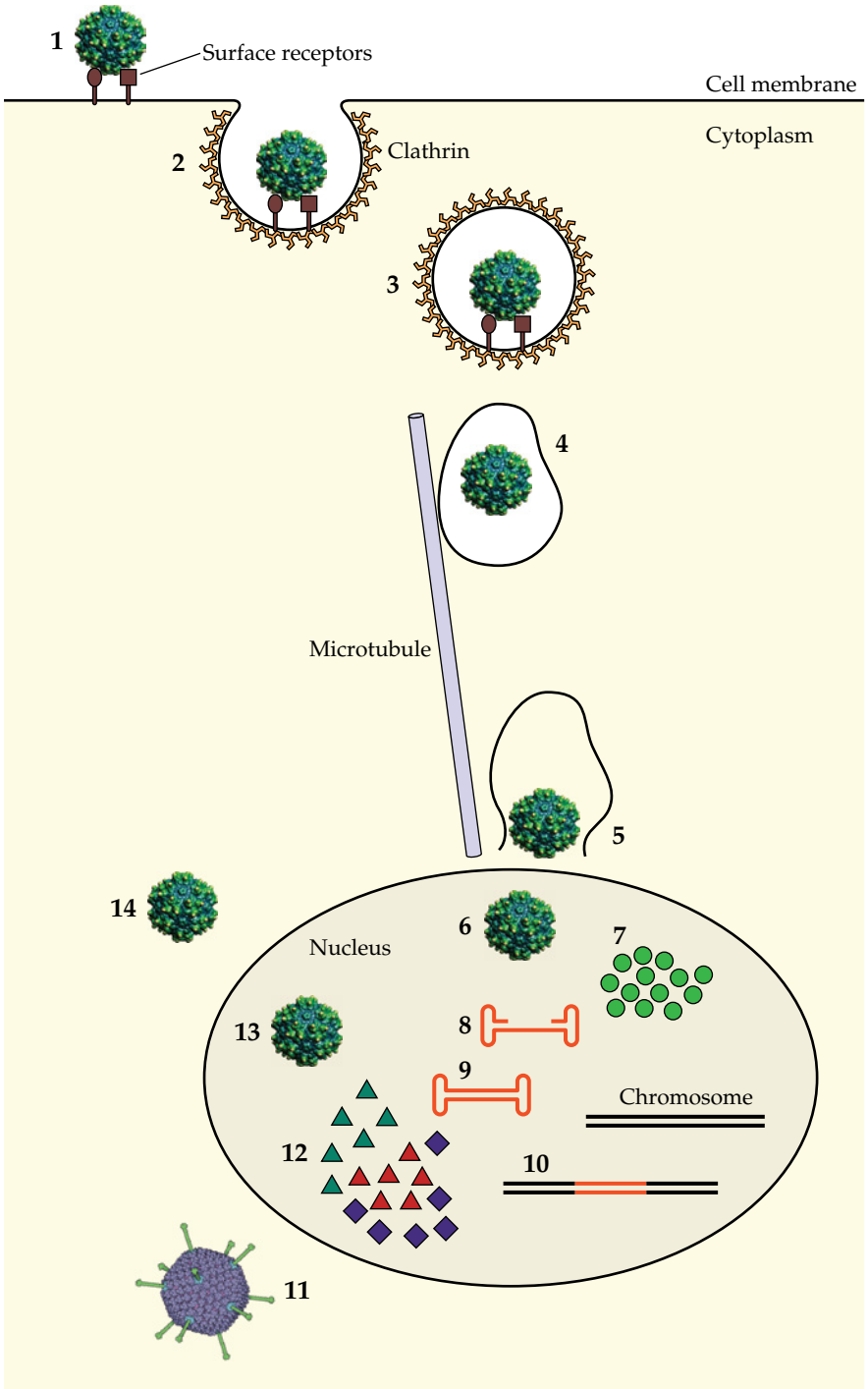


Figure 6.30 Schematic representation of the life cycle of AAV. The virus binds to specific surface receptors **(1)**, and receptor-mediated endocytosis is initiated **(2)**. The virus is internalized and wrapped in a clathrin-coated vesicle **(3)**. Within an endosome, the virus is routed along microtubules to the nucleus **(4)**, from which it escapes **(5)**. The intact virus enters the nucleus **(6)**, where it is uncoated **(7)**, releasing the viral genome **(8)**, which is duplicated **(9)**. In the absence of a helper virus, integration into a chromosome site is likely **(10)**. In the presence of a helper virus **(11)**, both AAV and helper virus proteins are synthesized **(12)** and lead to the assembly of virus particles **(13)**, which leave the nucleus **(14)** and exit the cell (not shown).

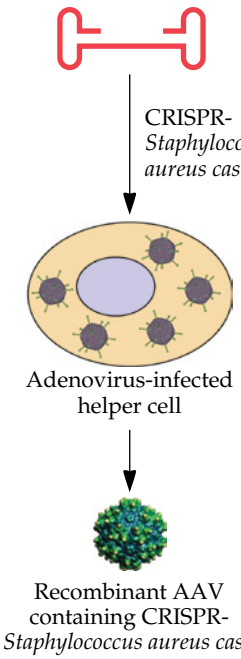


Figure 6.31 The CRISPR-*Staphylococcus aureus cas9* (with a much smaller gene size than the *Streptococcus pyogenes cas9* gene) is packaged in an AAV virion. The gray oval inside the adenovirus-infected helper cell represents the cell nucleus.

trials to include a combined phase I and phase II. This is because many of these trials are intended to develop treatments for relatively rare diseases for which large numbers of patients are not available. These clinical trials include efforts to treat cystic fibrosis, hemophilia A, arthritis, hereditary emphysema, Leber's congenital amaurosis, age-related macular degeneration, Duchenne muscular dystrophy, Parkinson's disease, Batten disease, Alzheimer's disease, spinal muscular atrophy, and congestive heart failure.

One of the major barriers regarding the use of AAV for gene therapy is the affordability of the drug, which is largely a consequence of the high cost of manufacturing. This is in part a result of the difficulty of purifying

Table 6.7 Some ongoing clinical trials involving AAV vectors, as of mid-2020[a]

Tissue target	Disease	AAV strain	Clinical phase
Brain	AADC deficiency	AAV2	I/II
Brain	Batten disease	AAV9	I/II
Brain	MPS-IIIB	AAV5	I/II
Brain	Parkinson's disease	AAV2	I/II
Spinal cord	SMA	AAV9	III
Spinal cord	Giant axonal neuropathy	AAV9	I
Eye	Achromatopsia	AAV2	I/II
Eye	Choroideremia	AAV2	II
Eye	LCA	AAV2	III
Eye	LHON	AAV2	II
Eye	RLBP1	AAV8	I/II
Eye	Wet AMD	AAV8	I
Eye	X-linked RP	AAV2	I/II
Eye	X-linked retinoschisis	AAV2	I/II
Liver	Crigler-Najjar syndrome	AAV8	I/II
Liver	FH	AAV8	I/II
Liver	GSD1a	AAV8	I/II
Liver	Hemophilia A	AAV8	I/II
Liver	Hemophilia B	AAV8	I/II
Liver	MPS-I	AAV6	I
Liver	MPS-II	AAV6	I
Liver	MPS-IIIA	AAV10	II/III
Liver	MPS-VI	AAV8	I/II
Liver	OTC deficiency	AAV8	I/II
Muscle	A1AT deficiency	AAV2	I
Muscle	CMT1A	AAV1	I/II
Muscle	DMD	AAV9	I/II
Muscle	HIV infections	AAV8	I
Muscle	Pompe disease	AAV9	I/I
Muscle	X-linked MTM	AAV	

[a]Note that in a number of instances there is more than one ongoing clinical trial for a particular disease (not shown in this table). Abbreviations: AADC, aromatic L-amino acid decarboxylase; MPS, mucopolysaccharidosis; SMA, spinal muscular atrophy; LCA, Leber's congenital amaurosis; LHON, Leber's hereditary optic neuropathy; RLBP1, retinaldehyde-binding protein 1; AMD, age-related macular degeneration; RP, retinitis pigmentosa; FH, familial hypocholesterolemia; GSD1a, glycogen storage disease type 1a; OTC, ornithine transcarbamylase; A1AT, α1-antitrypsin; CMT1A, Charcot-Marie-Tooth disease type 1A; MYM, myotubular myopathy.

AAV particles from cellular and viral impurities and the removal of AAV empty capsids.

Another barrier to using AAV includes neutralizing antibodies against AAV that circulate in the blood of a significant portion of the human population, which can block AAV gene delivery. For example, in human subjects with cardiac failure, ~60% were found to be positive for antibodies that neutralize AAV1. To avoid this problem, one group of scientists first mapped the majority of the antigenic epitopes on the surface of the AAV capsid to several variable regions. Moreover, several of these antigenic clusters were observed to be shared by different AAV strains. Analysis of different AAV strains/serotypes indicated that antibodies bound to specific sites on the capsid surface (Fig. 6.32), and most of the antigenic determinants were found to cluster around the three-fold symmetry axes on the surface of the AAV capsid. Thus, these researchers began to modify these antigenic epitopes on the capsid surface without altering the infectivity of the virus. Care was taken to avoid changing the amino acid residues responsible for binding the sialylated glycans that are responsible for cell surface attachment. Capsid surface modification was done by multiple rounds of directed evolution. The modified AAVs were then tested for loss of the ability to bind three different AAV-directed monoclonal antibodies. This work is still at an early stage; however, it demonstrates the potential of AAV variants as vectors for foreign DNA that avoid neutralization by circulating anti-AAV antibodies.

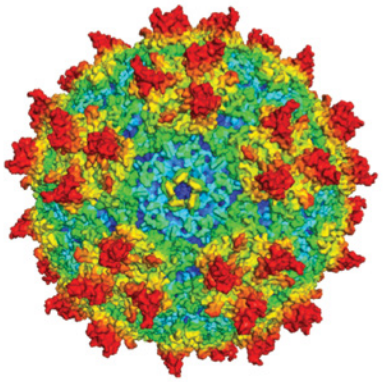


Figure 6.32 A schematic representation of the surface of the AAV capsid showing different proteins and their symmetry.

Nonviral Delivery Systems

Following the discovery of RNAi, a number of major pharmaceutical companies worldwide either acquired siRNA therapeutic technologies (typically by buying small companies that had already developed expertise in this technology) or established this technology in their own labs. However, as a consequence of the failure of existing methods to effectively deliver siRNAs directly to the sites of tumors or infections (and meet the companies' expectations), by 2010 several of these companies shut down their multimillion-dollar RNAi research laboratories. More recently, companies have experienced increasing success delivering RNAi (and other nucleic acid molecules) using nanomaterials. Thus, the global RNAi drug delivery market was estimated to have grown to approximately $39 billion U.S. by 2019 and was expected to grow to around $130 billion U.S. by 2024. These very optimistic projections reflect the recent approval (2018 to 2020) by the U.S. FDA of several RNAi drugs, including drugs that treat polyneuropathy and primary hyperoxaluria type 1.

Direct Injection

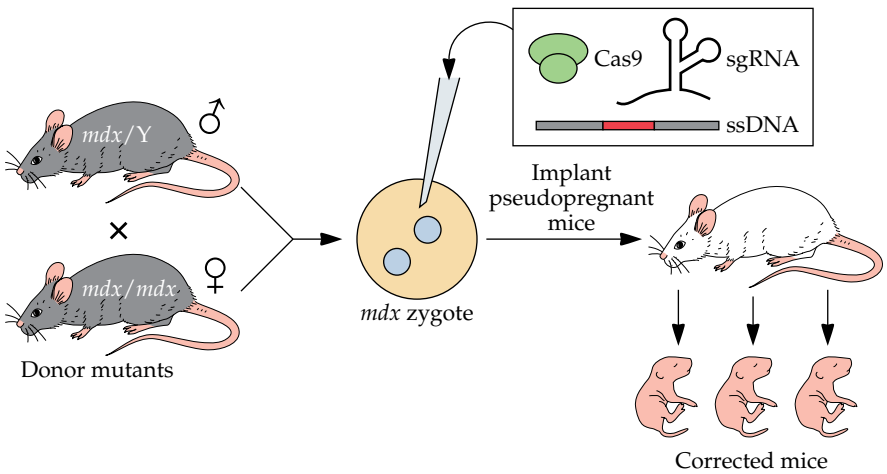
To alter the sequence or expression of either a defective endogenous gene or the gene of a pathogenic agent, naked DNA or RNA may be introduced directly into animal tissues. Nucleic acids may be injected on a macroscale into specific tissues, such as muscle, or on a microscale into specific cells or portions of those cells. Conceptually, microinjection of nucleic acids is the simplest possible nucleic acid delivery system; however, in practice, this technique is fraught with problems. Naked nucleic acids are difficult to deliver due to their rapid clearance, the presence of nucleases that

limit their serum half-life, a lack of organ-specific distribution, and the low efficiency of cellular uptake following systemic delivery. To be effective, the direct injection of nucleic acids requires exceptionally high levels of the input nucleic acid, with the possibility of unwanted side effects being very real. While the effectiveness of siRNAs for specifically inhibiting gene expression in animal cells in culture has been demonstrated on numerous occasions, it is much more difficult to efficiently deliver these RNAs to tissues *in vivo*. However, it may be possible to correct a mutation by directly injecting the guide RNA, corrective DNA, and the endonuclease for the CRISPR-Cas system (Fig. 6.33).

Duchenne muscular dystrophy, a recessive X-linked trait (i.e., the gene is located on the X chromosome), is the most common of the muscular dystrophies and occurs in about 1 in 3,500 males (and about 1 in every ~12,250,000 females). Onset of muscle weakness begins in childhood, and the condition progressively worsens, and often death occurs in the early twenties. Other muscular dystrophies have different times of onset and patterns of progression and affect particular muscle groups. There are no effective treatments for the different muscular dystrophies. The genetic defect in Duchenne muscular dystrophy is due to mutations of the dystrophin gene. This gene encodes a structural protein that is part of a dystrophin–glycoprotein complex, which maintains the structural integrity of the plasma membrane of muscle cells.

In a very limited number of circumstances, direct injection may be an efficient and effective means of introducing exogenous nucleic acids that change the genome of the injected animal (Fig. 6.33). Recently, a group of scientists employed the CRISPR-Cas9 system to correct a dystrophin gene mutation (*dmx*) in the germline of dystrophic mice. The sgRNA, Cas9, and

Figure 6.33 Germline correction of a mutation (on the mouse X chromosome) that causes Duchenne muscular dystrophy. A male and a female mouse that both carry the *mdx* gene (encoding a mutated dystrophin gene) are mated to produce a zygote in which the mutation is corrected using the CRISPR-Cas system. The Cas9 endonuclease, sgRNA, and ssDNA (single-stranded donor DNA carrying the correct *mdx* sequence) are introduced by microinjection. Modified zygotes are implanted into pseudopregnant mice. The offspring dystrophin genes have partially or completely corrected mutations, and they are disease free or disease limited.



corrective single-stranded donor DNA were directly injected in zygotes of mice that contained the defective gene on their X chromosome. The single-stranded donor DNA contained 90 base pairs of homologous sequences flanking the target site (the site that carries the mutation that is targeted for correction). The altered zygotes were then implanted into pseudopregnant mice that had been manipulated to be hormonally pregnant so that their uterus is receptive to zygote implantation. When the progeny were examined, they were found to be mosaics, individuals carrying a mixture of the mutant and the wild-type alleles, with 2 to 100% of the mutant genes having been corrected to the wild type (i.e., disease free). Although this study demonstrates the efficacy of direct injection of nucleic acids to correct gene mutations in mice, germline genome editing is currently not acceptable in humans.

Lipids

One approach used to overcome some of the problems associated with direct injection of therapeutic nucleic acids has been to chemically couple a therapeutic molecule such as an siRNA (at the terminal hydroxyl group of the sense strand RNA) to a lipid molecule such as cholesterol (Fig. 6.34). This was tested with an siRNA that was complementary to an mRNA encoding apolipoprotein B, a protein involved in the metabolism of cholesterol. Following intravenous introduction of the siRNA–cholesterol complex by injection into mice, the complex was specifically taken up by the liver, jejunum (part of the small intestine), heart, kidneys, lungs, and fat tissue cells. Once the complex was inside cells of these tissues, the sense strand of the siRNA was destroyed, and the antisense strand, together with the Argonaute protein, bound to the target mRNA.

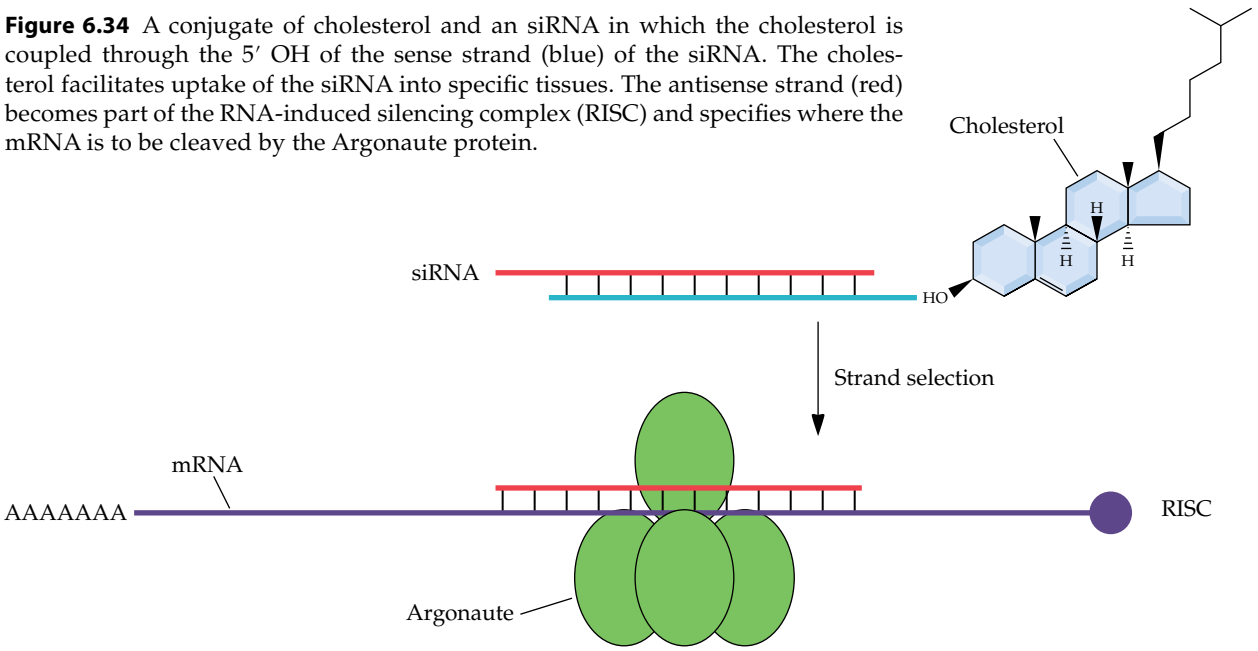


Figure 6.34 A conjugate of cholesterol and an siRNA in which the cholesterol is coupled through the 5′ OH of the sense strand (blue) of the siRNA. The cholesterol facilitates uptake of the siRNA into specific tissues. The antisense strand (red) becomes part of the RNA-induced silencing complex (RISC) and specifies where the mRNA is to be cleaved by the Argonaute protein.

With this approach, the level of apolipoprotein B was reduced by more than 50% in the mouse liver and by 70% in the jejunum. This resulted in a significant decrease in the plasma apolipoprotein B level, as well as in the total amount of cholesterol. This strategy is an important first step in the development of a method to therapeutically lower cholesterol levels in humans.

Several other lipid molecules, including some long-chain fatty acids and bile acids, have been used in place of cholesterol to mediate the uptake of siRNAs into cells. Not only does lipid conjugation increase cellular uptake of siRNAs, but it also significantly increases the plasma half-life of the siRNA. A critical factor in mediating the interaction between fatty acid-conjugated siRNAs and lipoprotein particles that deliver lipids to cells from the bloodstream is the length of the fatty acid alkyl chain (with longer chains being more hydrophobic). Thus, docosanyl (C22) and stearoyl (C18) conjugates bind more tightly to low-density lipoproteins and subsequently silence gene expression more effectively *in vivo* (in mice) than lauroyl (C12) and myristoyl (C14) conjugates (which prefer to bind to high-density lipoproteins). Studies are under way to improve the delivery of lipid-conjugated siRNAs to treat a wide range of diseases.

As mentioned briefly earlier in this chapter, scientists have had some success delivering nucleic acids using both synthetic and biological liposomes (Fig. 6.11). Biological liposomes are phospholipid-based spherical particles derived from human cells. If biological liposomes are derived from patients, they are likely to be recognized as self by one's immune system and thereby avoid an immune attack. However, biological liposomes tend to be difficult to harvest and maintain and are therefore not currently used to any significant extent. Notwithstanding the above-mentioned limitations of liposomal delivery, when apolipoprotein B siRNA-carrying liposomal particles were intravenously injected into monkeys, greatly reduced apolipoprotein B expression and serum cholesterol levels were observed.

One way to significantly improve on the results obtained by merely conjugating lipid molecules to siRNAs or encapsulating siRNAs in liposomes includes designing and (chemically) synthesizing lipopeptide nanoparticles (Fig. 6.35). These particles are composed of synthetic lipopeptides (lipoamino acids or lipopolypeptides), phospholipids, cholesterol, PEG-lipids, and siRNA. The modular design of lipopeptide nanoparticles enables researchers to fine-tune their chemical and physical properties while facilitating the entrapment of siRNA and the subsequent more accurate and effective delivery of the bound siRNAs. Lipopeptide nanoparticles are designed to mimic apolipoproteins by conjugating lipid tails to amino acids, peptides, and polypeptide head groups. Moreover, if they are found to be useful for delivering a range of different siRNAs, lipopeptide nanoparticles can be produced on a large scale without using cells or animals. In addition to the nanoparticles described here, more recently, some scientists have constructed nanoparticles that are composed of PEG-lipids, dendrimers (described later in this chapter), and siRNA. These dendrimer nanoparticles may be used to deliver therapeutic RNAs and can also serve to facilitate the delivery of RNAs that are used as vaccines.

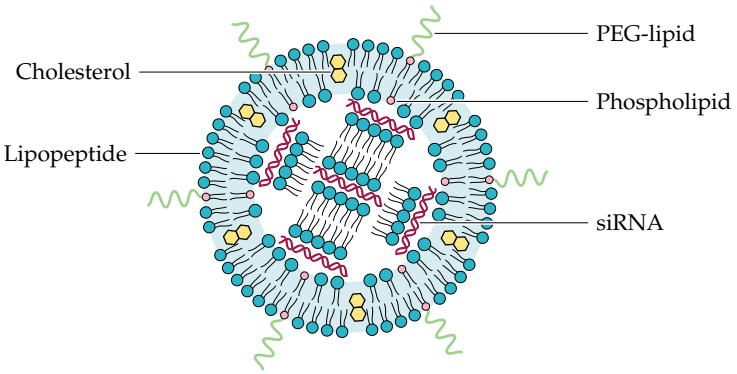


Figure 6.35 Schematic representation of a lipopeptide nanoparticle. PEG is polyethylene glycol.

Bacteria

Nonpathogenic bacteria that are normally found in association with various mammalian tissues and cells may be isolated, attenuated, and then genetically engineered to deliver therapeutic nucleic acid molecules. With these bacteria, toxin genes are removed, but the genes encoding functions that enable the bacterium to attach to and invade tissue, i.e., the features that are important for their use as a delivery vehicle, are retained. To date, bacterial nucleic acid delivery has been used mainly in cancer gene therapy and for DNA vaccination; it is also being tested for the treatment of genetic diseases, including cystic fibrosis. The nonpathogenic strains of bacteria currently used include *Listeria monocytogenes*, certain *Salmonella* strains, *Bifidobacterium longum*, and modified *Escherichia coli*. The engineered bacteria are used as vectors to deliver the nucleic acid therapeutic agents directly to the affected tissues. For example, a nonpathogenic strain of *E. coli* was transformed with the plasmid vector TRIP, containing the gene that encodes the protein invasin, which permits *E. coli* to bind to and enter β1-integrin-positive mammalian cells, and the gene *hlyA*, which encodes listeriolysin O, a protein that forms pores in entry vesicles and enables genetic material to escape (Fig. 6.36). The TRIP vector encodes an shRNA molecule directed against the mRNA produced by the cancer gene *CTNNB1* under the control of a bacterial promoter. As long as a bacterium is able to enter target mammalian cells and release shRNAs that are then processed to siRNAs by the host RISC (Fig. 6.14), the siRNAs may be directed against any specific mRNAs. The *E. coli* cells act as a vector to transport the shRNAs to where they are required, e.g., specific types of cancer cells. One important reason for the effectiveness of this approach is that bacterium-mediated RNAi expression may evade the host defense against exogenous DNA because the DNA is enclosed within the bacteria, potentially allowing for long-term expression of the transgene. This approach has been shown to work both in cancer cells in culture and *in vivo* with mice. Using live animals, the bacteria can be administered orally. The intrinsic toxicity of bacteria is the most significant risk in using bacteria as a delivery vehicle (i.e., bactofection). More recently, one group of scientists reported using an

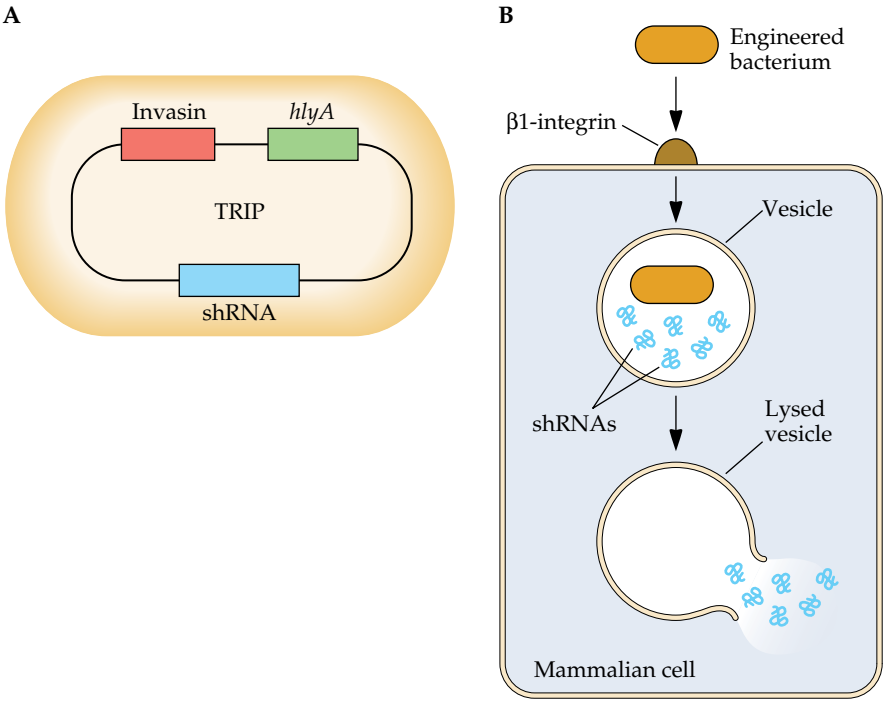


Figure 6.36 Use of a nonpathogenic strain of *E. coli* to deliver siRNAs to certain tissues. **(A)** Plasmid vector TRIP carrying a gene encoding the protein invasin, which permits *E. coli* to bind to and then enter mammalian cells that express the protein β1-integrin on their surface, an *hlyA* gene encoding listeriolysin O which lyses the vesicle surrounding the bacterium inside the mammalian cell, and an shRNA gene. **(B)** The engineered bacterium binds to the target mammalian cell, is taken up into a vesicle inside the cell, and synthesizes shRNAs before the vesicle is lysed and the shRNAs are released. The shRNAs are subsequently converted into siRNAs.

attenuated strain of *Salmonella* to deliver an siRNA targeting an immune checkpoint molecule involved in tumor immune escape to mice with melanoma (which were also treated with the drug pimozide). The combination of the siRNA and the drug had a significant therapeutic effect on the mice. Using attenuated strains of *Salmonella enterica* serovar Typhimurium in phase I clinical trials with several human cancers showed no significant side effects.

Another way in which bacteria may be used to target specific mammalian cells includes the use of bacterial minicells (Fig. 6.37A). In this case, a normal-size bacterial cell carries a mutation in a gene that determines the site of cell division; this mutation results in the production of minicells (only a fraction of the size of normal bacteria) that do not contain any chromosomal DNA. If minicell-forming bacteria are first transformed with plasmid DNA, the minicells, although they are devoid of chromosomal DNA, will contain plasmid DNA. The much smaller minicells may be physically separated from normal-size bacterial cells by centrifugation and/or cross-flow filtration. Minicells may be targeted to cancer cells using two different antibodies. One antibody targets a protein on the surface of the minicell, and another targets the surface of a cancer cell. The two types of antibodies are linked to one another through their Fc regions

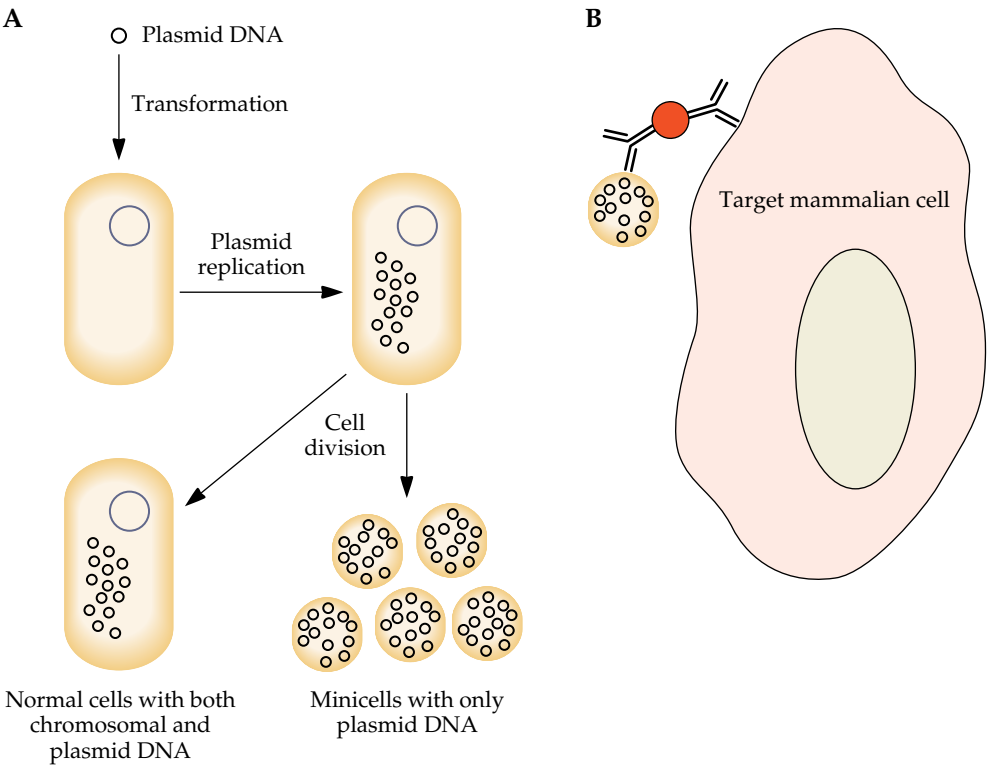


Figure 6.37 Production of bacterial minicells carrying plasmid DNA encoding either a mammalian cellular toxin or an shRNA. **(A)** Plasmid DNA is present in multiple copies, while chromosomal DNA is present in a single copy. Aberrant cell division results in the formation of both normal-size cells containing both chromosomal and plasmid DNA and minicells containing only plasmid DNA. **(B)** Normal-size cells and minicells are separated by centrifugation or cross-flow filtration. Minicells may be linked to the targeted mammalian cells using antibodies directed against proteins on the minicell and mammalian cell surfaces. The Fc portions of the antibodies are linked by the addition of protein A/G (a recombinant fusion protein that combines IgG Fc region binding domains of both protein A, from *S. aureus*, and protein G, from *Streptococcus* spp., shown in red).

by protein A/G (where this protein is a recombinant fusion of protein A originally from the cell wall of *S. aureus* and protein G from the surface of streptococcal bacteria) (Fig. 6.37B). When the target cell is a cancer cell, the minicell, which can produce protein encoded by the plasmid (because RNA polymerase and ribosomes are partitioned into the minicell), may be engineered to produce a compound that is toxic to the cancer cell. Alternatively, plasmids carried by minicells may be designed to produce specific shRNAs that are eventually converted into siRNAs by host RISC. In practice, minicells were shown to be able to successfully deliver a variety of chemotherapeutic agents to target cells, suggesting that this should be an effective means of delivery of plasmid-encoded products to target cells. While most of the early studies with minicells used strains of *E. coli*, researchers have shown that several strains of lactic acid bacteria (discussed in chapter 5) can also produce minicells, and these can also be used for drug delivery.

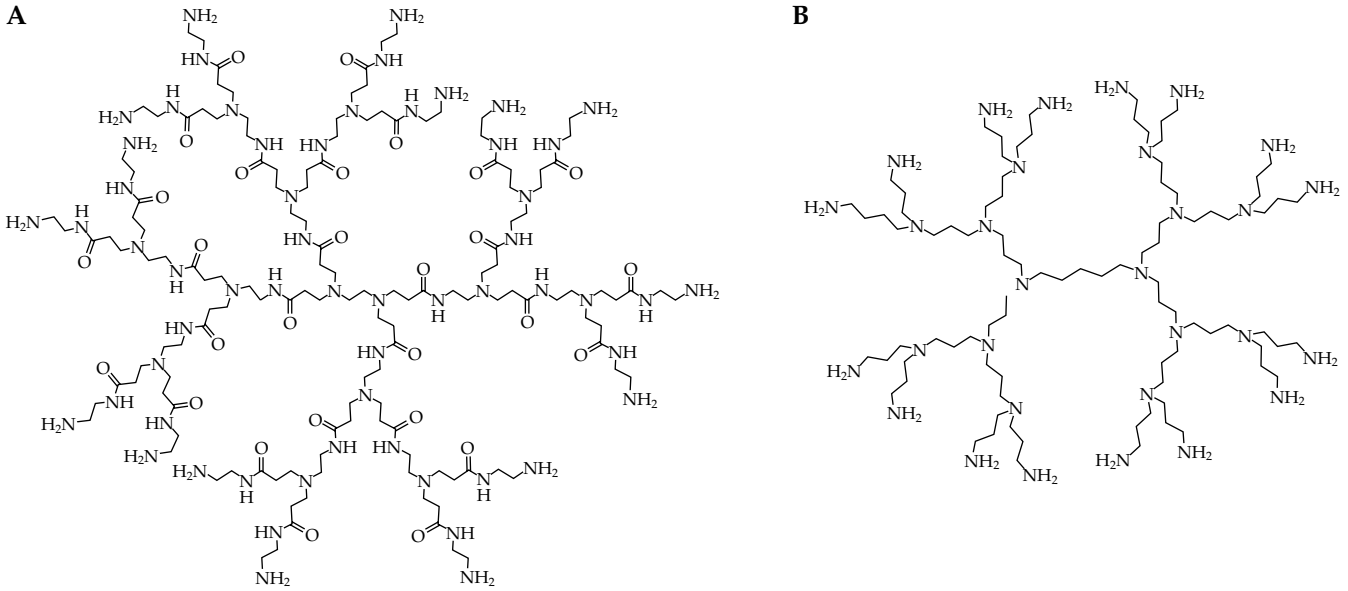
Dendrimers

In the past 10 to 15 years, numerous polycations and polymer micelles have been used for formulating genes, shRNAs, and siRNAs into complexes that have been termed "polyplexes." The polycations that have been used include histones, polylysine, cationic oligopeptides, polyethyleneimine, polypropyleneimine, dendrimers, poly[2-(dimethylamino)ethyl methacrylate], and chitosan.

Dendrimers are biocompatible, nonimmunogenic, water-soluble nanoparticles with sizes ranging from 1 to 15 nm. They possess terminal modifiable amine functional groups as the "sensors" for binding various targeting or guest molecules that they will ultimately deliver to target cells. Unlike classical polymers, dendrimers have a high degree of molecular uniformity, a narrow molecular weight distribution, specific size and shape characteristics, and a large number of potentially interactive terminal amino moieties (Fig. 6.38). Dendrimers are produced in an iterative sequence of reaction steps, with each additional iteration leading to a higher-generation dendrimer. Each new "generation" has a new layer with double the number of active sites (i.e., terminal amino groups) and approximately double the molecular weight of the previous generation. It is relatively easy to precisely control the size, composition, and chemical reactivity of dendrimers.

Dendrimers can form complexes with nucleic acid drugs, including plasmid DNA, shRNA, and siRNA, through electrostatic interactions, and they also bind to glycosaminoglycan molecules, such as heparin sulfate, hyaluronic acid, and chondroitin sulfate, that are typically found on cell surfaces. In addition, dendrimers are more efficient and safer than either cationic liposomes or other cationic polymers for *in vitro* gene transfer. The high transfection efficiency of dendrimers has been attributed to both their well-defined shape and the low pK_as of the amines. One dendrimer-based formulation for the prevention of genital herpes with activity against herpes virus (i.e., VivaGel) has successfully completed phase II clinical trials.

Figure 6.38 Structures of two commercially available dendrimers.



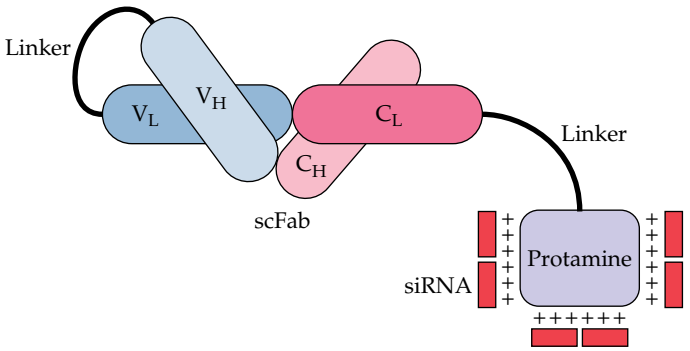


Figure 6.39 A single-chain Fab (scFab) fragment directed against a mammalian cell surface protein is fused to the positively charged polypeptide protamine, which binds noncovalently to negatively charged siRNAs. The Fab fragment delivers the siRNA to specific cells. Note that a conventional (two-chain) Fab fragment has also been used to deliver siRNAs (not shown).

Antibodies

Monoclonal antibodies (see chapter 5) may now be generated against nearly any target protein. These antibodies, or their variable regions, may readily be humanized and then produced in heterologous host cells. Moreover, at the DNA level, it is relatively straightforward to fuse the antibody gene(s) with the gene for another protein. With this in mind, a gene encoding a single-chain Fab fragment that binds specifically to an oncogenic protein called ErbB2 (encoded by the *ERBB2* gene, which is also called *HER2*), which is found on the surfaces of breast cancer cells, was fused to the gene for the positively charged nucleic acid-binding protein protamine (an arginine-rich small protein). The Fab portion of the fusion protein (at the N terminus) binds to the surfaces of cells expressing ErbB2, while the C-terminal end of the fusion protein carries the 51-amino-acid-long protamine, which readily binds to added siRNAs (Fig. 6.39). In another example of this system, an anti-HIV envelope Fab and an siRNA that is designed to cleave the HIV *gag* mRNA were employed. Using cells in culture, it was possible to reduce the amount of Gag protein (the protein of the nucleocapsid shell around the RNA of a retrovirus) by >70%. This system also works *in vivo* in mice when the construct is injected either intravenously or directly into tumors. The hope is that by using a combination of specific antibodies (or antibody fragments) that direct siRNAs only to certain cells, and siRNAs that selectively cleave specific target mRNAs, this system can be used to treat a wide range of diseases.

A variation on the theme of employing single-chain Fab fragment includes using a single-chain Fv molecule (which is somewhat smaller than a Fab molecule) that has been coupled to a polymer of nine arginine residues (molecular weight = 1,567.8), which is much smaller than the 51-amino-acid-long protamine molecule mentioned above. This variant procedure is equally as effective at both binding siRNAs (and other small RNAs) and binding to specific cell surface proteins.

Aptamers

The binding specificity to a small target antigen that is a central feature of the functioning of antibodies is also an intrinsic property of aptamers. Aptamers are able to bind to a variety of molecules, enabling the development of

aptamer-based delivery systems. Thus, conjugating aptamers, which bind to specific cell surface proteins, to siRNAs that are designed to reduce the expression of certain mRNAs provides another method of targeting siRNAs to specific tissues or cells. Also, since both aptamers and siRNAs are chemically synthesized RNA oligonucleotides, it is simple and straightforward to synthesize chimeric RNA molecules that include both the binding specificity of an aptamer and an siRNA that targets a specific mRNA (Fig. 6.40). For example, an aptamer that binds selectively to a prostate-specific membrane antigen (found on cancerous prostate cells) was first selected. Then, a 21-bp siRNA directed against mRNAs encoded by either of two genes that are necessary for prostate cancer cells to survive was added to the aptamer sequence. The activities of both aptamer binding and RNAi were maintained in the chimeric molecule. The targeting aptamer did not impair the ability of the siRNA to silence the target gene, and the presence of the siRNA did not affect the ability of the aptamer to bind to its target. This simple but highly effective approach should be amenable to treating a wide range of human diseases, provided that (i) silencing specific genes in a population produces therapeutic benefits and (ii) there are surface receptors (usually proteins) that distinguish the target cell population and allow the siRNA to be internalized by the cell.

Transposons

Transposons are discrete DNA elements that have the ability to move from one chromosomal location to another. They are typically excised from one chromosomal locus and then integrated into another location; this movement is catalyzed by an enzyme called a transposase. Being

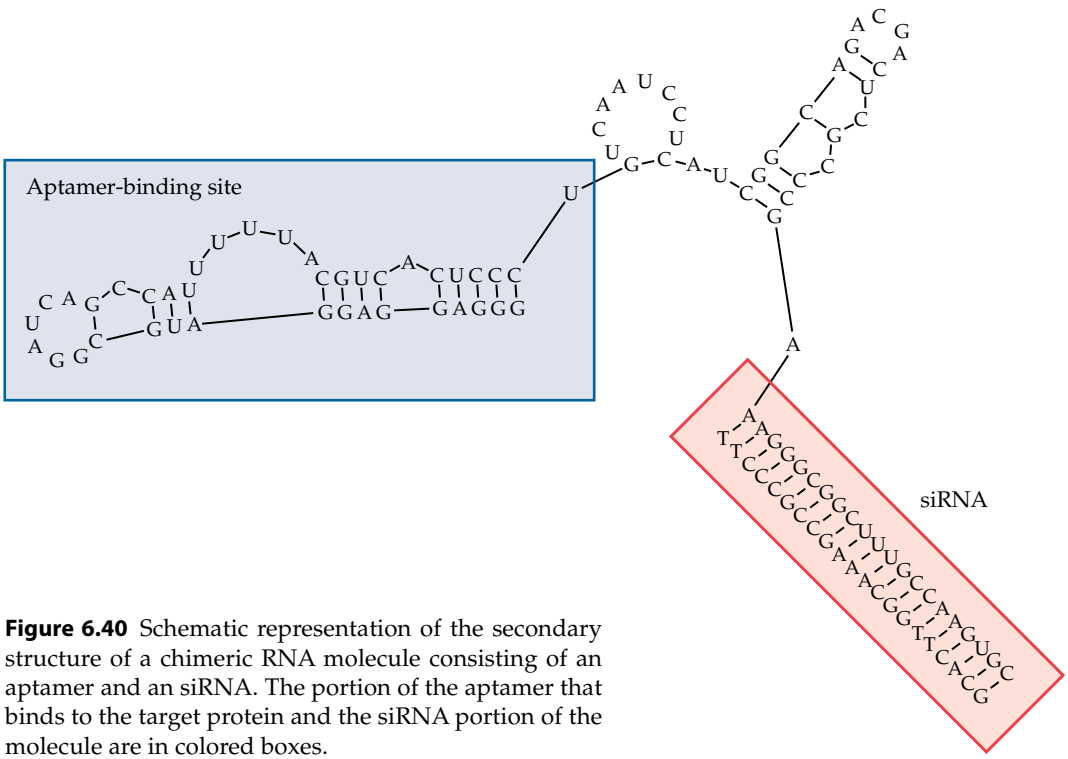


Figure 6.40 Schematic representation of the secondary structure of a chimeric RNA molecule consisting of an aptamer and an siRNA. The portion of the aptamer that binds to the target protein and the siRNA portion of the molecule are in colored boxes.

able to move DNA from one site to another makes transposons attractive as potential foreign gene delivery tools. The transposase can act in *trans* (i.e., it can be encoded on a separate DNA fragment) to excise virtually any target DNA sequence that is flanked by the terminal repeat DNA sequences, normally found at each end of the transposon. To use DNA transposons for gene delivery, a simple system has been developed (Fig. 6.41). This system consists of a plasmid that encodes and expresses the transposase, as well as a plasmid containing the target DNA to be integrated, which is flanked by the transposon terminal repeat sequences that are required for transposition. Once it is expressed, the transposase binds to the terminal repeat DNA sequences and catalyzes the excision of the gene of interest from the plasmid. The transposase then catalyzes the insertion of the target DNA into the genome of the host cell. Physically separating the transposase gene from the target gene-containing plasmid optimizes the stoichiometry of both components and hence the efficiency of the process.

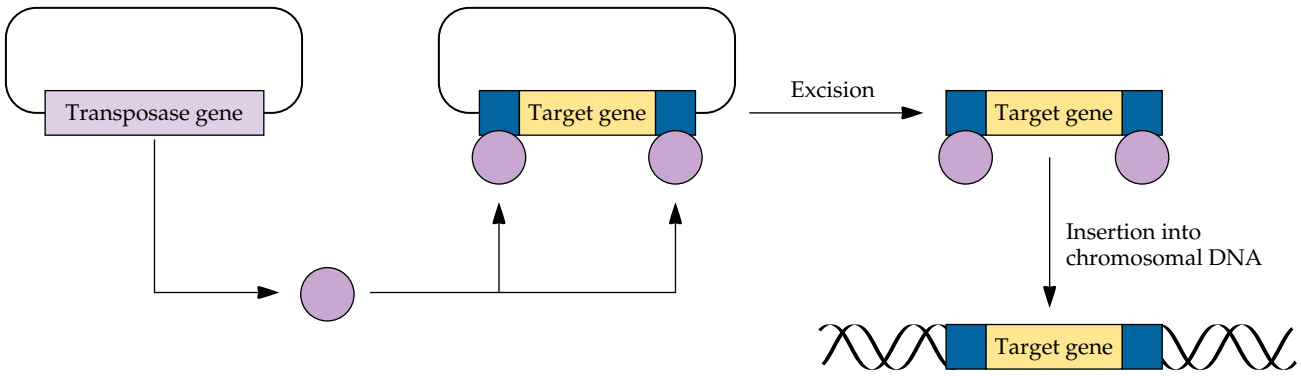
At present, three different transposon systems are being developed as gene therapy delivery vectors: *Sleeping Beauty*, *Tol2*, and *PiggyBac*. *Sleeping Beauty*, which has been used most frequently, was constructed by combining fragments of silent and defective Tc1/mariner elements from salmonid fish and shows significant transposition efficiencies in vertebrate cells. When tested, the original *Sleeping Beauty* transposon was at least 10-fold more efficient than other Tc1/mariner transposons. Moreover, to increase its activity, almost every amino acid in the transposase has been changed by DNA shuffling and PCR (see chapter 3). Following the screening of a large number of variants of the original transposon, one version, SB100X, was found to be more than 100-fold more active at catalyzing transposition in HeLa cells than the original *Sleeping Beauty* (Table 6.8).

Since *Sleeping Beauty* may integrate into any position in the chromosomal DNA, there is a risk of disrupting an important gene. However, in contrast to the case with many viral vectors, the terminal repeat sequences

Table 6.8 Relative efficiencies of different transposition systems

Transposon system	Relative transposition activity
Tc1/mariner	1
Sleeping Beauty	10
Sleeping Beauty with modified transposase (SB100x)	1,000

Figure 6.41 Representation of transposase-catalyzed movement of a target gene into cellular chromosomal DNA. The transposase gene and the target gene flanked by transposon inverted repeat sequences (shown in blue) are encoded on separate plasmids. The transposase (depicted as a purple circle) binds to the transposon inverted repeat sequences and catalyzes the excision of the transposon-flanked target gene from the plasmid and its subsequent insertion into chromosomal DNA.



of *Sleeping Beauty* have very low intrinsic promoter/enhancer activity. Thus, *Sleeping Beauty* cannot readily activate the expression of chromosomal genes that flank the transposon integration sites. However, the internal promoter/enhancer sequences that are used to drive the expression of the target gene may potentially activate expression of neighboring genes in the proximity of the integration site. To prevent this from happening, the expression cassette in a *Sleeping Beauty* transposon was flanked with insulator DNA sequences that significantly reduce the risk of activating the expression of neighboring genes.

The SB100X version of the *Sleeping Beauty* transposon has been used successfully to transfer genes into mice, rats, rabbits, and pigs, and this system is currently being tested in several separate phase I human clinical trials. In some of these trials, human T cells are genetically altered with *Sleeping Beauty* transposons and then transferred for adoptive immunotherapy (a way of using the immune system to fight cancer) into patients. In mid-2020, participants were being recruited for a phase II clinical trial to express transposon-engineered T cells expressing chimeric antigen receptors in patients with metastatic cancer. It has recently been claimed that, given the high costs associated with the manufacture of clinical-grade viral vectors and the limited number of available viral vectors, introduction of foreign genes with plasmid-based gene transfer systems based on transposable elements may have a number of significant advantages over the viral-based systems, including stable gene transfer without any possibility of toxicity.

Gene Therapy

Human somatic cell gene therapy is still very much in its infancy. Nevertheless, scientists have achieved considerable success in introducing genes and gene fragments into human cells in culture and in mouse models of numerous human diseases. Moreover, there has been limited success in treating several human diseases *in vivo* (and *ex vivo*). Many of the earlier experiments were performed using viral vectors to introduce human cDNAs. More recently, a wide range of nonviral vectors have been used to successfully deliver a range of nucleic acids that can modify human gene expression. Independent of how foreign genes are delivered, gene therapy may be used to augment or suppress the activity of a target gene, or a target gene may be altered by genome editing (Fig. 6.42). Here, some examples of the implementation of this technology are discussed (Box 6.3).

In mid-December 2020, it was reported that a group of researchers, using CRISPR technology, had been able to partially dilute some of the effects of sickle cell disease, including extreme fatigue, shortness of breath, dizziness, and irregular heartbeat. People with sickle cell disease have atypical hemoglobin molecules which distort red blood cells into a sickle shape, with individuals carrying one copy of the sickle cell hemoglobin gene being protected from malaria parasites. In this CRISPR treatment, researchers removed a patient's blood stem cells and, in the lab, disabled a genetic switch that normally turns off the synthesis of fetal hemoglobin. The patient then received chemotherapy to knock out their diseased cells, and the altered cells were then infused into the patient. Although this treatment is still in its infancy, it promises to be an important therapy for the millions of people worldwide with sickle cell disease.

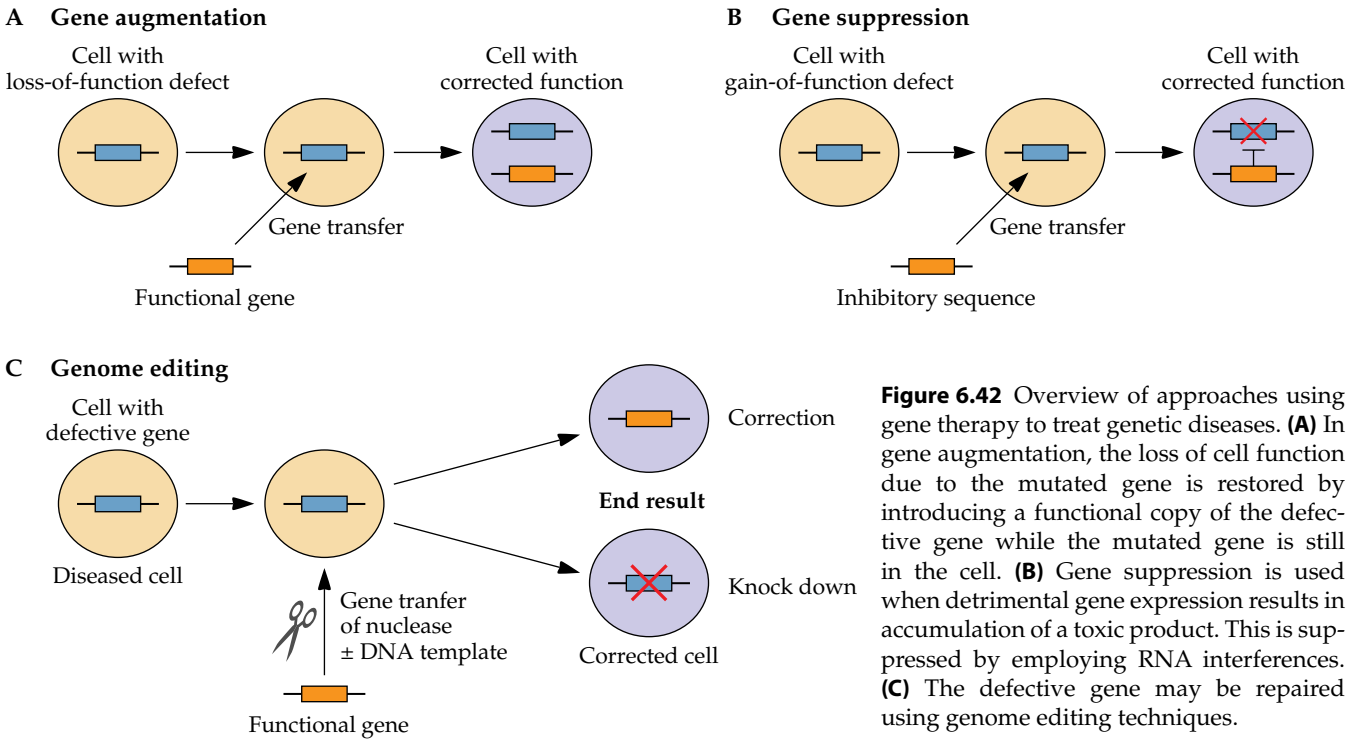


Figure 6.42 Overview of approaches using gene therapy to treat genetic diseases. **(A)** In gene augmentation, the loss of cell function due to the mutated gene is restored by introducing a functional copy of the defective gene while the mutated gene is still in the cell. **(B)** Gene suppression is used when detrimental gene expression results in accumulation of a toxic product. This is suppressed by employing RNA interferences. **(C)** The defective gene may be repaired using genome editing techniques.

First Gene Therapy for β-Thalassemia Approved

box 6.3

β-Thalassemia is an inherited (recessive) blood disorder that is caused by absent or reduced synthesis of the beta chains of hemoglobin. This disease is estimated to occur at a frequency of approximately 1 per 100,000 people globally (or around 80,000 affected people of the world's ~8 billion people). This disorder leads to the underproduction of hemoglobin and to microcytic anemia (small and less effective red blood cells). Therefore, individuals with this disease may require frequent blood transfusions throughout their lifetime.

In June 2019, the company Bluebird Bio was granted permission by the European Commission to market a gene therapy treatment for β-thalassemia. Given the relatively small number of individuals with this disease, the European Commission's approval was based on clinical data from only 32 adults and adolescents. This therapy provides affected individuals with a corrective gene (called Zynteglo) whose product (β-globin) combines with α-globin to produce functional hemoglobin, thereby curing this disease. Zynteglo is administered by *ex vivo* gene therapy in which a patient's bone marrow stem cells are harvested (outside their body) and then transduced (in a laboratory) with a β-globin gene delivered by a lentivirus vector. The transduced cells containing the corrective gene are reintroduced to the patient in a single injection, after which they become integrated into the patient's bone marrow and subsequently differentiate into red blood cells that produce normal hemoglobin.

Unfortunately, Zynteglo is only approved for the treatment of mild to moderate (but not severe) β-thalassemia. The apparent reason why this treatment was not effective with severe β-thalassemia relates to the fact that adding some cells that synthesize normal hemoglobin to a background of cells that produce only extremely limited amounts of functional hemoglobin is not sufficient to correct this disorder. Moreover, the cost for a course of this treatment is currently $1.8 million, an extremely high price, even for national health care providers. To make this high price somewhat more acceptable, the company has suggested a "pay-for-performance" plan in which the treatment would be paid for in five payments of $350,000 (over the course of a year), with continuing payments contingent on the treatment continuing to be effective.

Mitochondrial Diseases

As indicated in chapter 5, human mitochondrial DNA encodes 37 unique proteins. While defects in any of the large number of nuclear encoded proteins that are imported into the mitochondria may be corrected by enzyme replacement therapy, this is not readily achieved with defective proteins that are encoded in the mitochondrial DNA (defective mitochondrial genes occur in about 1 in every 5,000 individuals). Nevertheless, researchers have recently developed a technique wherein a defective mitochondrion may be replaced at the level of fertilization of the egg (Fig. 6.43). This strategy, which has been called mitochondrial replacement therapy, might also be called gene therapy using two mothers or the creation of three-parent babies. If mother number 1 has a mutation in one of her mitochondrial genes, the nucleus from one of her eggs may be inserted into an egg from mother number 2, whose nucleus has been removed and who does not have any defective mitochondrial genes. The modified egg may then be fertilized with the father's sperm, and once a zygote develops, it may be implanted into mother number 1. The infant that develops from this sort of manipulation will have a full set of nuclear chromosomal genes from mother number 1 and the father and normal mitochondrial genes from mother number 2. This genetic manipulation has been performed successfully in a very limited number of cases, with the first child being born following this manipulation in Mexico in 2016. The parents of this baby chose not to participate in a long-term follow-up study of the health of their child. At the present time, there are a number of safety concerns associated with this procedure, which is banned in the United States.

Prodrug Activation Therapy

Prodrug activation therapy (suicide gene therapy) is designed for localized eradication of tumor cells with minimal debilitating biological effects that often accompany chemo- or radiotherapy. In practice, prodrug activation therapy entails introducing into tumor cells a gene for an enzyme that is not part of the human repertoire and is under the control of a tumor-specific promoter. A prodrug is then delivered to the tumor mass. The enzyme is synthesized in the recipient tumor cells and subsequently converts the prodrug to its active form, thereby leading to tumor cell death. When a gene causes the death of its own cell, it is called a suicide gene. Moreover, the activated drug molecule can enter into neighboring cells, which very likely did not receive the suicide gene, and kill them (Fig. 6.44). The latter process has been designated the bystander effect.

Many prodrug-modifying enzyme combinations have been devised. For example, the ganciclovir [2-amino-9-(1,3-dihydroxypropan-2-yloxymethyl)-3H-purin-6-one]–herpes simplex virus thymidine kinase (HSV TK) duo has been thoroughly examined. Briefly, HSV TK phosphorylates ganciclovir to form ganciclovir monophosphate (Fig. 6.45). Although host cell kinases do not efficiently phosphorylate ganciclovir, they readily add phosphate groups to ganciclovir monophosphate to form ganciclovir triphosphate. Ganciclovir triphosphate inhibits DNA polymerase activity and terminates DNA synthesis, causing the death of a proliferating cell. Moreover, monophosphorylated ganciclovir can pass into unmodified cells through gap junctions and kill these cells as well. Generally, one HSV TK-expressing tumor cell destroys up to 10 unmodified neighbor cells.

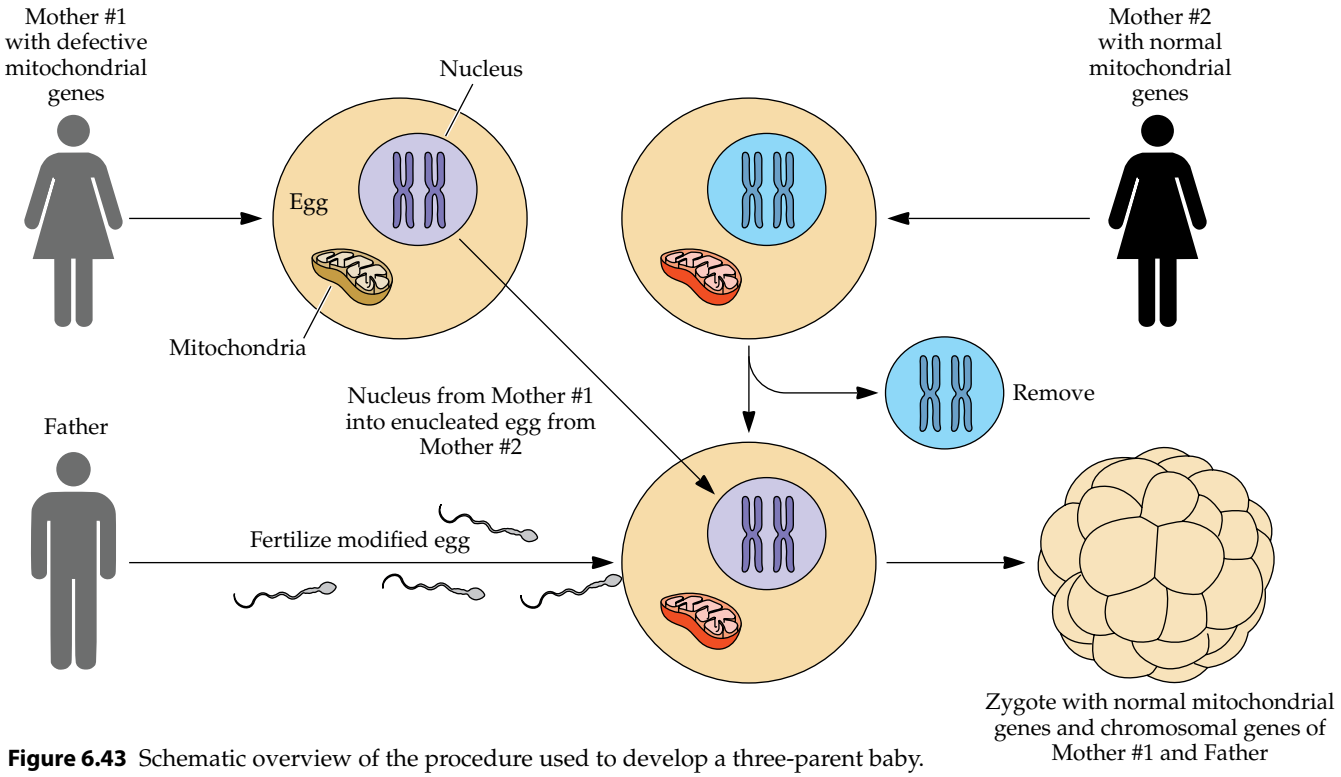


Figure 6.43 Schematic overview of the procedure used to develop a three-parent baby.

Various strategies have been devised to improve both the efficiency of gene transfer into tumor cells and the extent of penetration into a tumor mass, as well as augmenting the bystander effect of prodrug activation systems. For example, by creating a double suicide gene combination of 5-fluorouracil-cytosine deaminase and ganciclovir–HSV TK with both of the enzyme-coding genes under the control of a cancer-specific promoter, tumor destruction was increased with little effect on noncancerous cells. Clinical trials have shown that the ganciclovir–HSV TK treatment and other prodrug activation combinations can be effective in purging tumor cells but not to the extent of entirely eradicating a cancer. That is, prodrug activation therapy must be used in combination with other cancer therapies.

Promoterless Gene Targeting

Endonucleases such as that of the CRISPR-Cas system may have significant adverse effects, including (but not limited to) immunogenicity, off-target cleavage and mutagenesis, and induction of unanticipated chromosomal aberrations. Introduction of a foreign gene (or cDNA) may lead to similar problems. When a vector-borne promoter, driving expression of the therapeutic transgene and/or the nuclease, is integrated into target cell chromosomal DNA, it may lead to unwanted activation of nearby genes, including oncogenes.

Recombinant AAV has been used in a number of clinical trials and, recently, has become a viral vector of choice because it lacks pathogenicity and is devoid of viral genes. Unfortunately, nontargeted genomic integration

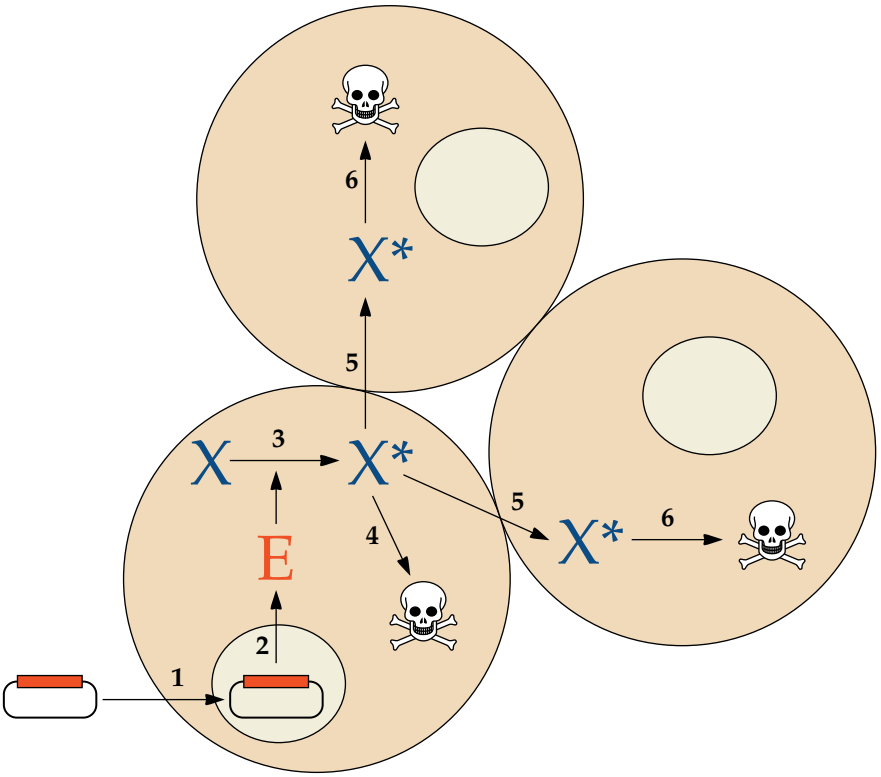
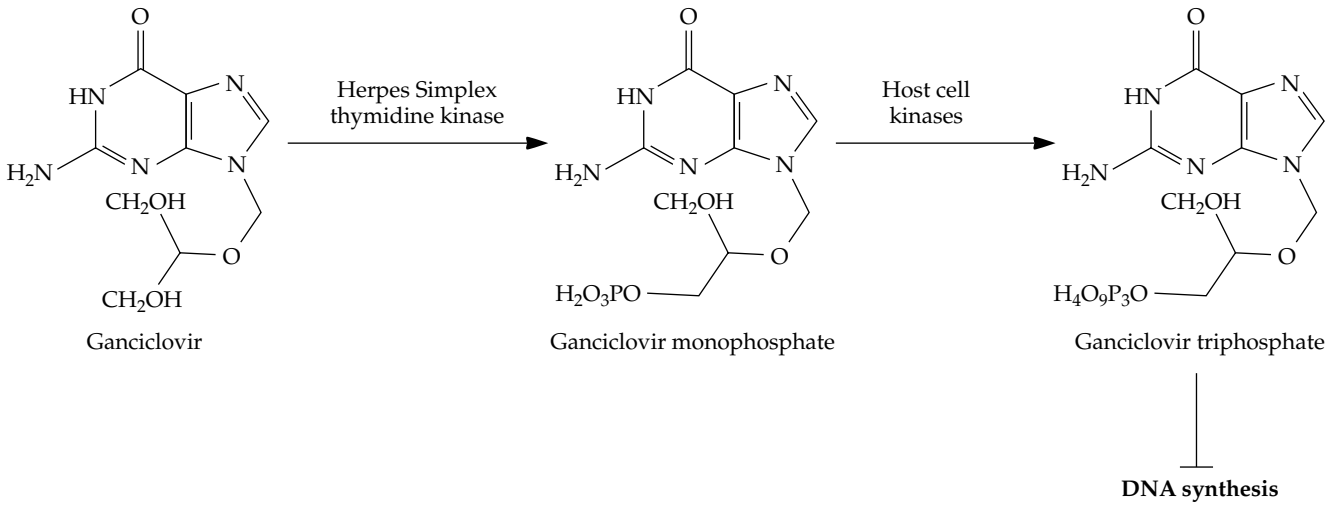


Figure 6.44 Overview of prodrug activation therapy. Tumor cells are either transfected or transduced with the gene (red) that encodes an enzyme (E) that activates a prodrug (X) **(1)**. The introduced prodrug-activating gene is expressed in a recipient cell **(2)**, and the enzyme converts the prodrug to its cytotoxic form (X*), which, in turn, destroys the cell (skull and crossbones) **(3** and **4)**. The activated prodrug (X*) enters adjacent cells **(5)** and destroys them **(6)**, creating a bystander effect.

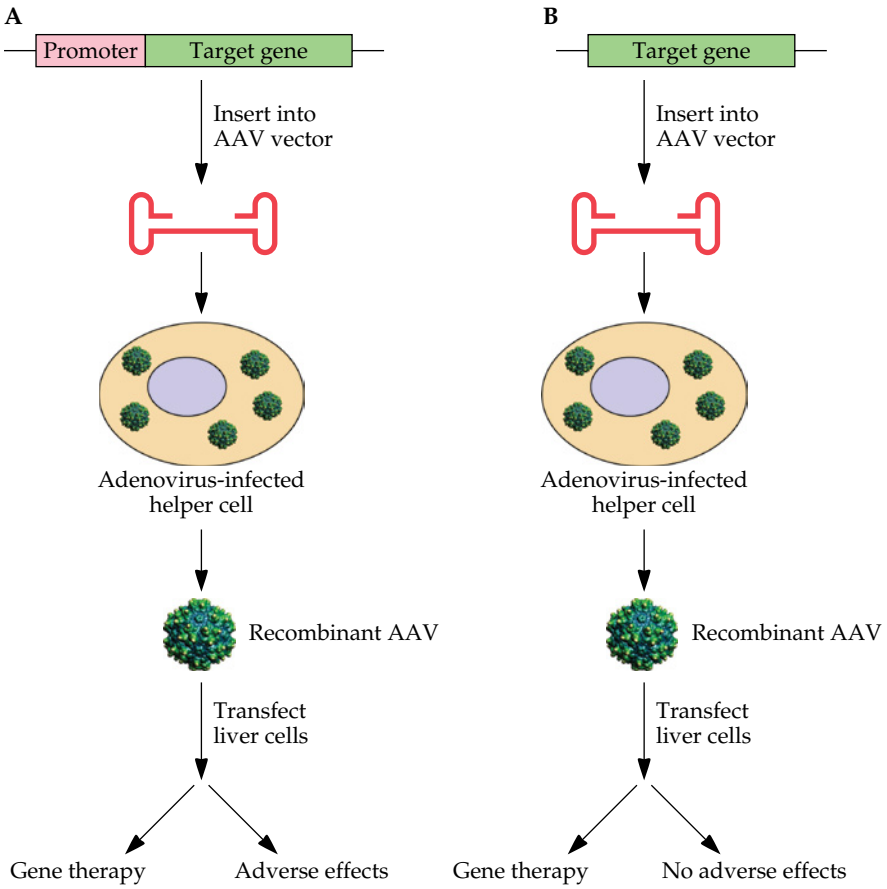
Figure 6.45 The conversion of ganciclovir to ganciclovir monophosphate by herpes simplex thymidine kinase and its subsequent conversion to ganciclovir triphosphate by host cell kinases, where ganciclovir triphosphate is an inhibitor of tumor cell DNA synthesis.



of recombinant AAV occurs at a measurable rate, and there are reports of such integrations inducing liver cancer in mice (Fig. 6.46A). One way to diminish the chance of activating oncogenes that are adjacent to or near the integrated target gene is to ensure that the introduced target gene does not encode a promoter (Fig. 6.46B). In this case, the introduced corrected gene can only be expressed if it has integrated at the correct site on the genome downstream of a promoter. The use of AAV helps to ensure this outcome since, as mentioned previously, different AAVs target different cell types.

To test this idea, scientists targeted the human F9 gene, which is defective in the X-linked recessive disease hemophilia B. The F9 gene encodes factor IX, a vitamin K-dependent plasma protein that participates in the intrinsic pathway of blood coagulation (see Fig. 5.45). In this case, AAV8 was used to target the promoterless gene to the liver of diseased mice. While the target gene integrated in only a limited number of sites, this experiment demonstrated that a therapeutic effect for *in vivo* gene targeting is possible without nucleases and without a vector-borne promoter. Moreover, the favorable safety profile of this strategy makes it a prime candidate for treating a range of genetic deficiencies.

Figure 6.46 Gene therapy with a promoterless gene (cDNA) delivered using an adenovirus-associated virus (AAV) vector. **(A)** When the target gene is under the control of a promoter introduced into the vector, adverse side effects may occur. **(B)** In the absence of a promoter, there are no adverse side effects.



summary

Several human disorders that result from the overproduction of a normal protein or from the presence of pathogens may be treated by using (i) nucleotide sequences that bind to a specific mRNA and prevent its translation, i.e., an antisense oligonucleotide; (ii) small RNA molecules, i.e., aptamers, that assume highly organized secondary and tertiary structures and bind tightly to and thereby inhibit the activity of a wide range of molecules, including proteins and amino acids; (iii) RNA or DNA sequences that bind and cleave specific RNA molecules, i.e., ribozymes or DNAzymes; (iv) small double-stranded RNA molecules that direct the sequence-specific degradation of mRNA, i.e., interfering RNAs; or (v) synthetic complexes of nanoparticles, RNase A, and oligodeoxyribonucleic acids that bind to and cleave specific mRNAs, i.e., nanozymes. In addition, zinc finger nucleases and the CRISPR-Cas nuclease system may be used to cleave DNA at target sites that is repaired to either disrupt the function of a gene or replace a defective gene. These techniques may all be used to either treat human genetic diseases or lessen or prevent diseases caused by pathogenic viruses and other disease-causing organisms.

The greatest impediment to the development of nucleic acid-based therapeutic agents is the difficulty in delivering these agents to their target tissue(s). Numerous viral systems have been developed for the delivery of therapeutic genes. Viral vectors take advantage of the ability of a virus to penetrate a specific cell, protect the therapeutic DNA from degradation, and direct it to the cell nucleus. Several viruses have been engineered for gene therapy applications. The key requirements for gene therapy are the efficient and safe delivery of a therapeutic gene to a specific tissue in sufficient numbers such that the introduced gene can alleviate a particular disorder. Somatic cell gene therapy is implemented in one of two ways: *in vivo* or *ex vivo*. For *in vivo* gene therapy, a therapeutic gene is introduced into a targeted tissue of the patient. *Ex vivo* gene therapy entails collecting and culturing stem cells from an affected individual, introducing the therapeutic gene into these cultured cells, growing the cells with the therapeutic gene, and then transplanting these cells back into the patient. In addition, special techniques have been developed to correct defective mitochondrial genomes. And bacteriophages have been developed as an alternative to antibiotics in order to control the proliferation of pathogenic bacteria.

In addition to the viral vectors mentioned above, nonviral approaches for the delivery of nucleic acid-based therapeutic agents include direct injection at the site of the pathology, packaging the nucleic acid into cationic liposomes, contained within nonpathogenic bacteria or bacterial minicells, and conjugating the nucleic acid to another molecule (e.g., lipid, cholesterol, antibody fragment, transposon, or aptamer).

REFERENCES

Acharya S, Mishra A, Paul D, Ansari AH, Azhar M, Kumar M, Rauthan R, Sharma N, Aich M, Sinha D, Sharma S, Jain S, Ray A, Jain S, Ramalingam S, Maiti S, Chakraborty D. 2019. *Francisella novicida* Cas9 interrogates genomic DNA with very high specificity and can be used for mammalian genome editing. *Proc Natl Acad Sci USA* **116:**20959–20968.

Aguiar S, van der Gaag B, Cortese FAB. 2017. RNAi mechanisms in Huntington's disease therapy: siRNA versus shRNA. *Transl Neurodegener* **6:**30.

Alemayehu D, Casey PG, McAuliffe O, Guinane CM, Martin JG, Shanahan F, Coffey A, Ross RP, Hill C. 2012. Bacteriophages fMR299-2 and fNH-4 can eliminate *Pseudomonas aeruginosa* in the murine lung and on cystic lung airways cells. *MBio* **3:**1–9.

Alexander LM, Oh J-H, Stapleton DS, Schueler KL, Keller MP, Attie AD, van Pijkeren J-P. 2019. Exploiting prophage-mediated lysis for biotherapeutic release by *Lactobacillus reuteri*. *Appl Environ Microbiol* **85:**e02335-18.

Anzalone AV, Randolph PB, Davis JR, Sousa AA, Koblan LW, Levy JM, Chen PJ, Wilson C, Newby GA, Raguram A, Liu DR. 2019. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576:**149–157.

Barzel A, Paulk NK, Shi Y, Huang Y, Chu K, Zhang F, Valdmanis PN, Spector LP, Porteus MH, Gaensler KM, Kay MA. 2015. Promoterless gene targeting without nucleases ameliorates haemophilia B in mice. *Nature* **517:**360–364.

Bikard D, Euler CW, Jiang W, Nussenzweig PM, Goldberg GW, Duportet X, Fischetti VA, Marraffini LA. 2014. Exploiting CRISPR-Cas nucleases to produce sequence-specific antimicrobials. *Nat Biotechnol* **32:**1146–1150.

Biscans A, Coles A, Haraszti R, Echeverria D, Hassler M, Osborn M, Khvorova A. 2019. Diverse lipid conjugates for functional extra-hepatic siRNA delivery *in vivo*. *Nucleic Acids Res* **47:**1082–1096.

Bramlage B, Luzi E, Eckstein F. 1998. Designing ribozymes for the inhibition of gene expression. *Trends Biotechnol* **16:**434–438.

Bunka DHJ, Stockley PG. 2006. Aptamers come of age - at last. *Nat Rev Microbiol* **4:**588–596.

Burmeister AR, Fortier A, Roush C, Lessing AJ, Bender RG, Barahman R, Grant R, Chan BK, Turner PE. 2020. Pleiotropy complicates a trade-off between phage resistance and antibiotic resistance. *Proc Natl Acad Sci USA* **117:**11207–11216.

Burnett JC, Rossi JJ. 2012. RNA-based therapeutics: current progress and future prospects. *Chem Biol* **19:**60–71.

Cafora M, Deflorian G, Forti F, Ferrari L, Binelli G, Briani F, Ghisotti D, Pistocchi A. 2019. Phage therapy against *Pseudomonas aeruginosa* infections in a cystic fibrosis zebrafish model. *Sci Rep* **9:**1527.

Cai H, Santiago FS, Prado-Lourenco L, Wang B, Patrikakis M, Davenport MP, Maghzal GJ, Stocker R, Parish CR, Chong BH, Lieschke GJ, Wong T-W, Chesterman CN, Francis DJ, Moloney FJ, Barnetson RSC, Halliday GM, Khachigian LM. 2012. DNAzyme targeting c-*jun* suppresses skin cancer growth. *Sci Transl Med* **4:**139ra82.

Cairns MJ, Hopkins TM, Witherington C, Wang L, Sun L-Q. 1999. Target site selection for an RNA-cleaving catalytic DNA. *Nat Biotechnol* **17:**480–486.

Cafora M, Deflorian G, Forti F, Ferrari L, Binelli G, Briani F, Ghisotti D, Pistocchi A. 2019. Phage therapy against *Pseudomonas aeruginosa* infections in a cystic fibrosis zebrafish model. *Sci Rep* **9:**1527.

Cathomen T, Joung JK. 2008. Zinc-finger nucleases: the next generation emerges. *Mol Ther* **16:**1200–1207.

Chen L, Rashid F, Shah A, Awan HM, Wu M, Liu A, Wang J, Zhu T, Luo Z, Shan G. 2015. The isolation of an RNA aptamer targeting to p53 protein with single amino acid mutation. *Proc Natl Acad Sci USA* **112:**10002–10007.

Chu TC, Twu KY, Ellington AD, Levy M. 2006. Aptamer mediated siRNA delivery. *Nucleic Acids Res* **34:**e73.

Cerchia L, de Franciscis V. 2010. Targeting cancer cells with nucleic acid aptamers. *Trends Biotechnol* **28:**517–525.

Cobb LH, Park J, Swanson EA, Beard MC, McCabe EM, Rourke AS, Seo KS, Olivier AK, Priddy LB. 2019. CRISPR-Cas9 modified bacteriophage for treatment of *Staphylococcus aureus* induced osteomyelitis and soft tissue infection. *PLoS One* **14:**e0220421.

Conde J, Artzi N. 2015. Are RNAi and miRNA therapeutics truly dead? *Trends Biotechnol* **33:**141–144.

Crooke ST, Geary RS. 2013. Clinical pharmacological properties of mipomersen (Kynamro), a second generation antisense inhibitor of apolipoprotein B. *Br J Clin Pharmacol* **76:**269–276.

Cullen BR. 2006. Enhancing and confirming the specificity of RNAi experiments. *Nat Methods* **3:**677–681.

Dedrick RM, Guerrero-Bustamante CA, Garlena RA, Russell DA, Ford K, Harris K, Gilmour KC, Soothill J, Jacobs-Sera D, Schooley RT, Hatfull GF, Spencer H. 2019. Engineered bacteriophages for treatment of a patient with a disseminated drug-resistant *Mycobacterium abscessus*. *Nat Med* **25:**730–733.

de Fougerolles A, Vornlocher H-P, Maraganore J, Lieberman J. 2007. Interfering with disease: a progress report on siRNA-based therapeutics. *Nat Rev Drug Discov* **6:**443–453.

Didigu CA, Wilen CB, Wang J, Duong J, Secreto AJ, Danet-Desnoyers GA, Riley JL, Gregory PD, June CH, Holmes MC, Doms RW. 2014. Simultaneous zinc-finger nuclease editing of the HIV coreceptors ccr5 and cxcr4 protects CD4+ T cells from HIV-1 infection. *Blood* **123:**61–69.

Dillon CP, Sandy P, Nencioni A, Kissler S, Rubinson DA, Van Parijs L. 2005. Rnai as an experimental and therapeutic tool to study and regulate physiological and disease processes. *Annu Rev Physiol* **67:**147–173.

Dong Y, Love KT, Dorkin JR, Sirirungruang S, Zhang Y, Chen D, Bogorad RL, Yin H, Chen Y, Vegas AJ, Alabi CA, Sahay G, Olejnik KT, Wang W, Schroeder A, Lytton-Jean AKR, Siegwart DJ, Akinc A, Barnes C, Barros SA, Carioto M, Fitzgerald K, Hettinger J, Kumar V, Novobrantseva TI, Qin J, Querbes W, Koteliansky V, Langer R, Anderson DG. 2014. Lipopeptide nanoparticles for potent and selective siRNA delivery in rodents and nonhuman primates. *Proc Natl Acad Sci USA* **111:**3955–3960.

Elbashir SM, Harborth J, Lendeckel W, Yalcin A, Weber K, Tuschl T. 2001. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411:**494–498.

Elbashir SM, Lendeckel W, Tuschl T. 2001. RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev* **15:**188–200.

Fichou Y, Férec C. 2006. The potential of oligonucleotides for therapeutic applications. *Trends Biotechnol* **24:**563–570.

Field AK. 1998. Viral targets for antisense oligonucleotides: a mini review. *Antiviral Res* **37:**67–81.

Furfaro LL, Payne MS, Chang BJ. 2018. Bacteriophage therapy: clinical trials and regulatory hurdles. *Front Cell Infect Microbiol* **8:**376.

Germer K, Leonard M, Zhang X. 2013. RNA aptamers and their therapeutic and diagnostic applications. *Int J Biochem Mol Biol* **4:**27–40.

Gopinath SCB, Lakshmipriya T, Chen Y, Arshad MKM, Kerishnan JP, Ruslinda AR, Al-Douri Y, Voon CH, Hashim U. 2016. Cell-targeting aptamers act as intracellular delivery vehicles. *Appl Microbiol Biotechnol* **100:**6955–6969.

Powell Gray B, Kelly L, Ahrens DP, Barry AP, Kratschmer C, Levy M, Sullenger BA. 2018. Tunable cytotoxic aptamer-drug conjugates for the treatment of prostate cancer. *Proc Natl Acad Sci USA* **115:**4761–4766.

Greenfield A, Braude P, Flinter F, Lovell-Badge R, Ogilvie C, Perry ACF. 2017. Assisted reproductive technologies to prevent human mitochondrial disease transmission. *Nat Biotechnol* **35:**1059–1068.

Gunaratne R, Kumar S, Frederiksen JW, Stayrook S, Lohrmann JL, Perry K, Bompiani KM, Chabata CV, Thalji NK, Ho MD, Arepally G, Camire RM, Krishnaswamy S, Sullenger BA. 2018. Combination of aptamer and drug for reversible anticoagulation in cardiopulmonary bypass. *Nat Biotechnol* **36:**606–613.

Gurney J, Brown SP, Kaltz O, Hochberg ME. 2020. Steering phages to combat bacterial pathogens. *Trends Microbiol* **28:**85–94.

Guo P, Coban O, Snead NM, Trebley J, Hoeprich S, Guo S, Shu Y. 2010. Engineering RNA for targeted siRNA delivery and medical application. *Adv Drug Deliv Rev* **62:** 650–666.

Haasnoot J, Westerhout EM, Berkhout B. 2007. RNA interference against viruses: strike and counterstrike. *Nat Biotechnol* **25:**1435–1443.

Harada LK, Silva EC, Campos WF, Del Fiol FS, Vila M, Dąbrowska K, Krylov VN, Balcão VM. 2018. Biotechnological applications of bacteriophages: state of the art. *Microbiol Res* **212-213:**38–58.

Hu G, Miao Y, Luo X, Chu W, Fu Y. 2020. Identification of a novel cell-penetrating peptide derived from the capsid protein of chicken anemia virus and its application in gene delivery. *Appl Microbiol Biotechnol* **104:**10503–10513.

Isaacs FJ, Dwyer DJ, Collins JJ. 2006. RNA synthetic biology. *Nat Biotechnol* **24:**545–554.

Izsvák Z, Hackett PB, Cooper LJN, Ivics Z. 2010. Translating Sleeping Beauty transposition into cellular therapies: victories and challenges. *BioEssays* **32:**756–767.

Kaiser J. 2020. Tweaking genes with CRISPR or viruses fixes blood disorders. *Science* **370:**1254–1255.

Kaur H, Bruno JG, Kumar A, Sharma TK. 2018. Aptamers in the therapeutics and diagnostics pipelines. *Theranostics* **8:**4016–4032.

Keefe AD, Pai S, Ellington A. 2010. Aptamers as therapeutics. *Nat Rev Drug Discov* **9:**537–550.

Kennedy EM, Cullen BR. 2017. Gene editing: a new tool for viral disease. *Annu Rev Med* **68:**401–411.

Khachigian LM. 2020. Deoxyribozymes as catalytic nanotherapeutic agents. *Cancer Res* 10.1158/0008-5472.CAN-18-2474.

Kim D-H, Behlke MA, Rose SD, Chang M-S, Choi S, Rossi JJ. 2005. Synthetic dsRNA Dicer substrates enhance RNAi potency and efficacy. *Nat Biotechnol* **23:**222–226.

Kim Y-G, Cha J, Chandrasegaran S. 1996. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc Natl Acad Sci USA* **93:**1156–1160.

Kobayashi-Ishihara M, Terahara K, Martinez JP, Yamagishi M, Iwabuchi R, Brander C, Ato M, Watanabe T, Meyerhans A, Tsunetsugu-Yokota Y. 2018. HIV LTR-driven antisense RNA by itself has regulatory function and may curtail virus reactivation from latency. *Front Microbiol* **9:**1066.

Kole R, Krainer AR, Altman S. 2012. RNA therapeutics: beyond RNA interference and antisense oligonucleotides. *Nat Rev Drug Discov* **11:**125–140.

Kono K, Torikoshi Y, Mitsutomi M, Itoh T, Emi N, Yanagie H, Takagishi T. 2001. Novel gene delivery systems: complexes of fusigenic polymer-modified liposomes and lipoplexes. *Gene Ther* **8:**5–12.

Lares MR, Rossi JJ, Ouellet DL. 2010. RNAi and small interfering RNAs in human disease therapeutic applications. *Trends Biotechnol* **28:**570–579.

Lee J-H, Canny MD, De Erkenez A, Krilleke D, Ng Y-S, Shima DT, Pardi A, Jucker F. 2005. A therapeutic aptamer inhibits angiogenesis by specifically targeting the heparin binding domain of VEGF165. *Proc Natl Acad Sci USA* **102:**18902–18907.

Leng Q, Woodle MC, Mixson AJ. 2017. Targeted delivery of siRNA therapeutics to malignant tumors. *J Drug Deliv* **2017:**6971297.

Lima WF, Prakash TP, Murray HM, Kinberger GA, Li W, Chappell AE, Li CS, Murray SF, Gaus H, Seth PP, Swayze EE, Crooke ST. 2012. Single-stranded siRNAs activate RNAi in animals. *Cell* **150:**883–894.

Liu MM, Tuo J, Chan CC. 2011. Gene therapy for ocular diseases. *Br J Ophthalmol* **95:**604–612.

Long C, McAnally JR, Shelton JM, Mireault AA, Bassel-Duby R, Olson EN. 2014. Prevention of muscular dystrophy in mice by CRISPR/Cas9-mediated editing of germline DNA. *Science* **345:**1184–1188.

Lu TK, Collins JJ. 2009. Engineered bacteriophage targeting gene networks as adjuvants for antibiotic therapy. *Proc Natl Acad Sci USA* **106:**4629–4634.

Ma C-C, Wang Z-L, Xu T, He Z-Y, Wei Y-Q. 2020. The approved gene therapy drugs worldwide: from 1998 to 2019. *Biotechnol Adv* **40:**107502.

Maloney DM, Chadderton N, Millington-Ward S, Palfi A, Shortall C, O'Byrne JJ, Cassidy L, Keegan D, Humphries P, Kenna P, Farrar GJ. 2020. Optimized OPA1 isoforms 1 and 7 provide therapeutic benefit in models of mitochondrial dysfunction. *Front Neurosci* **14:**571479.

Matzen K, Elzaouk L, Matskevich AA, Nitzsche A, Heinrich J, Moelling K. 2007. RNase H-mediated retrovirus destruction *in vivo* triggered by oligodeoxynucleotides. *Nat Biotechnol* **25:**669–674.

McNamara JO II, Andrechek ER, Wang Y, Viles KD, Rempel RE, Gilboa E, Sullenger BA, Giangrande PH. 2006. Cell type-specific delivery of siRNAs with aptamer-siRNA chimeras. *Nat Biotechnol* **24:**1005–1015.

Nimjee SM, Rusconi CP, Sullenger BA. 2005. Aptamers: an emerging class of therapeutics. *Annu Rev Med* **56:**555–583.

Oliveira S, Storm G, Schiffelers RM. 2006. Targeted delivery of siRNA. *J Biomed Biotechnol* **2006:**63675.

O'Rorke S, Keeney M, Pandit A. 2010. Non-viral polyplexes: scaffold mediated delivery for gene therapy. *Prog Polym Sci* **35:**441–458.

Osborn MF, Coles AH, Biscans A, Haraszti RA, Roux L, Davis S, Ly S, Echeverria D, Hassler MR, Godinho BMDC, Nikan M, Khvorova A. 2019. Hydrophobicity drives the systemic distribution of lipid-conjugated siRNAs via lipid transport pathways. *Nucleic Acids Res* **47:**1070–1081.

Parker CL, Jacobs TM, Huckaby JT, Harit D, Lai SK. 2020. Efficient and highly specific gene transfer using mutated lentiviral vectors redirected with bispecific antibodies. *MBio* **11:**e02990–e19.

Pendergrast PS, Marsh HN, Grate D, Healy JM, Stanton M. 2005. Nucleic acid aptamers for target validation and therapeutic applications. *J Biomol Tech* **16:**224–234.

Peng H, Borg RE, Dow LP, Pruitt BL, Chen IA. 2020. Controlled phage therapy by photothermal ablation of specific bacterial species using gold nanorods targeted by chimeric phages. *Proc Natl Acad Sci USA* **117:**1951–1961.

Pires DP, Vilas Boas D, Sillankorva S, Azeredo J. 2015. Phage therapy: a step forward in the treatment of *Pseudomonas aeruginosa* infections. *J Virol* **89:**7449–7456.

Pires DP, Cleto S, Sillankorva S, Azeredo J, Lu TK. 2016. Genetically engineered phages: a review of advances over the decade. *Microbiol Mol Biol Rev* **80:**523–543.

Radom F, Jurek PM, Mazurek MP, Otlewski J, Jeleń F. 2013. Aptamers: molecules of great potential. *Biotechnol Adv* **31:**1260–1274.

Ran FA, Cong L, Yan WX, Scott DA, Gootenberg JS, Kriz AJ, Zetsche B, Shalem O, Wu X, Makarova KS, Koonin EV, Sharp PA, Zhang F. 2015. *In vivo* genome editing using *Staphylococcus aureus* Cas9. *Nature* **520:**186–191.

Rossi JJ. 2004. Medicine: a cholesterol connection in RNAi. *Nature* **432:**155–156.

Rossi JJ. 2005. Receptor-targeted siRNAs. *Nat Biotechnol* **23:**682–684.

Sah H, Thoma LA, Desu HR, Sah E, Wood GC. 2013. Concepts and practices used to develop functional PLGA-based nanoparticulate systems. *Int J Nanomedicine* **8:**747–765.

Sander JD, Joung JK. 2014. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol* **32:**347–355.

Seow Y, Wood MJ. 2009. Biological gene delivery vehicles: beyond viral vectors. *Mol Ther* **17:**767–777.

Setten RL, Rossi JJ, Han SP. 2019. The current state and future directions of RNAi-based therapeutics. *Nat Rev Drug Discov* **18:**421–446.

Shahi S, Shanmugasundaram GK, Banerjea AC. 2001. Ribozymes that cleave reovirus genome segment S1 also protect cells from pathogenesis caused by reovirus infection. *Proc Natl Acad Sci USA* **98:**4101–4106.

Shim MS, Kwon YJ. 2010. Efficient and targeted delivery of siRNA in vivo. *FEBS J* **277:**4814–4827.

Siolas D, Lerner C, Burchard J, Ge W, Linsley PS, Paddison PJ, Hannon GJ, Cleary MA. 2005. Synthetic shRNAs as potent RNAi triggers. *Nat Biotechnol* **23:**227–231.

Snøve O Jr, Rossi JJ. 2006. Expressing short hairpin RNAs *in vivo*. *Nat Methods* **3:**689–695.

Song E, Zhu P, Lee S-K, Chowdhury D, Kussman S, Dykxhoorn DM, Feng Y, Palliser D, Weiner DB, Shankar P, Marasco WA, Lieberman J. 2005. Antibody mediated *in vivo* delivery of small interfering RNAs via cell-surface receptors. *Nat Biotechnol* **23:**709–717.

Soutschek J, Akinc A, Bramlage B, Charisse K, Constien R, Donoghue M, Elbashir S, Geick A, Hadwiger P, Harborth J, John M, Kesavan V, Lavine G, Pandey RK, Racie T, Rajeev KG, Röhl I, Toudjarska I, Wang G, Wuschko S, Bumcrot D, Koteliansky V, Limmer S, Manoharan M, Vornlocher H-P. 2004. Therapeutic silencing of an endogenous gene by systemic administration of modified siRNAs. *Nature* **432:**173–178.

Spraggon L, Cartegni L. 2013. Antisense modulation of RNA processing as a therapeutic approach in cancer therapy. *Drug Discov Today Ther Strateg* **10:**e139–e148.

Sundaram P, Kurniawan H, Byrne ME, Wower J. 2013. Therapeutic RNA aptamers in clinical trials. *Eur J Pharm Sci* **48:**259–271.

Taylor AI, Pinheiro VB, Smola MJ, Morgunov AS, Peak-Chew S, Cozens C, Weeks KM, Herdewijn P, Holliger P. 2015. Catalysts from synthetic genetic polymers. *Nature* **518:**427–430.

Tse LV, Klinc KA, Madigan VJ, Castellanos Rivera RM, Wells LF, Havlik LP, Smith JK, Agbandje-McKenna M, Asokan A. 2017. Structure-guided evolution of antigenically distinct adeno-associated virus variants for immune evasion. *Proc Natl Acad Sci USA* **114:**E4812–E4821.

Urnov FD, Rebar EJ, Holmes MC, Zhang HS, Gregory PD. 2010. Genome editing with engineered zinc finger nucleases. *Nat Rev Genet* **11:**636–646.

van Rij RP, Andino R. 2006. The silent treatment: RNAi as a defense against virus infection in mammals. *Trends Biotechnol* **24:**186–193.

Wang D, Tai PWL, Gao G. 2019. Adeno-associated virus vector as a platform for gene therapy delivery. *Nat Rev Drug Discov* **18:**358–378.

Wang T, Chen C, Larcher LM, Barrero RA, Veedu RN. 2019. Three decades of nucleic acid aptamer technologies: lessons learned, progress and opportunities on aptamer development. *Biotechnol Adv* **37:**28–50.

Wang Z, Liu H, Yang SH, Wang T, Liu C, Cao YC. 2012. Nanoparticle-based artificial RNA silencing machinery for antiviral therapy. *Proc Natl Acad Sci USA* **109:**12387–12392.

Wolfrum C, Shi S, Jayaprakash KN, Jayaraman M, Wang G, Pandey RK, Rajeev KG, Nakayama T, Charrise K, Ndungo EM, Zimmermann T, Koteliansky V, Manoharan M, Stoffel M. 2007. Mechanisms and optimization of *in vivo* delivery of lipophilic siRNAs. *Nat Biotechnol* **25:**1149–1157.

Xiang S, Fruehauf J, Li CJ. 2006. Short hairpin RNA-expressing bacteria elicit RNA interference in mammals. *Nat Biotechnol* **24:**697–702.

Xu C-F, Wang J. 2015. Delivery systems for siRNA drug development in cancer therapy. *Asian J Pharm Sci* **10:**1–12.

Yu D, Pendergraff H, Liu J, Kordasiewicz HB, Cleveland DW, Swayze EE, Lima WF, Crooke ST, Prakash TP, Corey DR. 2012. Single-stranded RNAs use RNAi to potently and allele-selectively inhibit mutant huntingtin expression. *Cell* **150:**895–908.

Zhang S, Wang J, Wang J. 2020. One-day TALEN assembly protocol and a dual-tagging system for genome editing. *ACS Omega* **5:**19702–19714.

Zhao T, Wei T, Guo J, Wang Y, Shi X, Guo S, Jia X, Jia H, Feng Z. 2019. PD-1-siRNA delivered by attenuated *Salmonella* enhances the antimelanoma effect of pimozide. *Cell Death Dis* **10:**164.

Zheng JH, Min J-J. 2016. Targeted cancer therapy using engineered *Salmonella* Typhimurium. *Chonnam Med J* **52:** 173–184.

Zhou J, Bobbin ML, Burnett JC, Rossi JJ. 2012. Current progress of RNA aptamer-based therapeutics. *Front Genet* **3:**234.

Zhou L-Y, Qin Z, Zhu Y-H, He Z-Y, Xu T. 2019. Current RNA-based therapeutics in clinical trials. *Curr Gene Ther* **19:**172–196.

Zhou W, Ding J, Liu J. 2017. Theranostic DNAzymes. *Theranostics* **7:**1010–1025.

Zipkin M. 2020. Big pharma buys into exosomes for drug delivery. *Nat Biotechnol* **38:**1226–1228.

review questions

1. What is *ex vivo* gene therapy? What is *in vivo* gene therapy?
2. How can antisense oligonucleotides be used as a therapeutic agent?
3. Provide an example of how an antisense cDNA can be used as a therapeutic agent.
4. What is an aptamer, how is it selected, and how is it used as a therapeutic agent?
5. How might aptamers be used instead of antibodies or antibody fragments?
6. How can the simultaneous use of two or more aptamers be used to prevent mutations in aptamer binding sites from decreasing aptamer therapeutic effectiveness?
7. How can aptamers be used to prevent the formation of dangerous blood clots following bypass surgery?
8. What are ribozymes, and how can they be used as human therapeutic agents?
9. What are DNAzymes, and how do they compare to ribozymes?
10. What are interfering RNAs, and how might they be used as human therapeutic agents?
11. How can the progression of age-related macular degeneration be limited using RNA therapeutics?
12. How can zinc finger nucleases be used to repair a cell's defective genes?
13. How can the CRISPR-Cas9 system be used to repair a cell's defective genes?
14. How can the propensity of the cell's repair system to sometimes add incorrect nucleotides using the CRISPR-Cas9 system be overcome?
15. How might bacteriophage be used as a therapeutic agent?
16. How can a bacteriophage be modified to increase its killing of pathogenic agents?
17. A variety of novel delivery strategies have been developed to introduce nucleic acids to their cellular targets. These include the use of lipids, dendrimers, antibodies, aptamers, nanoparticles, and lipopeptide nanoparticles. Briefly explain how these strategies may be used in practice.
18. Describe how the prodrug strategy works.
19. How are bacteria and bacterial minicells used to specifically deliver therapeutic nucleic acids to their cellular targets?
20. Discuss how transposons may be used as gene delivery agents.
21. Discuss a procedure that you might use to isolate a specific RNA aptamer against a mutant protein with a single amino acid substitution compared to the native protein.
22. Describe a noncovalent complex that you might use an aptamer to deliver an siRNA to its target cells.
23. Why might two or more ribozymes targeted against the same mRNA be utilized simultaneously?
24. What is a CRISPR-Cas complex, and how might it be used to cleave a target DNA?
25. How might a CRISPR-Cas combination be used to kill antibiotic-resistant pathogenic bacteria?
26. What is a nanozyme, and how does it work?
27. How might you correct a mutation (on the mouse X chromosome) that causes Duchenne muscular dystrophy?
28. Describe a procedure that you might use to correct a mitochondrial mutation in the progeny of a woman who has this mutation.
29. When introducing corrective genes (cDNAs) into a human cell, what strategy might you use to prevent the activation of adjacent or nearby oncogenes?

Vaccination
Current and Future Vaccines
Subunit and Peptide Vaccines
Herpes Simplex Virus
Bovine Herpes Virus-1
Cholera
Influenza
SARS
COVID-19
Staphylococcus aureus
Human Papillomavirus
Foot-and-Mouth Virus
Streptococcus
Peptides
Malaria
Delivery
Genetic Immunization: DNA Vaccines
Delivery
Cancer
Zika Virus
Dental Caries
Engineered Attenuated Vaccines
Herpes Simplex Virus
Cholera
Salmonella Species
Leishmania Species
Vector Vaccines
Vaccines Directed against Viruses
Vaccines Directed against Bacteria
Bacteria as Antigen Delivery Systems
Monoclonal Antibody Passive Immunity
Influenza virus
SUMMARY
REFERENCES
REVIEW QUESTIONS

7 Vaccines

Vaccination

Vaccination can protect a recipient from various pathogenic agents by establishing an immunological resistance to infection. An injected or oral vaccine induces the host to generate antibodies against the disease-causing organism; therefore, during subsequent exposures, the infectious agent is inactivated (neutralized or killed), its proliferation is prevented, and the disease state is not established.

Over 220 years ago, in 1796, Edward Jenner experimentally tested the folklore-based notion that human infection with the agent that caused a mild cattle disease called cowpox would protect infected individuals against the human disease smallpox. Smallpox is an extremely virulent disease with a high death rate. Jenner inoculated James Phipps, an 8-year-old boy, with exudate from a cowpox pustule. In two separate trials after the initial vaccination, the boy was fully protected against human smallpox.

With the advent of vaccination, antibiotics, and effective public health measures, a large number of epidemic diseases have, for the most part, been brought under control (Table 7.1). Today, millions of people are alive and well due to the more than 85 vaccines that are now licensed for use against a wide range of pathogenic agents. It is estimated that vaccines currently prevent 2 to 3 million deaths per year worldwide. The causative agents of diphtheria, measles, mumps, poliomyelitis, rubella, smallpox, and others gave rise to more than 39 million infections in the 20th century in Canada and the United States. Vaccines have since reduced considerably the frequencies of these diseases (typically by a thousand-fold or more). This public health intervention owes its success not only to the

Molecular Biotechnology: Principles and Applications of Recombinant DNA,
Sixth Edition. Bernard R. Glick and Cheryl L. Patten.
© 2022 American Society for Microbiology.

Table 7.1 Annual cases in Canada from various diseases before and after the introduction of vaccines against the causative agents of the diseases

Disease	Annual no. of cases before vaccine was introduced	No. of cases in 2002	Decrease of disease incidence following immunization (%)
Polio	20,000	0	100
Diphtheria	9,000	0	100
Rubella	69,000	16	99.9
Mumps	52,000	197	99.6
Haemophilus influenzae type b infection	2,000	48	97.6
Whooping cough	25,000	2,557	90.0
Measles	300,000	7	99.99

development of effective vaccines but also to an efficient infrastructure for vaccine production, regulatory and safety procedures, and vaccine delivery. Without question, vaccines are the most economical and effective way to protect against serious epidemics. Most existing vaccines target viral diseases; this is because bacterial diseases are often treated with antibiotics, and vaccines directed against parasitic diseases are generally still at an early stage of development. However, as antibiotic-resistant bacteria become more of a public health problem, additional efforts are being directed toward the development of vaccines against pathogenic bacteria.

For many human and animal diseases, there are no vaccines. Today, more than two billion humans suffer from diseases that theoretically could be curtailed by vaccination. However, when protective measures are unavailable, devastating new outbreaks can occur. In 2014, an Ebola outbreak struck Liberia, Guinea, Sierra Leone, Nigeria, the United States, and Mali, producing, over the next 2 years, more than 30,000 infections and more than 11,000 deaths. In addition, new diseases for which vaccines might be useful continue to emerge. For example, in 2020, the world experienced a global pandemic in which >60 million people became infected with the newly identified coronavirus SARS CoV-2, with >1.5 million deaths resulting. The Global Alliance for Vaccines and Immunization estimates that more than 1.5 million children die every year from vaccine-preventable diseases. Highly effective vaccines are also not yet available for many leading diseases, including human immunodeficiency virus (HIV)/AIDS, tuberculosis, and malaria, which incur more than 5 million deaths globally each year. However, as discussed later in this chapter, considerable progress has been made in efforts to develop vaccines against these diseases.

In recent years, in some more developed countries, a small, but vocal, minority have refused to have their children vaccinated. These individuals argue that many of the previously common illnesses have been vanquished, and they fear the potential side effects of the vaccinations (most not supported by scientific evidence) more than the disease itself. In addition, many of these people question modern medicine and instead prefer to rely upon traditional or natural therapies. In fact, the small number of individuals who are not vaccinated benefit from the immunization of the vast majority of other people in society, which prevents the spread of the pathogen through a community. However, in communities where vaccination levels decrease below a certain threshold level, there is a real danger of some traditional diseases (whose incidence is currently quite low) making a comeback. For example, in March 2015, according to the

U.S. Centers for Disease Control and Prevention (CDC), 188 people from 24 states and the District of Columbia were reported to have measles. Most of those cases were linked to an amusement park in California. At the same time, Canadian public health officials reported that 119 cases of measles in the Province of Quebec were linked to the same amusement park in California and that all of the confirmed cases included individuals who were not vaccinated.

Measles symptoms typically develop 10 to 12 days after exposure to an infected person and generally last 7 to 10 days. These symptoms commonly include fever, high temperature, cough, runny nose, and inflamed eyes. In addition, individuals infected with the measles virus often display a red rash that usually covers the entire body. While the immediate risk of death from measles is relatively low—it occurs in about 0.2% of infected individuals—following a measles infection, individuals are much more susceptible to a host of other pathogens. This is because the measles virus can eliminate 11 to 73% of the antibody repertoire of an infected individual. Thus, the disease susceptibility and morbidity of individuals who develop measles are much greater than that of other individuals in the population.

Notwithstanding the devastation brought about by the worldwide COVID-19 pandemic, even before a vaccine for this virus existed, the small but seemingly well-organized antivaccination movement had begun campaigning against a COVID-19 vaccine. These individuals, using social media, claim (without any evidence) that when they are available, the vaccines will kill millions of people. While overall, most people support the use of vaccines, this sort of opposition could undermine efforts to establish herd immunity to the new coronavirus.

Current and Future Vaccines

The majority of current vaccines consist of either a killed (inactivated) (Fig. 7.1A) or a live, nonvirulent (attenuated) (Fig. 7.1B) form of an infectious agent. Traditionally, the infectious agent is grown in culture, purified, and either inactivated or attenuated without losing the ability to evoke an immune response that is effective against the virulent form of the pathogen. Notwithstanding the considerable success that has been achieved in creating effective vaccines against diseases such as German measles, diphtheria, whooping cough, tetanus, smallpox, and poliomyelitis, there are a number of limitations to the traditional mode of vaccine production.

- Not all infectious agents can be grown in culture, so no vaccines have been developed for a number of diseases.
- Production of animal and human viruses requires animal cell culture, which is expensive.
- Both the yield and rate of production of animal and human viruses in culture are often quite low, making vaccine production costly.
- Extensive safety precautions are necessary to ensure that laboratory and production personnel are not exposed to a pathogenic agent.
- On rare occasions, batches of vaccine may not be killed or may be insufficiently attenuated during the production process, thereby introducing virulent organisms into the vaccine and inadvertently spreading the disease.
- Attenuated strains may revert, a possibility that requires continual testing to ensure that the reacquisition of virulence has not occurred.

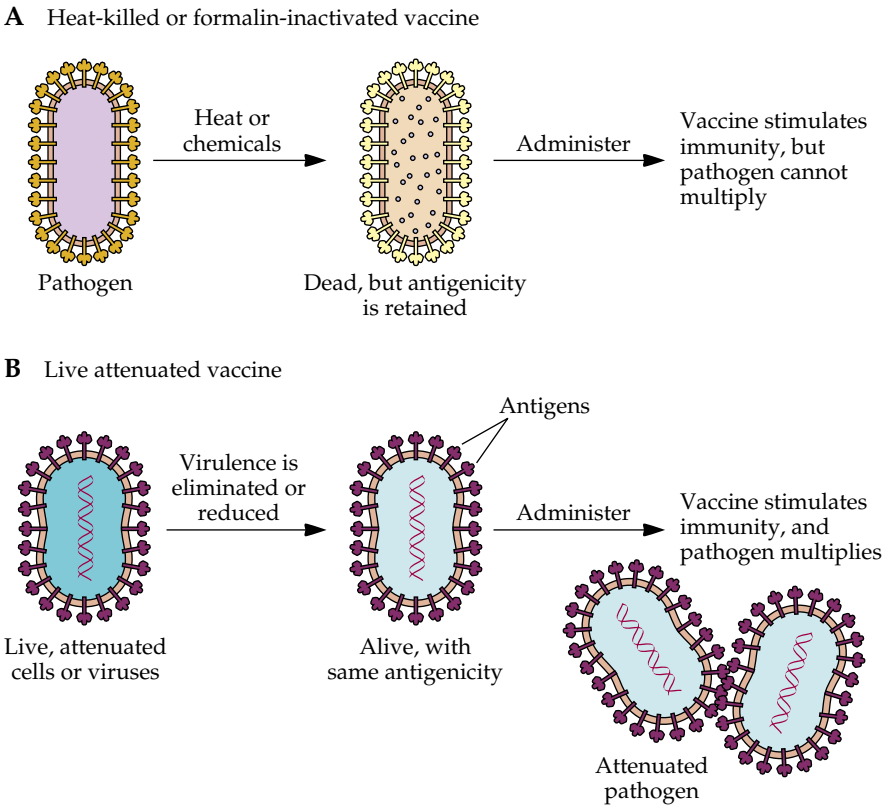


Figure 7.1 Inactivated and attenuated forms of a vaccine. The main difference between inactivated and attenuated forms of a vaccine is that while an inactivated vaccine stimulates immunity and cannot multiply, a live attenuated vaccine can both stimulate immunity and multiply.

- Not all diseases are preventable through the use of traditional vaccines.
- Most traditional vaccines have a limited shelf life and often require refrigeration to maintain potency. This requirement creates storage problems in countries with large, unelectrified rural areas.

Within the last 2 to 3 decades, molecular biotechnology has provided a means of creating a new generation of vaccines that overcome the drawbacks of traditional vaccines. The availability of gene cloning has enabled researchers to contemplate various novel strategies for vaccine development.

- Virulence genes may be deleted from an infectious agent that retains the ability to stimulate an immunological response. In this case, the genetically engineered agent may be used as a live vaccine without concern about reversion to virulence, because it is not possible for a whole gene to be reacquired spontaneously during growth in culture.
- Live nonpathogenic carrier systems that carry antigenic determinants of an unrelated pathogenic agent may be created. In this form, the carrier system facilitates the induction of a strong immunological response directed against the pathogenic agent.
- For infectious agents that cannot be maintained in culture, the genes for the proteins that have critical antigenic determinants can be isolated, cloned, and expressed in an alternative host system, such as

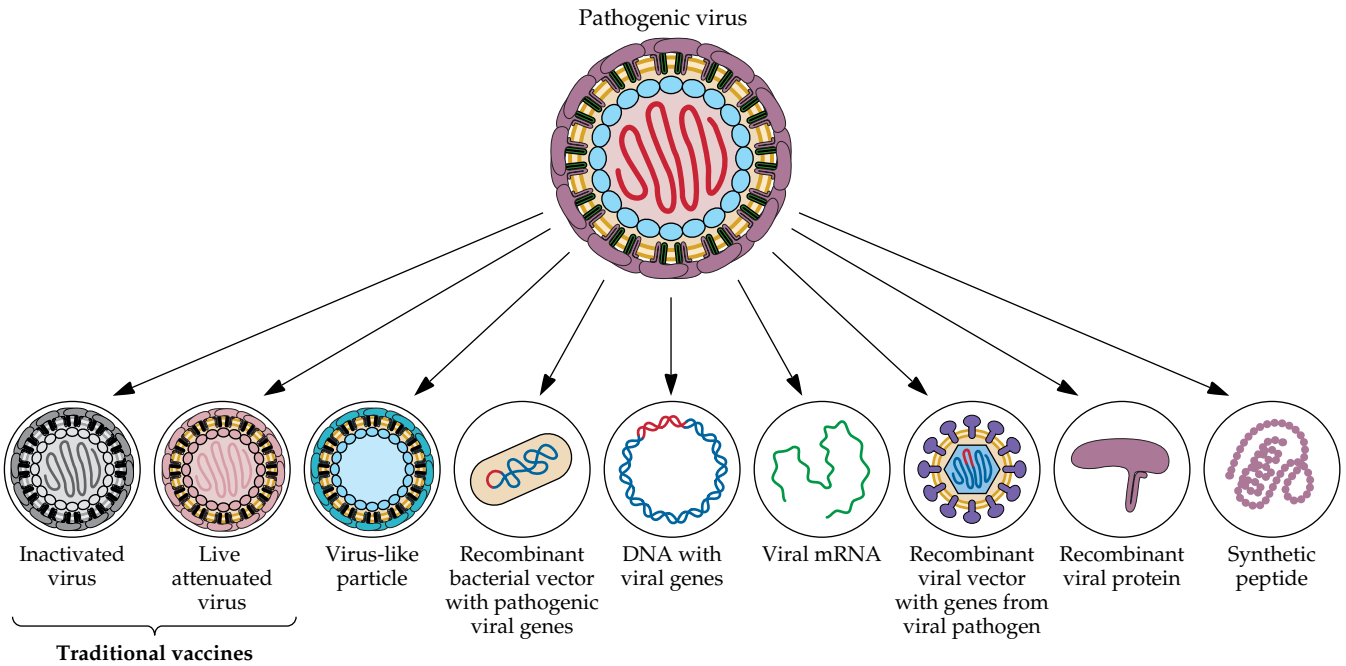
Escherichia coli, yeast, or a mammalian cell line. The proteins from the cloned genes can be formulated into a vaccine.

- The nucleic acids encoding key elements of an infectious agent may also be used as a vaccine.

There are a relatively wide range of approaches that are used and considered in the development of recombinant vaccines. These techniques include (i) the use of a portion of the pathogenic agent in a subunit vaccine, e.g., a specific protein or a virus-like particle, (ii) a small synthetic peptide that mimics the immunogenicity of the whole pathogen, (iii) a portion of the pathogen's DNA (or cDNA) or mRNA that encodes a key element in the pathogen's functioning, or (iv) a portion of the pathogen's DNA (or cDNA) carried on either a bacterial or viral vector (Fig. 7.2). Examples of each of these types of vaccines are discussed later in this chapter.

Because of less stringent regulatory requirements, the first vaccines that were produced using recombinant DNA techniques were for animal diseases, such as foot-and-mouth disease, rabies, and scours, a diarrheal disease of pigs and cattle. In addition, many more animal vaccines are currently being developed. For human diseases, a large number of recombinant vaccines are currently in various stages of development, including some that have been recently approved and others that are in clinical trials. The global recombinant vaccine market was estimated to be ~$9 billion in 2018 and is projected to reach ~$21 billion by 2026; however, this estimate is independent of the market for various COVID-19 vaccines that are currently being developed. The 2021 projected revenue from Pfizer's COVID-19 mRNA vaccine is ~$33.5 billion.

Figure 7.2 Schematic overview of different types of vaccines that may be directed against human or animal viruses. Traditional vaccines include both inactivated (killed) and live attenuated viruses. Subunit vaccines include recombinant viral proteins and virus-like particles. Also shown are peptide vaccines, DNA and RNA vaccines, and the use of bacterial and recombinant viral vectors to express genes from a viral pathogen.



Unfortunately, in comparison to the many hundreds of new therapeutic agents, only a limited number of recombinant vaccines have been commercialized. Why, according to the vaccine producers, does it take so long for many new vaccines to come to the marketplace? First, while there were 25 major vaccine manufacturers worldwide in 1970, in 2020 there were only 12. Second, some companies argue that there is limited financial incentive to develop new vaccines; in 2018, the worldwide market for all types of preventive vaccines was approximately $38.4 billion, with most of the market being for traditional vaccines (this amount is much smaller than the size of the therapeutics market discussed in chapters 5 and 6). Third, the United States and other governments are major purchasers of vaccines, forcing lower prices and thereby decreasing the potential profit. Fourth, the transition from conventional to newer processes for vaccine production is expensive and time-consuming (not including the large amount of time and money that are required for clinical trials), so that it is preferable to continue using a more established technology. On the positive side, the focus of the larger vaccine manufacturers on large-scale products has provided smaller biotechnology companies with a number of niche opportunities to develop and market new products. Finally, since most vaccines are intended to protect large populations, the very large amounts of money that companies often charge to treat a single individual with some of the newer therapeutic agents are unrealistic for the pricing of a new vaccine. Ironically, it is precisely in many poorer countries, where most individuals cannot afford to pay very much for treatment or immunization, that vaccines are needed the most.

Subunit and Peptide Vaccines

Traditional vaccines generally consist of either killed or attenuated forms of the pathogenic agent. The antibodies elicited by these vaccines initiate an immune response to inactivate (neutralize) pathogenic organisms by recognizing and binding to proteins on the outer surface of the agent. So, do vaccines need to contain the whole organism, or will specific molecules of pathogenic organisms suffice? For disease-causing viruses, purified outer surface viral proteins, either capsid or envelope proteins (Fig. 7.3), are often sufficient for eliciting neutralizing antibodies in the host organism. Vaccines that use components of a pathogenic organism rather than the whole organism are called "subunit" vaccines.

Safe and well-characterized subunit vaccines offer the potential to target immune responses toward specific epitopes. Since they are composed of well-defined components, they enable the production of custom vaccines beyond that provided by traditional whole-attenuated-pathogen approaches. Using recombinant DNA technology, the generation of libraries of different vaccine components permits the selection of vaccine constituents to stimulate immune responses that yield protective immunity against a specific pathogen. In addition to the identification of antigens and their components that can elicit protective or therapeutic immunity, the use of appropriate adjuvants that stimulate potent, antigen-specific immune responses is also important. The development of new subunit vaccine components is a rather important area of investigation upon which much future vaccine development will be based.

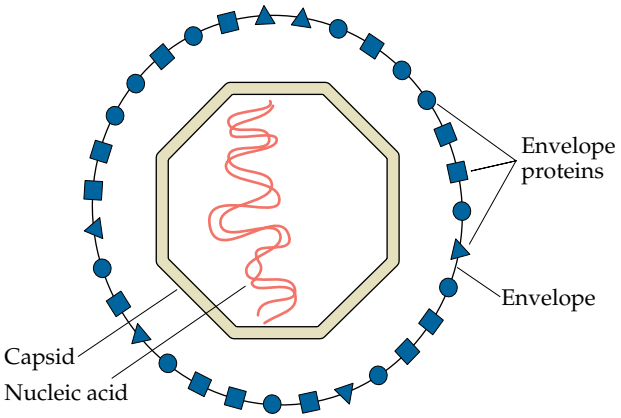


Figure 7.3 Schematic representation of an animal virus. Viruses generally consist of a relatively small nucleic acid genome (3 to 200 kb of either double- or single-stranded DNA or RNA) within a viral protein capsid that is sometimes, depending on the virus, surrounded by a protein-containing viral envelope (membrane).

There are advantages and disadvantages to the use of subunit vaccines. On the positive side, using a purified protein(s) as an immunogen ensures that the preparation is stable and safe, is precisely defined chemically, and is free of extraneous proteins and nucleic acids that can initiate undesirable side effects in the host organism. On the negative side, purification of a specific protein can be costly, and in certain instances, an isolated protein may not have the same conformation as it does *in situ* (within the viral capsid or envelope), with the result that its antigenicity (immunogenicity) is decreased. Obviously, the decision to produce a subunit vaccine depends on an assessment of both biological and economic factors.

Herpes Simplex Virus

Herpes simplex viruses 1 and 2 (HSV-1 and HSV-2) are, respectively, the causes of cold sores and genital herpes. HSV-1 and -2 persist in the body by becoming latent and hiding from the immune system in the cell bodies of neurons. Episodically, new virions are produced that cause lytic infections in the skin around the mouth or the genital epithelium. It is estimated that there are currently around 500 million people worldwide that are infected with HSV-2, with approximately 20 million new cases occurring annually. HSV has also been implicated as a cancer-causing (oncogenic) agent, in addition to its more common roles in causing sexually transmitted disease, severe eye infections, and encephalitis. Therefore, prevention of HSV infection by vaccination with either killed or attenuated virus may potentially put the recipient at risk for cancer. Current evidence suggests that protection against HSV would be best achieved by treatment with a subunit vaccine, which would not be oncogenic.

The primary requirement for creating any subunit vaccine is identification of the component(s) of the infectious agent that elicits antibodies that react against the intact form of the infectious agent. The HSV-1 envelope glycoprotein D (gD) is such a component (Fig. 7.4), because after

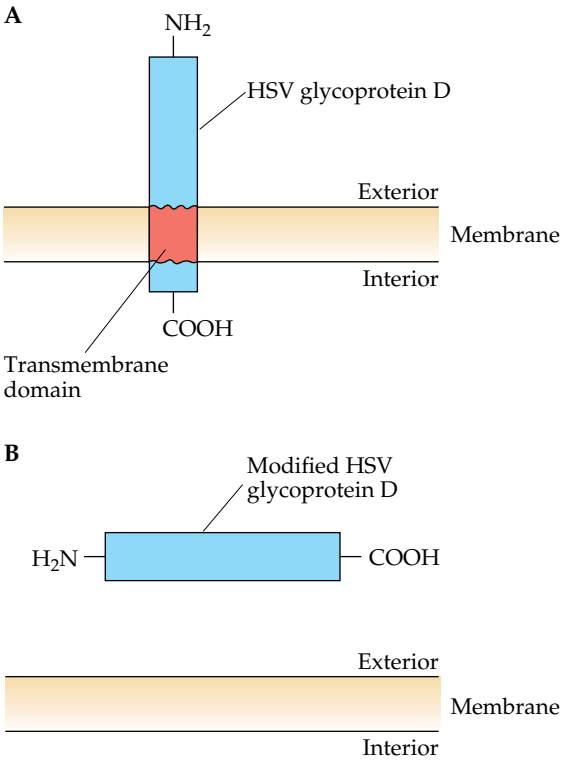


Figure 7.4 (A) Location in the envelope of HSV-1 gD with the transmembrane domain. **(B)** Extracellular location of a soluble gD without the transmembrane domain.

injection into mice, it elicits antibodies that neutralize intact HSV. The HSV-1 gD gene was isolated and then cloned into a mammalian expression vector and expressed in Chinese hamster ovary (CHO) cells (Fig. 7.5), which, unlike the *E. coli* system, properly glycosylate foreign eukaryotic proteins. The complete sequence of the gD gene encodes a protein that becomes bound to the mammalian host cell membrane (Fig. 7.4A). However, a membrane-bound protein is technically much more difficult both to produce in large amounts and to purify than a soluble one. Consequently, the gD gene was modified by removing the nucleotides encoding the C-terminal transmembrane-binding domain (Fig. 7.4B). The modified gene was then transformed into CHO cells, where the product was glycosylated and secreted into the external medium (Fig. 7.5). In laboratory trials with mice, the modified form of gD was effective against both HSV-1 and HSV-2. Unfortunately, this approach with mice in laboratory trials was not effective in human clinical trials.

Bovine Herpes Virus-1

Bovine herpes virus type 1 is an enveloped virus that is the cause of several diseases of cattle, including rhinotracheitis, vaginitis, balanoposthitis, conjunctivitis, and enteritis, and is a contributing factor in shipping fever. This is an economically significant disease for the global cattle

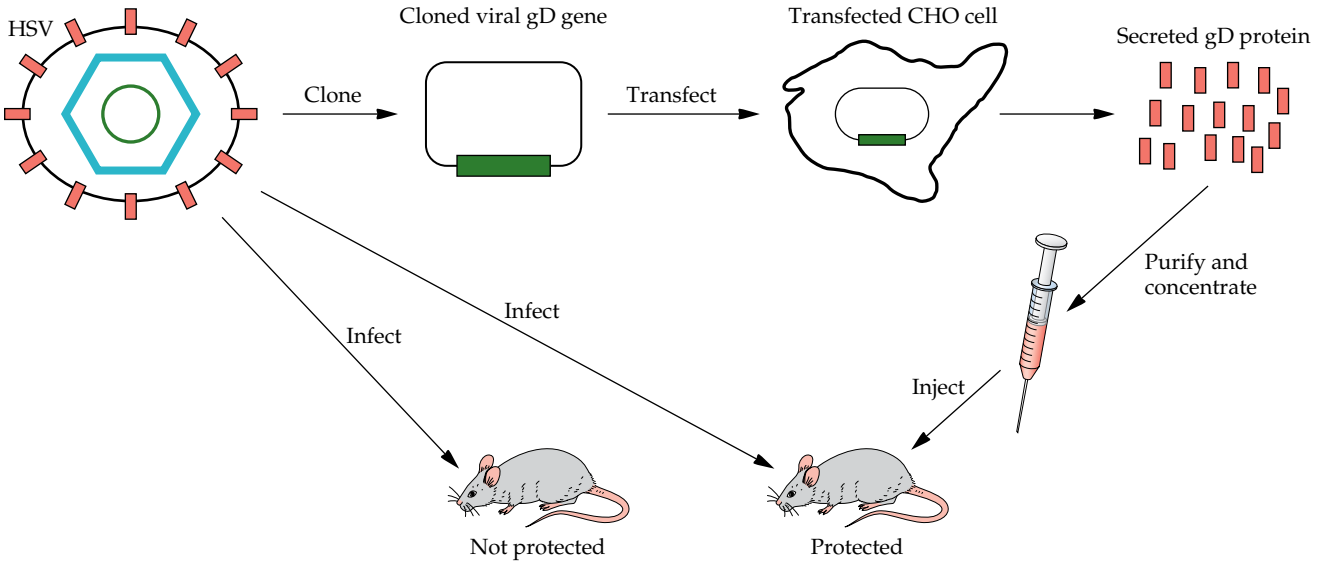
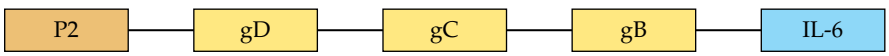


Figure 7.5 Schematic representation of the development of a subunit vaccine against HSV. The isolated HSV gD protein gene is used to transfect CHO cells. Then, the transfected cells are grown in culture and produce gD protein. Mice inoculated with the purified gD protein are protected against infection by HSV.

industry. Traditional vaccines against this virus, both inactivated and live attenuated, are available. Unfortunately, these vaccines often cause pregnant cattle to abort their potential offspring. Thus, a safe and effective subunit vaccine would be highly beneficial to the cattle industry. Knowing that the major envelope glycoproteins gD, gC, and gB induce a protective immune response against the virus, one group of researchers designed a unique protein that contained elements of all three of these proteins. Selected peptides (10 to 25 amino acids) from each of these three glycoproteins that were each able to fold into an epitope that is recognized by bovine antibodies were used to construct a unique subunit vaccine. Each of the peptides, in triplicate, was separated by short linker peptides (6 amino acids), and the entire construct was fused with bovine interleukin-6 (Fig. 7.6). When rabbits were vaccinated with this protein, they were significantly protected against a challenge with bovine herpes virus type 1. The data were also consistent with the safety and high level of immunogenicity of this multiepitope protein, making it a potential candidate for use as a vaccine to control bovine herpes virus type 1 infections.

Figure 7.6 Schematic representation of the DNA construct encoding a unique bovine herpes virus-1 subunit vaccine. P2 represents the universal $CD4^+$ T cell epitope of tetanus toxin P2, which enhances the immunogenicity of the multi-epitope protein. The abbreviations gD, gC, and gB represent bovine herpes virus-1 glycoprotein D, glycoprotein C, and glycoprotein B, respectively. IL-6 stands for bovine interleukin-6.



Cholera

Cholera is a major global public health problem and is an important threat in developing countries, particularly in areas of high population density and poor sanitation. Cholera, which manifests as mild to severe diarrhea and concomitant loss of fluids and electrolytes, can lead to death within 24 h if left untreated. Without treatment, severe infection by the bacterium *Vibrio cholerae*, the causative agent of cholera, has a mortality rate of 30 to 50%. In 2017, the World Health Organization (WHO) recorded 1.3 million cholera cases and ~6,000 deaths worldwide. However, the estimated actual burden of cholera is estimated to be approximately 2 to 5 million cases and 100,000 to 140,000 deaths per year. The disease is endemic to parts of Africa, Asia, the Middle East, and South America.

The bacterium *V. cholerae* colonizes the small intestine and secretes large amounts of the pathogenic hexameric enterotoxin. This protein complex consists of one A subunit, which has ADP ribosylation activity and stimulates host adenylate cyclase activity (leading to ion and water efflux into the intestinal lumen), and five identical B subunits that bind specifically to an intestinal mucosal cell receptor (Fig. 7.7). The A subunit has two functional domains: the A_1 peptide, which contains the toxic activity, and the A_2 peptide, which (noncovalently) joins the A subunit to the B subunits. Until a few years ago, a traditional cholera vaccine consisting of phenol-killed *V. cholerae* was in common use. This vaccine generated only moderate protection, typically lasting from about 3 to 6 months. In 2004, a vaccine (Dukoral) came into use that contained a mixture of a traditional vaccine consisting of heat-inactivated *V. cholerae* Inaba classic strain, heat-inactivated Ogawa classic strain, formalin-inactivated Inaba El Tor strain and formalin-inactivated Ogawa classic strain, and a recombinant cholera toxin B subunit (which greatly increases the efficacy of the vaccine). This vaccine is taken orally (two doses 1 week apart), and an additional booster immunization is not required for about 2 to 3 years. Unfortunately, Dukoral is considered to be primarily a travelers' vaccine that is used by people from industrialized countries who are traveling to cholera-endemic areas. This is because Dukoral (i) costs $6 a dose and (ii) requires coadministration with a buffer to protect the recombinant cholera toxin B subunit from the acidic environment of the stomach. However, since 2018, a single-dose traditional oral killed whole-cell cholera vaccine has become available for use in the developing world. Importantly, the cost of this vaccine, Euvichol-Plus, is $1.20 per dose.

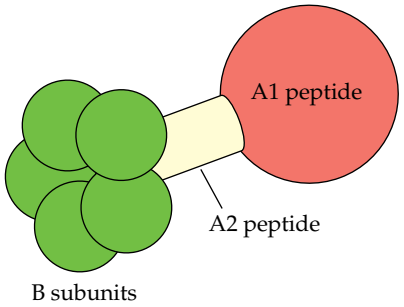


Figure 7.7 Schematic representation of the cholera toxin protein. The cholera toxin B and A2 peptides are noncovalently attached to one another.

Influenza

At present, there are 25 to 50 million documented global cases of influenza that are estimated to result in around 300,000 to 600,000 deaths each year. In 1918 to 1919, it is estimated by the CDC that ~500 million people became infected with influenza virus, resulting in around 50 million deaths worldwide. Fortunately, traditional inactivated or live attenuated influenza vaccines were developed in the 1930s, with large-scale availability in the United States beginning in 1945. The influenza vaccine is generally effective against three or four types of influenza virus. However, the virus evolves rapidly, and a current year's vaccine is often based on the previous year's virus so that the vaccine's effectiveness in the United States from 2010 to 2020 (in preventing the disease) is estimated to average ~35 to 40%. Various strains of influenza A have been responsible for six separate pandemics during the past 150 years.

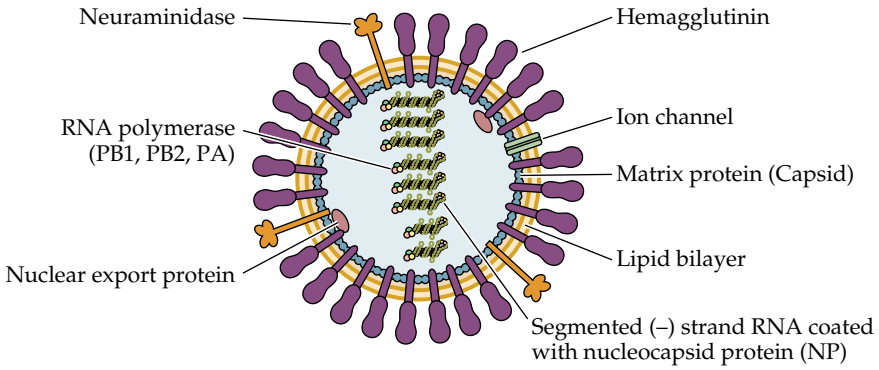
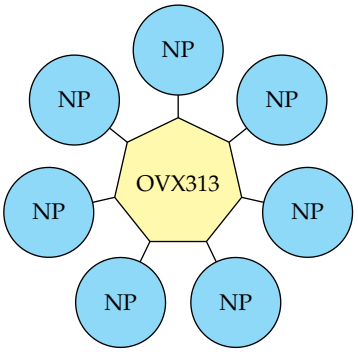


Figure 7.8 Schematic representation of an influenza virus. The surface of this enveloped virus is covered with the proteins hemagglutinin and neuraminidase. The genome of this virus consists of eight negative-stranded RNAs.

Influenza viruses belong to the *Orthomyxoviridae* family, are enveloped, and contain multiple copies of the glycoproteins hemagglutinin and neuraminidase on their surface, and their genomes consist of six to eight linear negative-sense RNA strands (Fig. 7.8). The protein hemagglutinin is a lectin (a carbohydrate binding protein) that is responsible for the binding of the virus to the surface of target human and animal cells, while the enzyme neuraminidase is involved in the release of progeny virus from the infected cells. These two surface proteins are targets for both antiviral drugs and antibodies directed against the whole virus. However, both evolve rapidly so that antibodies elicited by subunit vaccines targeting them are not always effective. Given the variability of hemagglutinin and neuraminidase, researchers have sought to develop subunit influenza vaccines based on other influenza proteins. In this regard, studies have found that some vaccines directed against the influenza nucleoprotein, a conserved internal influenza protein, protected mice against lethal doses of several different influenza strains. In one study, the cDNA for the influenza nucleoprotein was fused to a DNA fragment encoding a small oligomerization domain from the chicken complement inhibitor C4 binding protein, also called OVX313 (Fig. 7.9). This fusion protein was created because of earlier observations that oligomeric forms of many protein antigens are more immunogenic than the monomeric forms of those antigens. The oligomeric fusion protein significantly increased the immunogenicity of the nucleoprotein in mice as well as the mouse interferon response when mice were challenged with influenza virus. Most importantly, the oligomeric fusion protein significantly protected mice against a lethal challenge of several different influenza virus strains. Based on the results obtained with mice and knowing that the nucleoprotein is highly conserved in influenza A viruses, the fusion protein is a good candidate for a universal human influenza vaccine. It has been suggested that for added effectiveness, the fusion protein might be coadministered with antigens that elicit an antibody response to hemagglutinin or neuraminidase.

Figure 7.9 Schematic representation of a fusion protein of a small oligomerization domain (from the chicken complement inhibitor C4 binding protein, OVX313) and seven nucleoproteins. The fusion protein is significantly more immunogenic in mice than the nucleoprotein by itself.



SARS

The first new infectious disease of the 21st century was identified as severe acute respiratory syndrome, or SARS, and the first case of SARS was reported in the Guangdong Province of China in November 2002. In 2003,

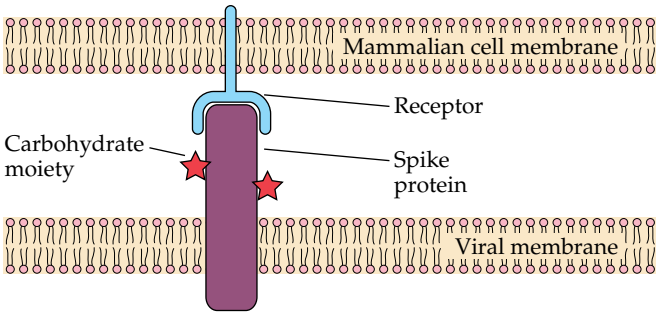
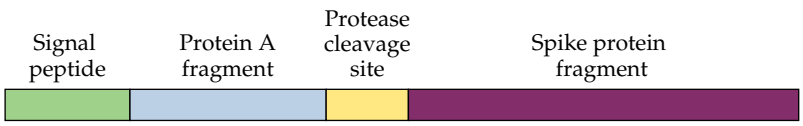


Figure 7.10 Simplified representation of the binding of the SARS virus glycosylated spike protein S1 subunit to a mammalian cell outer surface protein receptor, angiotensin-converting enzyme 2.

there were more or less simultaneous outbreaks in several major cities, including Hong Kong, Singapore, and Toronto. The disease rapidly spread to 29 countries on 5 continents. However, with the implementation of strict infection control procedures, by mid-July 2003, the outbreak was effectively contained. However, this was not before a total of 8,096 SARS cases and 774 associated deaths were reported.

The SARS coronavirus virus (SARS-CoV) is an enveloped virus that contains a single-stranded plus-sense RNA genome of ~30 kb. The viral spike protein, which is inserted into the viral membrane, binds to a receptor protein that is present on the surfaces of mammalian host cells (Fig. 7.10). The spikes of SARS-CoV consist of trimers of an S protein which is composed of two subunits, S1, containing the receptor binding domain, and the S2 subunits, which mediate fusion between the viral and host cell membranes. After the virus enters the body, it binds via the S1 subunit to the receptor, the angiotensin-converting enzyme 2 (ACE2), on primary target cells, including pneumocytes (surface epithelial cells of the alveoli) in the respiratory system. Following the binding of the virus, the viral and cell membranes fuse, facilitating the entry of the virus into the cell. In particular, it was found that amino acids 318 to 510 of the S1 subunit could bind efficiently to the host cell receptor protein. Following the determination of the complete nucleotide sequence of the SARS virus genome in 2003, it was relatively straightforward to express a codon-optimized version of this 192-amino-acid protein fragment in CHO cells. In addition to encoding the 192-amino-acid S1 peptide, the DNA construct introduced into CHO cells included a mammalian secretion signal, an N-terminal (*Staphylococcus aureus*) protein A purification tag, and a tobacco etch virus protease cleavage site (Fig. 7.11). The recombinant protein synthesized in CHO cells was secreted into the growth medium, purified by affinity

Figure 7.11 The main features of a portion of the recombinant plasmid construct used to transfect CHO cells and to produce the fragment of the SARS spike protein that interacts with the host cell receptor.



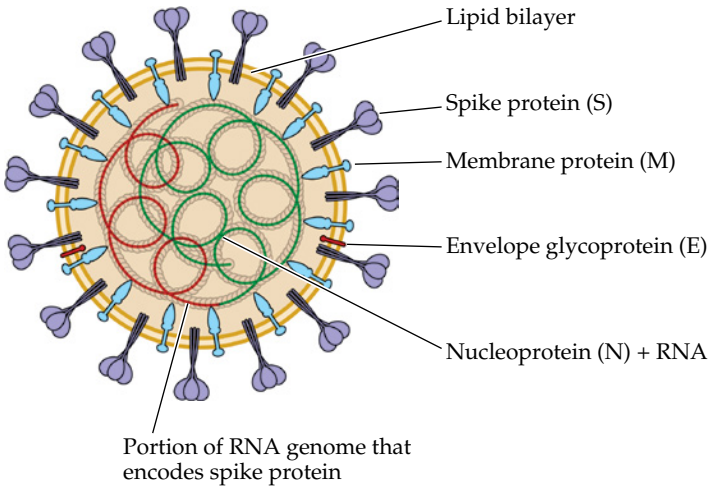
chromatography on a column containing immobilized IgG (which binds strongly to protein A), and then digested with tobacco etch virus protease to remove the protein A purification tag. The fully glycosylated form of this subunit vaccine candidate elicits a strong immune response in mice and rabbits.

COVID-19

Coronavirus disease 2019 (COVID-19) is an acute respiratory disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (Fig. 7.12). Similar to the SARS virus (discussed above), the SARS-CoV-2 virus binds to angiotensin-converting enzyme 2 (ACE2) receptors on the target cells and is then internalized by endocytosis delivering the viral genome for replication in the host cell. The first confirmed case of this disease was on 17 November 2019 in Hubei, China. In the following year, more than 82 million cases were reported across 218 countries and territories worldwide, resulting in more than 1.8 million deaths. Symptoms of this viral infection include fever, cough, fatigue, shortness of breath, and loss of smell and taste (Fig. 7.13). In addition, individuals already in ill health or over the age of 80 may acquire acute respiratory disease syndrome, multiorgan failure, septic shock, and blood clots, often resulting in death. By the middle of May 2021 two companies reported successful clinical trials of a vaccine to fight this disease, with scientists from several hundred laboratories around the world having undertaken a massive effort to develop both effective antiviral compounds and vaccines. It is expected that within the next year there will be several additional highly effective vaccines that will protect people from this virus.

Some of the vaccines being developed by different groups of scientists around the world include the following. (i) Traditional vaccines employing a killed or attenuated virus. (ii) Subunit vaccines in which either the viral spike protein or a portion of the spike protein (or another viral protein) is overproduced and used to inoculate individuals (an approach that

Figure 7.12 Schematic representation of the SARS-CoV-2 virus. The portion of the viral RNA encoding the spike protein is highlighted.



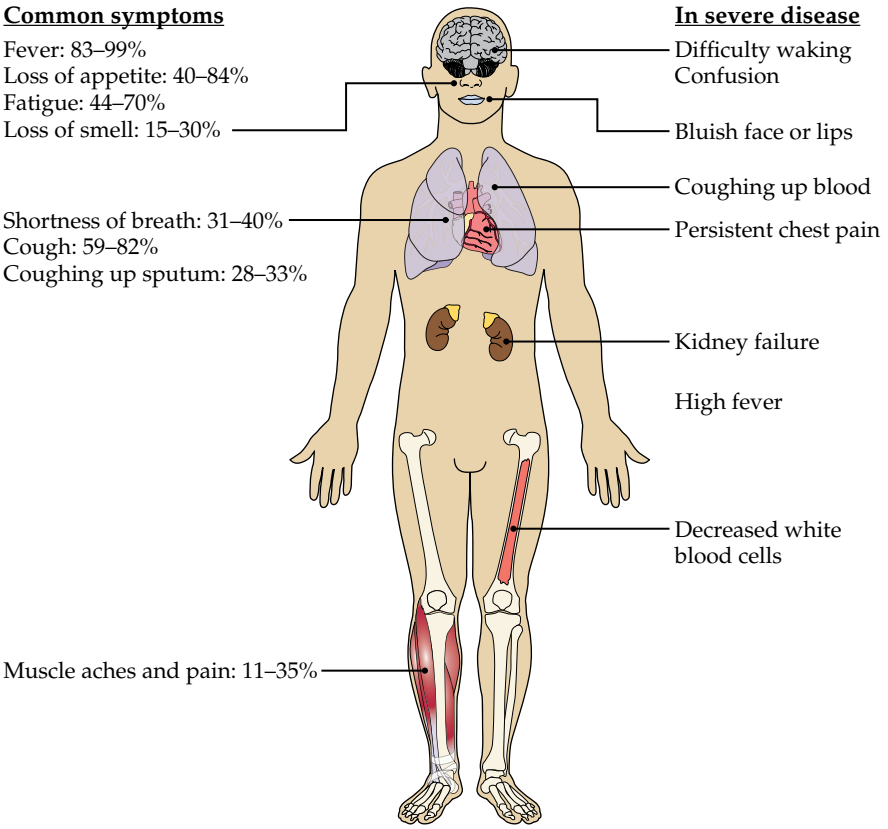


Figure 7.13 A schematic representation of how the human body may be affected by SARS-CoV-2.

might require the use of adjuvants). This approach has been successful in mice, rats, and nonhuman primates. (iii) Synthetic lipid particles into which spike proteins that have been overproduced by baculoviruses and insect cell cultures are inserted, with the spike proteins facing outside the particle. (iv) Nucleic acid vaccines in which DNA or RNA encoding a portion of the virus is enclosed in lipid nanoparticles (see chapter 6) and then introduced into human cells that in turn produce copies of the encoded viral protein. Transient side effects lasting 12 to 24 hours to the lipid nanoparticle in 2 to 5% of inoculated individuals include fatigue, muscle pain, joint pain, and headache. (v) Nonpathogenic virus-based vectors (such as adeno-associated virus, discussed in chapter 6, or the yellow fever vaccine virus) which carry a nucleic acid encoding one or more SARS-CoV-2 virus proteins with the expectation that this vector virus will express the encoded proteins and provoke a strong immune response against them. (vi) Virus-like particles (as described for human papillomavirus later in this chapter) which mimic the structure of the coronavirus but lack its genetic material. The vaccines that were licensed in 2020 to protect people against the effects of this virus both contain mRNA of the viral spike protein delivered in lipid nanoparticles. Both of these vaccines require two inoculations 3 to 4 weeks apart. In addition to the many types of vaccines mentioned above, several groups of researchers have developed monoclonal antibodies directed against

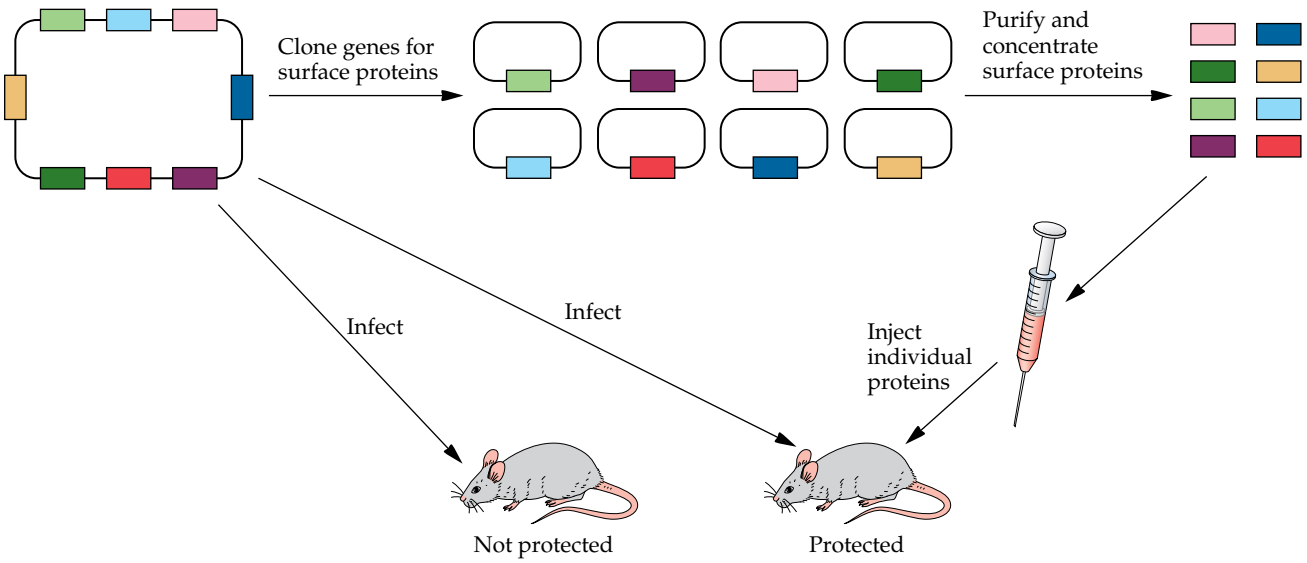
SARS-CoV-2 virus proteins, with the expectation that those individuals who show adverse reactions to the developed vaccines might be treated with these monoclonal therapeutic agents.

Staphylococcus aureus

The Gram-positive bacterium *S. aureus* colonizes the skin or mucous membranes of about 30% of the human population and can cause serious diseases such as pneumonia, toxic shock syndrome, and sepsis. It is also a major cause of hospital-acquired infection. Methicillin-resistant *S. aureus* (MRSA) strains have emerged and spread rapidly in the community and are now the most common cause of community-associated skin and soft tissue infections. Some strains are resistant to vancomycin, considered an antibiotic of last resort. In view of the lack of treatment options for serious *S. aureus* infections, the discovery of a safe and effective *S. aureus* vaccine has the potential for great benefit in public health care.

Traditional whole-cell attenuated or killed *S. aureus* vaccines have been developed; however, these vaccines have not been particularly effective. Similarly, subunit vaccines composed of individual bacterial surface proteins generate immune responses that afford only partial protection when tested in experimental animals. A more effective subunit vaccine has recently been developed to protect individuals against *S. aureus* by combining several of the bacterium's antigens (Fig. 7.14). Starting with one disease-causing strain of *S. aureus*, 23 bacterial outer

Figure 7.14 Development of a vaccine against *S. aureus*. The genes encoding several bacterial outer surface proteins from a disease-causing strain of *S. aureus* were cloned into plasmid vectors. The proteins were expressed and purified, and mice were separately immunized with each of the purified proteins. The mice were then injected with a live disease-causing strain of *S. aureus*. Many of the recombinant surface proteins generated an immune response that afforded partial protection against staphylococcal disease, with some proteins affording more protection than others. Mice that were not injected with any of the surface proteins but were infected with *S. aureus* were not protected from disease.



surface proteins (i.e., integral membrane proteins) were identified from genomic DNA sequence data. Then, the coding regions of these proteins, minus the signal sequences (to prevent the protein's insertion into a cell membrane), were amplified by PCR and cloned (separately) into plasmid vectors that enabled the proteins to be expressed in *E. coli* with a poly-histidine tag at the N terminus of each protein (to facilitate the purification of the overexpressed soluble proteins). The proteins were expressed and purified, and mice were separately immunized with each of the 23 purified proteins. The immunized mice were subsequently challenged by injections of live disease-causing *S. aureus*. Many of the recombinant surface proteins generated an immune response that afforded partial protection against staphylococcal disease, with some proteins affording more protection than others. Unfortunately, over the long term, immunization with individual surface proteins afforded only modest protection. However, a mixture of the four proteins that individually generated the most effective antibodies was used to immunize mice and was found to completely protect the immunized mice against the pathogen. The experimental design ensured that only common, and not strain-specific, *S. aureus* surface proteins were used to immunize mice. Thus, it is not surprising that the tetravalent subunit vaccine that was developed was effective against five different clinical isolates (strains) of *S. aureus*. This work represents an important first step in the development of an *S. aureus* vaccine.

Another group of researchers selected five conserved *S. aureus* antigens, which were known to have different roles in *S. aureus* pathogenesis, as the starting point for the development of a novel subunit vaccine. The five conserved antigens were the secreted factors α-hemolysin, Ess extracellular A and Ess extracellular B, and the two surface proteins ferric hydroxamate uptake D2 and conserved staphylococcal antigen 1A. The five combined vaccine antigens provided a high level of protection (close to 100%) in four different mouse models when challenged with several different disease-causing *S. aureus* strains. These data demonstrate that the rational selection of mixtures of conserved antigens can lead to promising vaccine formulations against *S. aureus*. However, these results remain to be confirmed in humans.

Human Papillomavirus

Human papillomavirus (HPV) is a nonenveloped virus that is sexually transmitted. While most infections are benign and often asymptomatic, persistent infection with some strains of human papillomavirus is associated with the development of cervical and related cancers, as well as genital warts. Over 100 different types of HPV exist, and 15 types have carcinogenic potential. Most genital warts are caused by HPV-6 and HPV-11, while HPV-16 and HPV-18 account for about 70% of invasive cervical cancers. Thus, a vaccine that prevents HPV-16 and HPV-18 infection could significantly reduce the incidence of cervical cancer. Moreover, a vaccine that is directed against several different types of human papillomaviruses could effectively prevent nearly all HPV-induced cervical cancers. To put this into perspective, cervical cancer is the second most commonly diagnosed cancer among women worldwide, accounting for more than 250,000 deaths per year.

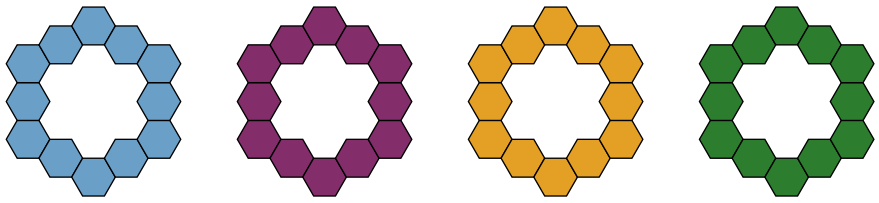
It was previously shown that the viral major capsid protein L1 can self-assemble into virus-like particles (see Fig. 7.2) that resemble papillomavirus virions, and these particles are highly immunogenic, inducing neutralizing antibodies directed against the whole live virus. The gene for the L1 protein from each of four virus types was cloned and expressed in a recombinant *Saccharomyces cerevisiae* (yeast) strain. Following separate fermentations of each of the four recombinant yeast strains, the viral capsid proteins assembled into virus-like particles (i.e., the viral capsid without any other viral proteins or the viral nucleic acid). The demonstration that the L1 protein self-assembled into virus-like particles was key to the development of an HPV vaccine. These virus-like particles were then purified and combined to form a quadrivalent vaccine.

In June 2006, the U.S. Food and Drug Administration (FDA) approved a vaccine that protects women against infection by the papillomavirus types most frequently associated with cervical cancer and genital warts, i.e., HPV-6, -11, -16, and -18. By the end of 2006, this vaccine had been approved for use in more than 50 countries worldwide. The vaccine, called Gardasil, is quadrivalent; i.e., it contains virus-like particles assembled from the L1 proteins of HPV-6, -11, -16, and -18 (Fig. 7.15 and Box 7.1). Prior to its approval, in four phase II and III clinical trials, prevention of HPV-16- and HPV-18-related cervical cancers was 99 to 100%. In the combined data from the phase II and III studies, efficacy against external genital lesions from all 4 HPV types was 95 to 99%.

In addition to Gardasil, a second HPV vaccine, Cervarix, was developed and then approved by the U.S. FDA in 2009. This vaccine is a combination of HPV-16, -18, and -20 L1 proteins that are produced in insect cells in culture. Like Gardasil, this vaccine requires three injections.

It may take many years before the long-term effectiveness of these vaccines in reducing cervical cancer is known. However, recent data for about 86,000 individuals vaccinated with the quadrivalent HPV vaccine beginning in 2007 from six sexual health clinics in Australia indicate that vaccine coverage rates were excellent and averaged about 80% for all three doses of the quadrivalent vaccine. In that case, the proportion of women under 21 years of age who had genital warts was reduced from 11.5% in 2007 to 0.85% in 2011. Only 13 cases of genital warts were observed in women younger than 21 years in these six health clinics in 2011. The proportion of women aged 21 to 30 years who presented with genital warts was reduced from 11.3% in 2007 to 3.1% in 2011. Unfortunately, the

Figure 7.15 Schematic representation of the virus-like particles assembled from cloned and overproduced L1 proteins from the capsids of four different strains of human papillomavirus. These virus-like particles are the constituents of a commercial subunit vaccine against the virus.



box 7.1

A Vaccine To Prevent Cervical Cancer

On 15 September 2007, the headline on the front page of *The Globe and Mail*, a Toronto, Canada, newspaper, read, "Should your daughter get the needle?" The article that followed related how the Canadian federal government, in conjunction with the Ontario provincial government, was funding a program that would offer free vaccinations against human papillomavirus (the most prevalent sexually transmitted disease worldwide such that three out of four adults will be infected with human papillomavirus during the course of their lifetime) to girls in grade 8 (typically 12- and 13-year-olds). The vaccine, which had received approval in the United States a year earlier, provides immunized women with immunity against the viruses that are responsible for approximately 70% of all cervical cancers and 90% of genital warts. Grade 8 was chosen because, according to officials, it is before most girls become sexually active. The vaccine, sold under the brand name Gardasil, is given by needle in three doses over 6 months and is approved for females between the ages of 9 and 26. The rationale for giving the vaccine at such a young age is related to the fact that once a woman has been exposed to the four strains of the virus for which the vaccine provides protection, the vaccine will no longer be effective. In addition, the immune response is most robust when the vaccine is given at younger ages. The three doses of vaccine cost about $300 to $400, although those inoculated through this program receive the vaccine free of charge.

Boys can also get human papillomavirus infections, and in 2015 the U.S. Centers for Disease Control and Prevention recommended that all boys and girls aged 11 or 12 should routinely receive the vaccine.

According to *The Globe and Mail*, "For many parents it's a no-brainer: anything that will protect their daughters from cancer . . . is worth the risks." However, at the same time, a small but vocal minority has expressed serious reservations about this program. On one hand, there are individuals who do not trust the medical establishment, the pharmaceutical companies, and/or the government. Others have expressed concerns that some girls will naively believe that this vaccine will protect them against any and all sexually transmitted diseases and use this as a rationale or excuse for becoming sexually active at an early age. Still others have questioned the potential side effects from the vaccine, despite the fact that extensive clinical trials have shown that they are quite rare. Notwithstanding the concerns of some individuals, the vaccine was initially offered through school inoculation programs to young females in the Canadian provinces of Newfoundland and Labrador, Prince Edward Island, Nova Scotia, and Ontario. However, by September 2008, all of the other provinces in Canada had decided to implement the program of immunizing girls in grade 8 (this voluntary, school-based vaccination program is administered by qualified registered nurses from local Public Health Units). The real benefits of the program (hopefully an enormous reduction in cervical cancer) may not be known for several decades; in the meantime, the debate will continue. Also, since the vaccine does not protect against all strains of human papillomavirus, it is essential that women continue to get an annual Pap test.

rate of diagnoses of genital warts in women over 30 was unaffected. The proportion of men under 21 years old with genital warts also decreased appreciably, from 12.1% in 2007 to 2.2% in 2011.

Foot-and-Mouth Virus

Foot-and-mouth disease (sometimes called hoof-and-mouth disease) is a highly contagious viral disease of cloven-hoofed animals, including cattle, sheep, goats, water buffalo, antelope, deer, bison, and pigs. Foot-and-mouth disease virus (FMDV) has a devastating impact on cattle and swine and is extremely virulent, but for the most part, it has been possible to limit infections with the virus by using formalin-killed FMDV preparations as a vaccine (i.e., a traditional killed-virus vaccine). Approximately 1 billion doses of this killed-virus vaccine are used worldwide each year. The availability of the vaccine notwithstanding, in 2001, there was a major outbreak of foot-and-mouth disease in Europe in which approximately six million

cattle were slaughtered, and their carcasses were incinerated in an effort to prevent the virus from spreading. A similar outbreak of foot-and-mouth disease in 2011 in South Korea resulted in approximately three million pigs being slaughtered. Foot-and-mouth disease is considered to be endemic in several parts of Asia and in most of Africa and the Middle East.

The virus responsible for this disease is a picornavirus (the same family as poliovirus), which has a single-stranded positive RNA genome that is approximately 7,700 nucleotides long. FMDV is a nonenveloped virus with a viral capsid that contains 60 copies each of viral proteins (VP) 1 to 4 (Fig. 7.16). The viral proteins are synthesized as a single 95-kDa polyprotein that is cleaved to yield VP1 to -4, which are assembled into a capsid. The major antigenic determinant that induces neutralizing antibodies is VP1, and although purified VP1 is a much less potent antigen than intact viral particles, it can still act as a subunit vaccine and elicit neutralizing antibodies by itself and therefore can protect animals from infection by FMDV.

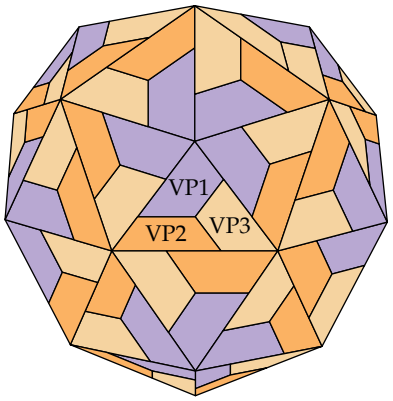
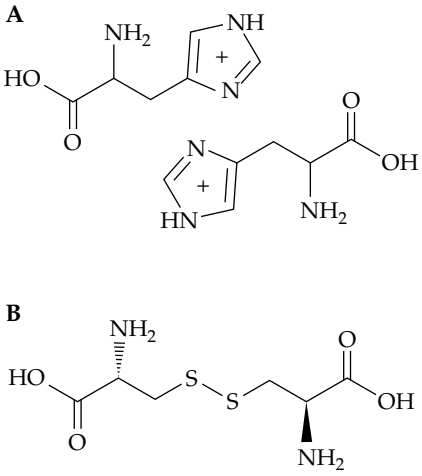


Figure 7.16 Schematic representation of the capsid surface of a picornavirus. The relative positions of VP1, VP2, and VP3 are shown. VP4 acts as an anchor below the capsid surface holding these proteins together.

FMDV is unstable at a pH of <7.0 and elevated temperatures, with the capsid dissociating into 12 pentamers and the genome being released. This pH and temperature instability results in killed FMDV (the vaccine preparation) being unstable and having a limited shelf life, which is a significant problem in parts of the world where it is not always possible to store the vaccine at cold temperatures. In an effort to overcome the instability of FMDV particles, researchers first determined the detailed three-dimensional structure of the FMDV capsid. Careful examination of this structure suggested that viral particle instability might result from an exposed histidine residue on VP2 interacting with (and repelling) the same histidine residue on an adjacent VP2 subunit (Fig. 7.17A). The two histidine residues repelled one another, especially at low pH values as a consequence of the intrinsic pK_a of the histidine ring (i.e., 6.0), which carries a positive charge at low pH. To overcome this instability, researchers changed the problematic histidine residue (by directed mutagenesis) to a cysteine residue so that the two proximal cysteine residues might form a (covalent and therefore strong) disulfide bond (Fig. 7.17B). In fact, the modified VP2 readily assembled into FMDV particles, and when used to inoculate cattle, these modified particles yielded a comparable level of immunity to the killed whole-cell vaccine. Importantly, however, the modified viral particles (in essence, empty capsids) provided a safe and efficacious alternative to the existing vaccine, an alternative whose manufacture did not require biocontainment and one that was expected to have a greatly increased stability during storage and transport.

Figure 7.17 **(A)** Two proximal histidine residues at low pH showing a positive charge on each molecule so that these residues (when inserted into a protein molecule) would naturally repel one another. **(B)** Two proximal cysteine residues readily form a stable covalent disulfide bridge.



Streptococcus

Group B *Streptococcus* (*Streptococcus agalactiae*) is a Gram-positive bacterium that is a part of the normal flora of the gut and the genital tract. In pregnant women, it is sometimes harmful to both mother and (eventually) to the baby. In fact, maternal colonization by this organism may result in neonatal death due to it causing pneumonia, sepsis, and meningitis in newborns and young infants. Globally, group B *Streptococcus* infection is reported to cause ~150,000 preventable stillbirths and infant deaths annually. There is not a vaccine against group B *Streptococcus* at the present time. Currently, the most promising vaccine candidates are the structural

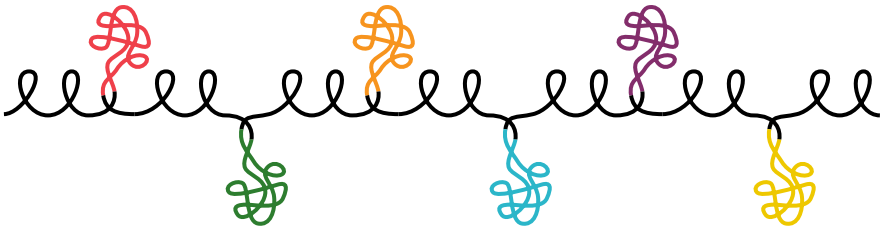


Figure 7.18 Schematic representation of a chimeric protein consisting of the six separate protective domains (each domain is shown in a different color) of each of the six variants of the group B *Streptococcus* type 2a pilus backbone proteins.

subunits of the pili protruding from the bacterial surface; in this bacterium the pili play a key role in bacterial virulence and pathogenesis. The backbone protein (one of the major protein components) of the type 2a pilus of group B *Streptococcus* has six antigenically different but structurally similar variants. Antibodies specifically recognize one of the domains that make up the backbone protein in each of the variant forms. Moreover, when they were tested, each antibody against a particular variant form of the pilus backbone protein protected mice against one of the different strains of group B *Streptococcus*. With this in mind, researchers constructed a chimeric protein that consisted of the antigenic domains of each of the six variants (Fig. 7.18). They then showed that when this chimeric protein was administered as a vaccine, it protected mice against all six of the type 2a pilus-carrying strains. This chimeric protein is currently in early-stage clinical trials.

Group A *Streptococcus* (*Streptococcus pyogenes*) causes infections of the throat and skin, with an estimated >700 million cases of mild diseases such as pharyngitis (strep throat), impetigo, and scarlet fever per year and >600,000 cases of serious invasive infections (including pneumonia, toxic shock syndrome, and necrotizing fasciitis) per year. While infections caused by group A *Streptococcus* are within the top 10 global infectious disease causes of death on a yearly basis, there is no commercially available vaccine for these infections. One group of researchers recently took an innovative bioinformatics approach toward developing a universal vaccine for group A *Streptococcus* infections. They assembled and then sequenced the genomes of all of the 2,083 known global isolates of group A *Streptococcus* (Fig. 7.19) and identified a core genome of 1,306 genes that was common to all of these strains. By removing all of the repeated genes (i.e., the same gene present in more than one copy per genome) from consideration, they were left with a core genome of 416 unique genes. From the unique core genes, they identified all of the genes whose DNA sequences were highly conserved and also were present in more than 99% of the bacterial isolates. From the small number of remaining genes, the researchers modeled the structures of the proteins encoded by those genes and (based on the knowledge of what makes a protein antigenic) predicted which proteins might act as major immunological antigens. Having winnowed the genomes of 2,083 different bacterial strains, each with 1,800 to 1,900 different genes, down to just a handful of genes, it is now necessary to synthesize the proteins encoded by those genes and test them for the ability to act, either separately or together, as a universal group A *Streptococcus* subunit vaccine.

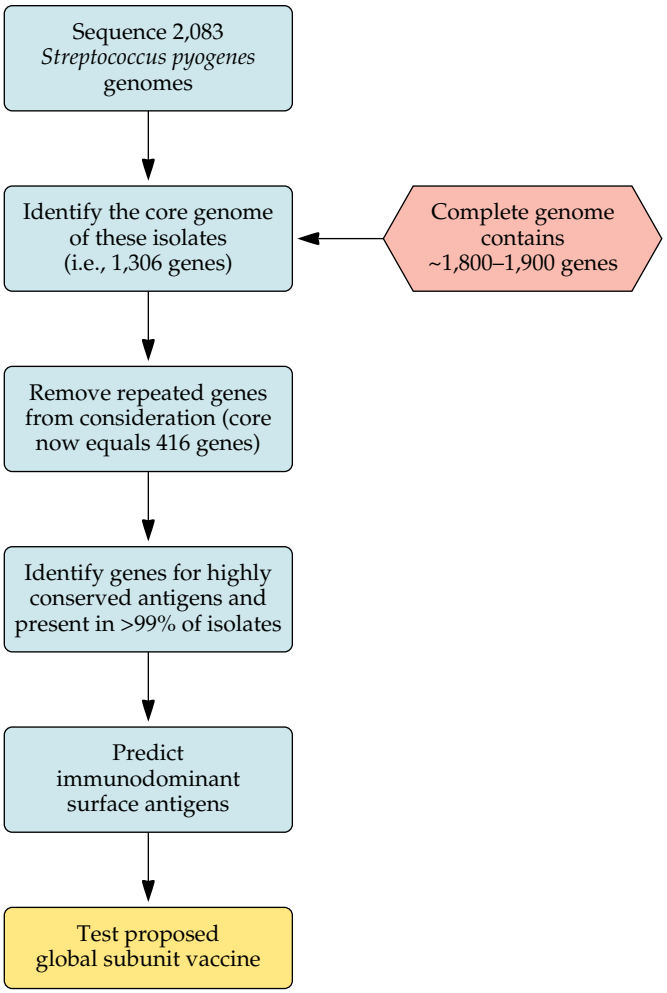
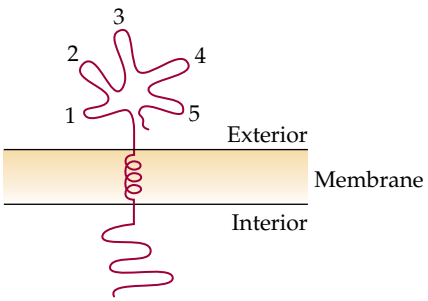


Figure 7.19 Flow chart summarizing some of the major steps in the identification of group A *Streptococcus* universal immunodominant surface antigens. The identified antigens must now be synthesized and tested as effective subunit vaccines.

Peptides

The question arises as to whether a small discrete portion (domain, peptide) of a protein can act as an effective subunit vaccine and induce the production of neutralizing antibodies. One might expect that only the portions of a protein that are accessible to antibody binding, that is, on the exterior surface of the virus, would be immunologically important and that those located in inaccessible regions inside the virus particle could be ignored if they do not contribute to the conformation of the immunogenic domain (Fig. 7.20). If this argument has validity, it is possible that short peptides that mimic epitopes (antigenic determinants) will be immunogenic and may be used as peptide vaccines. This approach requires that an epitope consists of a short stretch of contiguous amino acids, which does not always occur naturally, and the peptide must be able to assume the same conformation as the epitope in the intact pathogen. A single epitope may not be sufficiently immunogenic.

Figure 7.20 Generalized envelope-bound protein with external epitopes (1 to 5) that may elicit an immune response.



Malaria

Malaria is a parasitic disease transmitted to humans and other vertebrates by mosquitoes and is potentially fatal. Following a 7- to 14-day incubation period after infection, malaria is characterized by fever, chills, abdominal pain, vomiting, muscle and joint pain, fatigue, and malaise. Historically, the malaria parasite is thought to have killed more humans than any other single infectious agent, with the risk of death increasing significantly with age and among children. Today, along with HIV and tuberculosis, malaria remains one of the "big three" infectious diseases, every year exacting a heavy toll on human life and health in parts of Central and South America, Asia, and sub-Saharan Africa, where up to 90% of malaria deaths occur. According to the World Health Organization, in 2018, ~405,000 deaths, mostly among children under 5 years of age, resulted from ~228 million cases of malaria. Malaria parasites are bigger, more complicated, and more difficult to deal with than the viruses and bacteria that have been controlled with vaccines. Despite decades of research toward a vaccine for malaria, this goal has remained somewhat elusive.

The genus *Plasmodium*, a group of unicellular eukaryotes that are obligate parasites of multiple species of insects and vertebrates, consists of approximately 125 known species of parasitic protozoa, 5 of which (*Plasmodium falciparum*, *P. vivax*, *P. ovale*, *P. malariae*, and *P. knowlesi*) infect humans and cause malaria. Because *P. falciparum* elicits the most severe malaria disease and the most deaths (99.7% of the total), it is the target of most vaccine development efforts. The *Plasmodium* life cycle is very complex. Sporozoites from the saliva of a biting female mosquito are transmitted to either the blood or the lymphatic system and then migrate to the liver and invade liver cells (hepatocytes) (Fig. 7.21). In hepatocytes, the parasite grows, divides, and propagates as haploid merozoites before budding off in merosomes containing hundreds or thousands of merozoites. These merosomes lodge in pulmonary capillaries and slowly disintegrate there, generally over 2 or 3 days, releasing the merozoites. The merozoites invade the red blood cells (erythrocytes), where the parasite proliferates, producing new merozoites, which then leave the red blood cells and travel within the bloodstream to invade new red blood cells. The parasite eventually forms gametocytes, which may be ingested by feeding mosquitoes. Fusion of the gametes that develop from gametocytes leads to the formation of new sporozoites in the mosquito that can infect new individuals, spreading the disease. In the life cycle of the malaria parasite, it is the asexual blood-stage multiplication that is responsible for most of the acute symptoms of the disease.

In areas where malaria is endemic, some individuals show considerable resistance to the disease despite the fact that when their blood is examined, they are found to carry the parasite. This resistance to the worst symptoms of malaria was shown to be a result of an "antibody-dependent cellular-inhibition" mechanism that inhibits parasite development. In other words, some individuals who were infected with the malaria parasite made antibodies against a parasite protein that prevented the growth of the parasite. Following a detailed study, it was determined that the protective antibodies targeted merozoite surface protein 3. When this protein was examined in different strains of *Plasmodium*, it was observed that while the N-terminal part of the protein varied considerably from one strain to another, the C-terminal end of the protein was highly conserved among the various isolates of the parasite. It was therefore decided to chemically

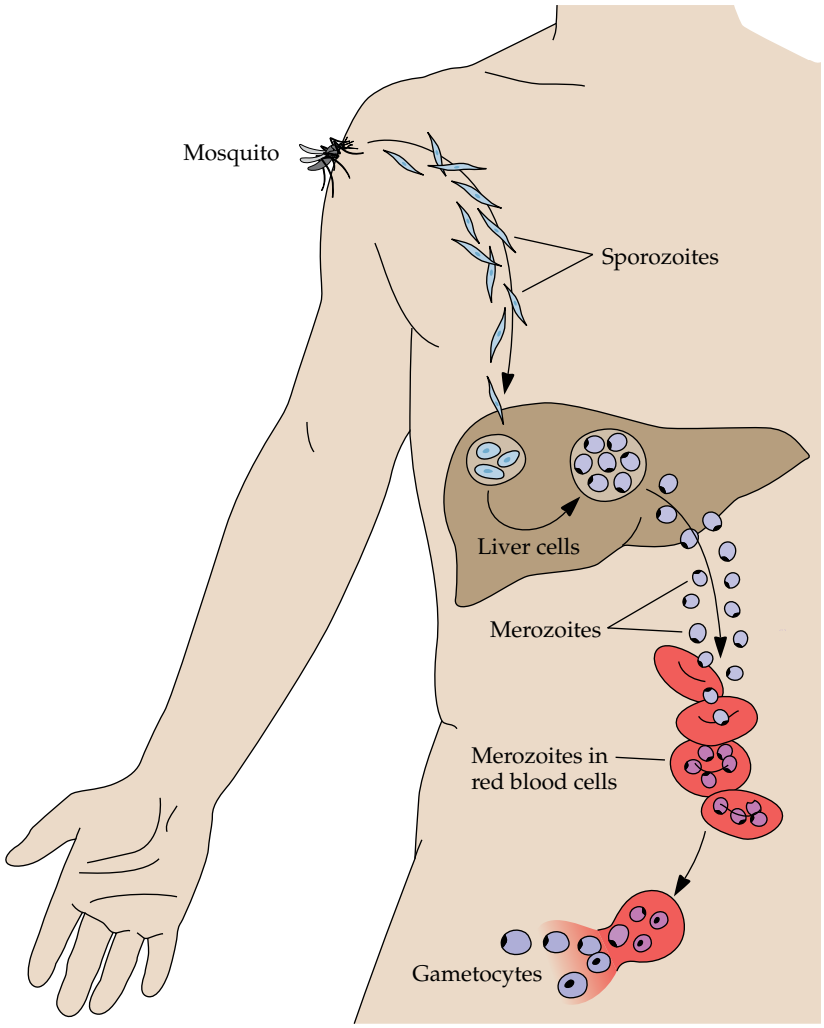


Figure 7.21 Infection of an individual with *Plasmodium falciparum* (a malaria-causing parasite) introduced by a mosquito.

synthesize peptides that corresponded to small portions of the C terminus of merozoite surface protein 3 (Fig. 7.22). The peptides were tested against serum from individuals who were resistant to the parasite, and the antibodies were affinity purified based upon their interaction with one or more of the peptides. The antibodies that bound to the peptides were then tested in an antibody-dependent cellular-inhibition assay. Antibodies directed against peptides B, C, and D (Fig. 7.22) had a major inhibitory effect on parasite growth. Based on the ability of peptides B, C, and D to bind to and select protective antibodies, a peptide representing amino acid residues 181 to 276 of merozoite surface protein 3 was chemically synthesized. This peptide is currently being tested in clinical trials as a novel malaria vaccine. In fact, during the past 10 years, there have been nine clinical trials of synthetic malaria peptide vaccines designed to target different stages of the malaria parasite life cycle. While more research needs to be done, in the future, synthetic peptide vaccines could become highly specific, relatively inexpensive, safe, and effective alternatives to traditional vaccines.

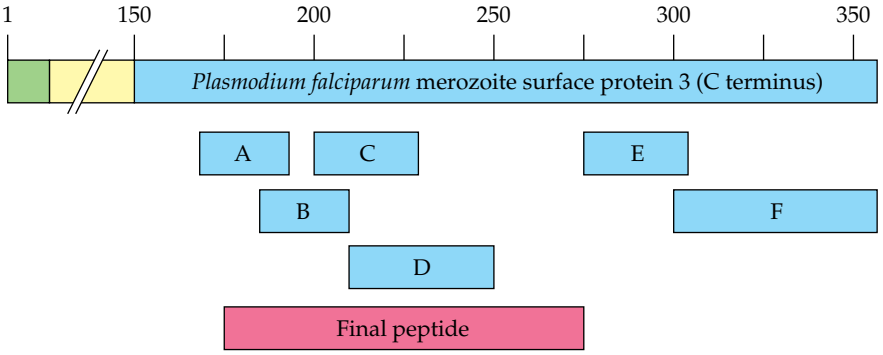


Figure 7.22 Schematic representation of *Plasmodium falciparum* merozoite surface protein 3 and peptides corresponding to portions of the C terminus. The peptides, labeled A to F, are drawn to scale, with the numbers above the whole protein indicating the amino acid number (counting from the N terminus). The "final peptide" is currently being tested in clinical trials for efficacy as a malaria vaccine.

Although a vaccine containing a merozoite antigen would protect against the erythrocyte stage of infection, a preerythrocyte-stage vaccine would block infection and prevent the parasites from reaching the blood and causing disease. In this regard, in November 2012, a malaria vaccine that targeted the sporozoite stage was tested in phase III clinical trials. The peptide vaccine, RTS,S/AS01, comprises the antigen RTS,S and a liposomal adjuvant (AS01) to boost the immune response. The antigen is a hybrid of peptides derived from the repeat (R) and T-cell epitope (T) regions of the immunodominant circumsporozoite protein, which covers the exterior of the sporozoite and the stable surface antigen from hepatitis B virus (S) (Fig. 7.23). The results of a small phase IIIa clinical trial of this peptide vaccine with children ages 5 to 17 months were promising, finding 55.8% protection against malaria. However, in a larger subsequent trial with African infants, the peptide provided only 31% protection. Notwithstanding the less than hoped for levels of protection from this vaccine, it became the first approved malaria vaccine (Mosquirix), and pilot implementation began in 2019 in several countries where mortality from malaria is high.

Two major difficulties in developing effective malaria vaccines are (i) the complexity of the parasite's life cycle, which suggests a large number of possible *P. falciparum* protein targets, and (ii) the diversity in protein composition among different *P. falciparum* strains. Until very recently, antimalarial vaccines were largely composed of a single protein antigen, and only a very few of the parasite proteins have been tested as vaccine candidates. However, the genome of *P. falciparum* encodes more than 5,000 different proteins, with a little more than 400 being involved in the blood-stage development of the parasite, and around 100 are surface exposed and thus antibody susceptible. To expand the repertoire of potential subunit vaccines against *P. falciparum*, one group of researchers systematically screened a group of 29 previously untested blood-stage surface-exposed proteins to determine whether any of them might provide the basis for a unique malaria vaccine. The enormity of the task permitted only about one-third of the protein antigens that are exposed on the *P. falciparum* surface during blood-stage development of the parasite to be tested. However, it is the researchers' intention to test additional protein antigens in the future.

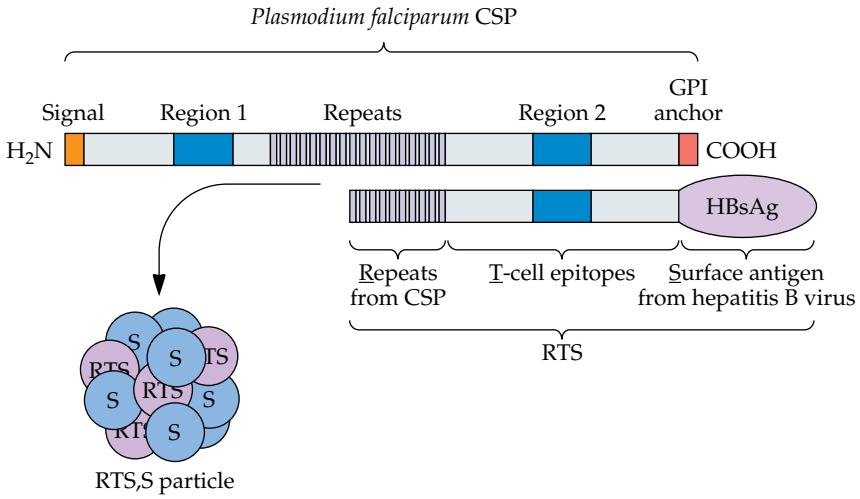


Figure 7.23 Schematic representation of the RTS,S hybrid vaccine particle used in the treatment of malaria. The central repeat (R) region is derived from the *Plasmodium falciparum* circumsporozoite protein (CSP) and is genetically fused to the T-cell epitope (T) and surface antigen (S) regions derived from the hepatitis B virus. The hybrid protein, RTS, expressed in transformed *Saccharomyces cerevisiae* yeast cells, self-assembles into virus-like particles similar to native hepatitis B surface antigen (HBsAg), with exposure of the circumsporozoite epitope on the exterior surface of these particles. The glycosylphosphatidylinositol (GPI) anchor attaches the CSP to the lipid bilayer of a cell membrane.

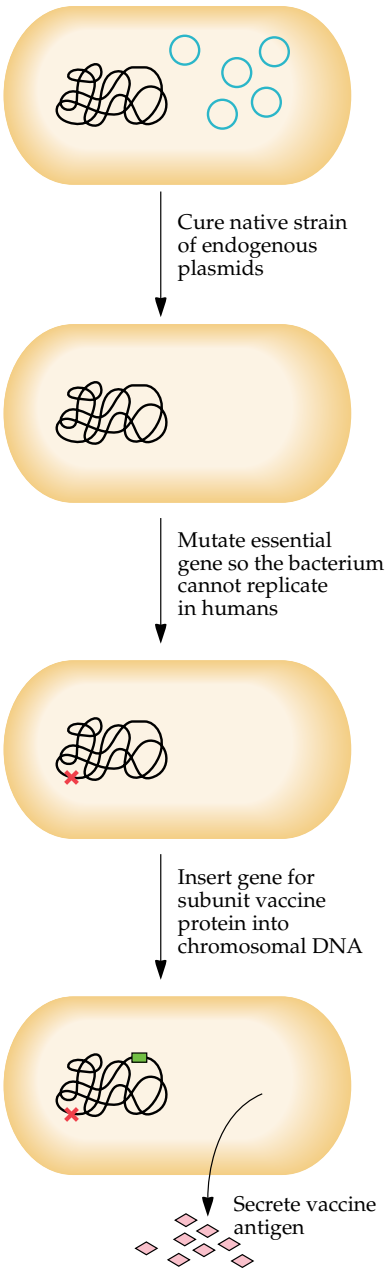
Following extensive testing, these workers identified several unique protein targets that were able to inhibit erythrocyte invasion by two different *P. falciparum* strains. In addition, the combination of two or more of some of these proteins was even more effective than any individual protein. This result opens up the possibility of developing a malaria vaccine that contains multiple protein antigens; this approach should overcome any changes to individual protein antigens because changes are unlikely to occur simultaneously in several protein antigens.

In another recent study, researchers identified 11 different parasite proteins that were recognized specifically by antibodies in sera from malaria-resistant individuals. One of these proteins, glutamic-acid-rich protein, which is expressed on the surface of erythrocytes, was expressed during the parasite's early trophozoite (feeding) stage. In preliminary experiments, these researchers found that monkeys that were immunized with either this purified protein or an mRNA vaccine encoding this protein were moderately protected against malaria caused by a high burden of blood-stage *P. falciparum*. It remains to be determined whether combining this protein antigen with others that were found in the malaria-resistant individuals might form the basis of an effective vaccine.

Delivery

In the past 10 to 15 years, lactic acid bacteria, which are "generally regarded as safe," or GRAS, by the U.S. FDA, have been used to facilitate the delivery of human therapeutic proteins (see chapter 5). Lactic acid bacteria immunize humans via the mucosal route (gastrointestinal, respiratory, and

Figure 7.24 Overview of the use of lactic acid bacteria to deliver vaccine antigens. A lactic acid bacterial strain that colonizes the target mucus membrane (generally gastrointestinal, respiratory, or urogenital tract) is isolated and characterized. The strain is then cured of its endogenous plasmids before an essential chromosomal gene is mutated so that the bacterium lacks an essential nutrient and cannot replicate outside of the lab. A gene encoding the target vaccine antigen is inserted into the chromosomal DNA of the mutant and, following protein synthesis, the target vaccine antigen is secreted outside the bacterium.



urogenital tracts), which increases the effectiveness against pathogens that use mucosal surfaces as their major route of entry into the human body. These bacteria may also be used to deliver (nonhuman) proteins that can act as mucosal subunit vaccines. At the present time, there are very few approved mucosal vaccines, and most of these are traditional killed or attenuated bacteria or viruses. These vaccines require refrigeration or may carry the risk of reversion to virulence. Thus, there is a need for effective mucosal delivery systems that avoid these problems. Delivery of antigens by lactic acid bacteria often has the added benefit of inducing both mucosal and systemic immune responses. Moreover, many lactic acid bacterial strains have the ability to act as adjuvants that enhance the immune response induced by the antigen that they have been engineered to encode. Figure 7.24 provides a schematic overview of the development of a lactic acid bacterial subunit vaccine. In this procedure, it is essential that the gene encoding the target antigen be inserted into the bacterial chromosomal DNA rather than on a plasmid. This ensures that no antibiotic resistance genes, often used as plasmid selectable markers, will be transferred to other bacteria and subsequently proliferate in the environment. In addition to the steps outlined in Fig. 7.24, to ensure efficient gene expression, researchers have paid particular attention to the use of inducible promoters and sometimes have had to engineer the codon usage of the target gene to be similar to the codon usage of the host bacterium (see chapter 3). Although lactic acid bacterial vaccines have been shown to be effective with mice (Table 7.2), none of these vaccines is currently in clinical trials.

Scientists are currently working on a wide range of techniques for the effective delivery of various vaccines to individuals being immunized. One recent technique that is still being tested is based on a device called a Nanopatch. Traditionally, and still in common use today, a needle and syringe are used to inject a liquid vaccine into muscle (Fig. 7.25A). However, it has been hypothesized that vaccines might be much more effective if they were delivered into the narrow layer (dermis) that is just beneath the surface of the skin. This layer contains a high density of

Table 7.2 Lactic acid bacterial subunit vaccines successfully tested in mice

Vaccine target	Antigen	Disease target
Helicobacter pylori	Urease B	Stomach ulcers
Tetanus toxoid	Tetanus toxoid fragment C	Tetanus
Streptococcus pneumoniae	PspA, PssA	Bacterial pneumonia
Streptococcus pyogenes	Conserved sequences of the C-repeat region of the M-protein	Skin infections
HIV	V2-V4 loop of gp120, GAG protein	AIDS
Enterotoxigenic *E. coli*	K99 fimbriae	Traveler's diarrhea
Rotavirus	Viral protein 7, 8	Infant's diarrhea
Group B *Streptococcus*	Pilus	Infected pregnant women
Human papilloma virus-16	Protein L2, E6, E7	Cervical cancer
Bacillus anthracis	Protective antigen	Anthrax
Giardia lamblia	Cyst wall protein-2	Giardiasis
Yersinia pseudotuberculosis	Low calcium response antigen	Far East scarlet-like fever
Borrelia burgdorferi	Outer membrane surface protein A	Lyme disease
Listeria monocytogenes	Listeriolysin O	Listeriosis
Influenza virus A	Hemagglutinin, nucleoprotein	Influenza
Staphylococcus aureus		

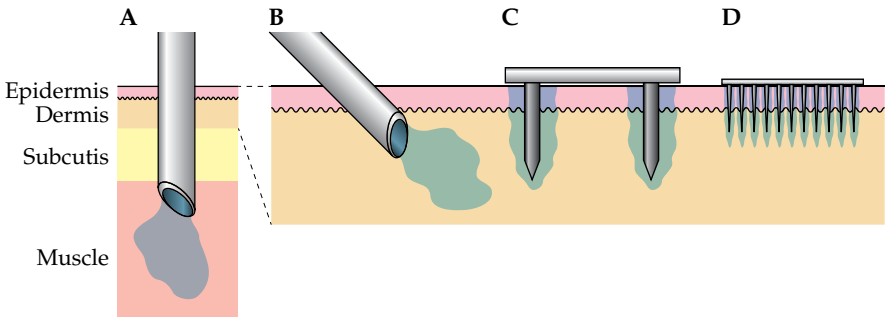


Figure 7.25 Use of various injection techniques to deliver vaccine antigens. **(A)** Intramuscular injection; **(B)** intradermal injection (using a small needle and syringe), where there is a high concentration of antigen-presenting cells; **(C)** the use of microneedles to deliver a vaccine to the dermis; and **(D)** the use of a Nanopatch, where thousands of microprojections are used to deliver the vaccine.

antigen-presenting cells that are essential to generate a protective immune response. Intradermal injection can achieve comparable immune responses to intramuscular injection while using only about one-tenth of the vaccine dose. Unfortunately, given the relative sizes of the dermis and the needle, it is technically difficult to do this effectively. As an alternative to intradermal injection, some researchers developed microneedles to more precisely deliver the vaccine to the skin (Fig. 7.25B and C). More recently, other scientists developed a Nanopatch technique wherein a very small and densely packed array (1 cm^2 of silicon with ~20,000 microprojections on its surface, with individual microprojections invisible to the naked eye) is used to directly target the vaccine to the dermis (Fig. 7.25D). Results with the Nanopatch show that it is up to 10 times more effective (in terms of the decreased amount of vaccine that it required to elicit a similar immune response) than any other vaccine delivery methods, and this increased efficiency comes without the reliance on an added adjuvant, and with only a single vaccination. In several different animals the implementation of this technology could lower the amount of vaccine that is required to produce an immune response by 10- to 100-fold and at the same time alleviate problems regarding fear of the needle (especially in children). This technology has recently been commercialized, and its efficacy in delivering several different vaccines is being tested. In addition to the described benefits of the Nanopatch, vaccines coated onto this device are in a dry format and are therefore stable even in the absence of refrigeration.

In the 21st century the world has faced several coronavirus outbreaks, including SARS in 2002, MERS (Middle East respiratory syndrome) in 2012, and COVID-19 in 2019/2020. For the three coronaviruses causing these diseases, the spike (S) protein that exists as a trimer on the surface of the viral envelope is a key determinant of viral transmission and infectivity. Thus, the S protein is an important target for vaccine development. In one study, to ensure that cloned S protein monomers would trimerize and hence mimic its form in the native viral structure, DNA encoding a 27-amino-acid sequence "trimerization domain" was added to the DNA coding for a portion of the sequence for the S protein from either the MERS or the COVID-19 virus. This subunit vaccine was delivered to mice using microneedle arrays. The results to date (mid-2020) indicate that this approach resulted in the development of a strong antivirus antibody response in injected mice and is considered to be a promising strategy for developing a human vaccine against these viruses.

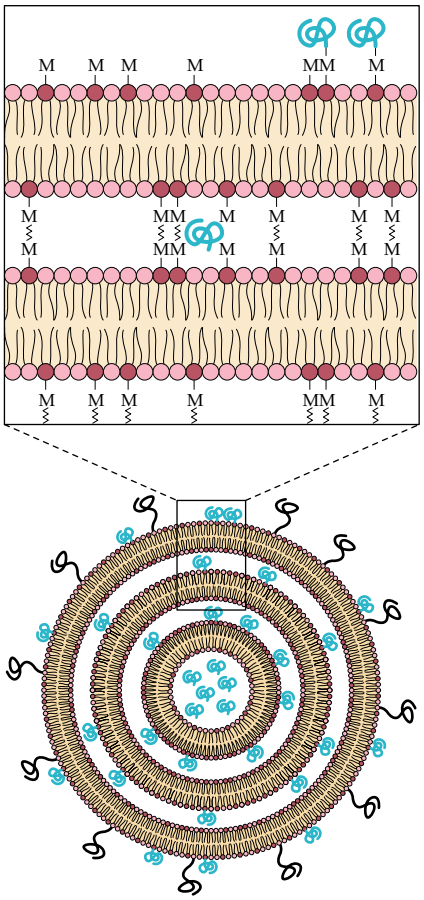


Figure 7.26 Schematic illustration of interbilayer-crosslinked multilamellar vesicles with surface-conjugated VMP001. Sulfhydryl residues on the surface of the VMP001 protein were covalently coupled to the lipid vesicle's surface maleimide (M) moieties. Two different types of lipids were used to prepare this vesicle.

It is important to develop effective strategies to enhance the usefulness of subunit and peptide vaccines. This is because these vaccines typically elicit a much weaker immune response than live vaccines. In fact, it is believed that one of the key requirements to making subunit and peptide vaccines more effective is the development of better adjuvants. This has led some researchers to consider employing nanoparticles as both a delivery system and an adjunct to increase the body's immune response to a vaccine. In one set of experiments to test this possibility, researchers used a complex nanoparticle composed of multilamellar lipid vesicles to deliver a recombinant *Plasmodium vivax* circumsporozoite antigen, VMP001, entrapped in an aqueous core and anchored to lipid bilayer surfaces (Fig. 7.26). It is worth noting here that *P. vivax* is the second most serious malaria-causing parasite (after *P. falciparum*), infecting 130 to 390 million people every year. Immunization with these nanoparticle complexes together with monophosphoryl lipid A (which stimulates the immune response) yielded high-titer and high-avidity (affinity for the antigen) antibody responses in mice against the VMP001 protein. In addition, the immune response against this protein lasted for more than 1 year at 10-fold lower doses than when conventional adjuvants were used together with purified VMP001 protein. While these data are still preliminary, they indicate that the use of nanoparticles is both a promising vaccine delivery and adjuvant strategy. Another group of researchers has recently reported that mice that were vaccinated with peptide vaccines representing portions of the *P. vivax* circumsporozoite antigen VMP001 generated high titers of antibodies against this protein and had significantly delayed onset of the appearance of malarial parasites in the blood.

Genetic Immunization: DNA Vaccines

DNA vaccination is a relatively new technique for protecting animals and humans against disease by injecting them with genetically engineered DNA instead of a protein antigen. The gene on the injected DNA directs cells to produce a protein antigen, effectively acting as a subunit vaccine. As of mid-2020, no DNA vaccines have been approved for human use in the United States. One of the first DNA vaccines to be approved (in 2005) protects horses from West Nile virus. In addition, animal vaccines have been approved to prevent infectious hematopoietic necrosis virus disease in salmon and canine melanoma.

When viruses, cells, or antigens are injected into the body as a vaccine, they may elicit a vigorous immune response initially but subsequently may become less effective over time. Without further antigen stimulation, the immune system often returns to its preimmunization state of activity. Thus, to increase the durability of an immune response after vaccination, a steady supply of an antigenic stimulus is required. To facilitate this requirement, various vectors containing a gene(s) that encodes specific antigens have been used to provide a constant source of those antigens to stimulate an immune response.

Delivery

In early studies, researchers introduced cloned antigen cDNAs into mouse cells by injecting large amounts of the plasmid carrying the target DNA directly into the muscles of test animals. However, effective "genetic

immunization" by direct injection into muscles requires approximately 100 μg per mouse. Alternatively, scientists coated very small gold microprojectiles (the size of a typical bacterium) with *E. coli* plasmid DNA carrying an antigen gene which was under the transcriptional control of an animal virus promoter. A biolistic system was used to deliver the microprojectiles into cells in the ears of mice (see chapter 10 for a more detailed description of the biolistic system). The biolistic system (sometimes called a gene gun) requires 3 to 4 orders of magnitude less DNA (i.e., only 10 to 100 ng per mouse) than the direct injection delivery system. One important and distinctive feature of genetic immunization is that the costly and time-consuming procedure of either purifying an antigen or creating a recombinant vaccine delivery vehicle is bypassed. Moreover, proteins produced by this procedure are more likely to be correctly posttranslationally modified than are proteins that are produced by different host organisms. This is because the proteins are produced and posttranslationally modified within the cells of the immunized animal/human.

An advantage of genetic immunization, besides bypassing the need for purified protein antigens, is that it triggers a response against only the protein encoded on the plasmid and not against the plasmid itself (Table 7.3). In addition, when plasmid DNA is introduced into a mammalian system, only those genes (or cDNAs) under the control of eukaryotic regulatory signals will be transcribed and translated. Antibiotic resistance genes for maintaining the plasmid in a bacterial host (typically *E. coli*), which lack eukaryotic regulatory signals, will not be transcribed or translated, and the same vector can be used to deliver different proteins to an individual at the same time, or the administration of the same gene can be repeated a number of times.

In one early series of experiments, mice were injected in the quadriceps (a large muscle group on the front of the thigh) of both legs with an *E. coli* plasmid carrying the cDNA for influenza A virus nucleoprotein under the transcriptional control of either a Rous sarcoma virus or a cytomegalovirus promoter. Although the expression of the nucleoprotein was too low to visualize on gels, nucleoprotein-specific antibodies were observed in the blood of the test mice 2 weeks after the initial injection. In comparison to control mice, the nucleoprotein-injected mice were significantly protected against the lethal effects of influenza virus infection (Fig. 7.27). Moreover, the nucleoprotein-injected mice were also protected against a different strain of influenza virus. This cross-protection against different influenza virus strains is in sharp contrast to traditional influenza virus vaccines, which are directed against surface antigens of the virus, so that each

Table 7.3 Advantages of genetic immunization over conventional vaccines

Cultivation of dangerous agents is not required.
Since genetic immunization does not utilize any viral or bacterial strains, there is no chance that an attenuated strain will revert to virulence.
Since no organisms are used, attenuated organisms that may cause disease in young or immunocompromised animals are not a problem.
Approach is independent of whether the microorganism is difficult to grow or attenuate.
Production is inexpensive because protein does not need to be produced or purified.
Storage is inexpensive because of the stability of DNA.
One plasmid could encode several antigens/vaccines, or several plasmids could be mixed together and administered at the same time.

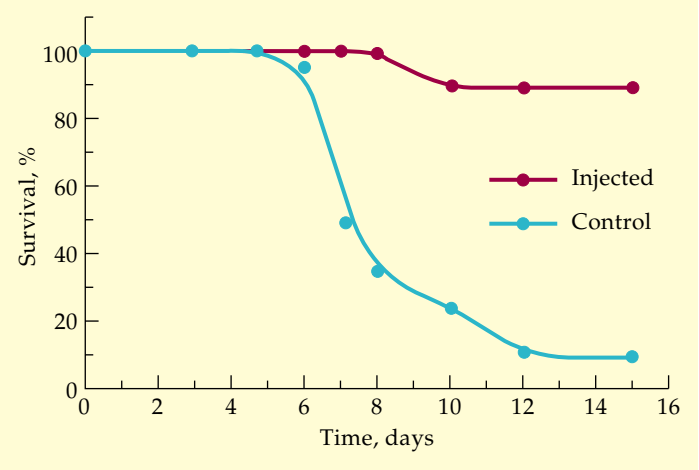


Figure 7.27 Survival of DNA-immunized mice. Injected mice were immunized with DNA that contained the influenza A virus nucleoprotein gene under the control of the Rous sarcoma virus promoter on an *E. coli* plasmid. The control mice were injected with plasmid DNA (without the nucleoprotein) only. The *x* axis represents the number of days after the animals were challenged with the live influenza virus.

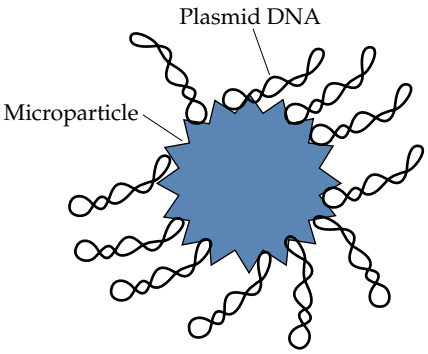
vaccine is specific to only a single strain of influenza virus. In addition, traditional vaccines work only as long as the antigens on the surface of the virus do not change. Unfortunately, the genes for the influenza virus surface antigens mutate at a high rate, which creates significant differences among strains. Although the core (interior) components of the virus, such as the nucleoprotein, are relatively invariant, they can activate the immune system by a mechanism that is different from that of surface antigens.

The fate of the introduced DNA is not known, and it could have the undesirable effect of integrating into the genome of the host cell, possibly disrupting an important gene. However, this risk is currently considered to be extremely low. It is more likely that the DNA will exist for a short period as a nonreplicating extrachromosomal element before it is degraded. Several human clinical trials using DNA vaccines are currently ongoing.

DNA Microparticles

One of the problems with the use of DNA vaccines in large animals and humans compared to mice is that the transfection efficiency of introduced plasmid DNA is often insufficient to generate a protective immune response. One approach to deliver foreign DNA to animal cells utilizes biodegradable microscopic (0.3- to 1.5-μm) polymeric particles with a cationic surface that binds the plasmid DNA (Fig. 7.28). Plasmid DNA is bound to the surfaces of these "microparticles" and is slowly released over a period of 2 to 3 weeks after inoculation of an animal. Using microparticles, it was possible to achieve the same biological effect as with naked DNA with about 250-fold less DNA, demonstrating the potential of this approach. In addition, the level of antibodies induced by the expression of plasmid-encoded genes bound to microparticles was significantly enhanced by (i) addition of the vaccine adjuvant aluminum phosphate and (ii) the use of nanoparticles 0.05 μm in diameter that were coated with poly-L-lysine. In contrast to naked DNA,

Figure 7.28 Schematic representation of the binding of plasmid DNA to the cationic surface of a polymeric microparticle.



DNA bound to microparticles induced potent cytotoxic T-lymphocyte responses at a low dose. In addition to binding plasmid DNA to cationic particles, some researchers have successfully delivered DNA vaccines using the Nanopatch (Fig. 7.25).

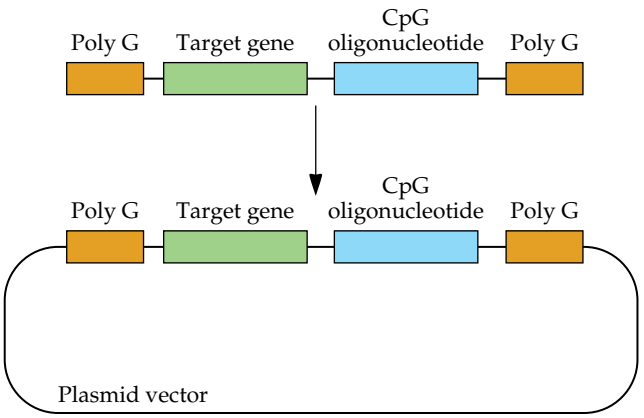
CpG Dinucleotides

Notwithstanding the lack of immunogenicity of the plasmid vector, the presence of one or more unmethylated CpG-rich oligonucleotide fragments (with CpG dinucleotides, the C is often methylated) on the plasmid can stimulate the immune system in inoculated mammalian and aquatic animals. Importantly, in some instances the inclusion of CpG oligonucleotide motifs may replace the use of adjuvants (which can often irritate the animal or person being vaccinated) in the delivery of DNA vaccines (Fig. 7.29). Unlike native microbial DNA, synthetic unmethylated CpG oligonucleotides have a partially or completely phosphorothioated backbone. They also often have a poly G tail at the 3′ end or the 5′ end. The phosphorothioated backbone protects the oligonucleotide from being degraded by nucleases when the plasmid is introduced into an inoculated animal, while the poly G tail enhances the cellular uptake of the plasmid. In one experiment, a plasmid vector that was used to inoculate fish with a DNA vaccine against viral hemorrhagic septicemia was modified to include several CpG oligonucleotide motifs, with the result that the immunogenicity of the vaccine was significantly increased. In this experiment, when vectors with zero, two, or four regions of unmethylated CpG-rich oligonucleotide were compared in terms of the immune response that was elicited by a DNA vaccine against viral hemorrhagic septicemia, the plasmid that contained four regions of unmethylated CpG-rich oligonucleotide was the most immunogenic.

Mucosal Immunity

While several different delivery methods have been developed for DNA vaccines, to date, most have been delivered either by intramuscular or intradermal injection. Although these vaccines can induce a potent immune response, they do not induce mucosal immunity. Mucosal immunity can

Figure 7.29 Construction of plasmids used for delivery of DNA vaccines. **(A)** Both the CpG oligonucleotide and the target gene are flanked by Poly G regions. **(B)** The target gene and the CpG oligonucleotide regions are independent of one another. Plasmids may contain several CpG oligonucleotide regions.



prevent pathogens from entering the body, while systemic immunity deals with pathogens only once they are inside the body. This is an important consideration because mucosal surfaces, including the respiratory, intestinal, and urogenital tracts, are the major sites of infection of many microorganisms. Because of the protective barriers of the mucosal surfaces, traditional antigen-based vaccines are largely ineffective unless they are administered with specific agents that penetrate or bind to the mucosa, i.e., mucosal adjuvants.

Mucosal immunity induces a response separate and distinct from systemic immunity. The antibodies produced as part of the mucosal immune response restrict not only mucosal pathogens, but also microorganisms that initially invade mucosal surfaces and then cause systemic disease. Many mucosal vaccines are live attenuated organisms that infect mucosal surfaces and are effective at inducing mucosal responses. Of these, oral polio vaccines and both attenuated *Salmonella enterica* serovar Typhi Ty21a and *V. cholerae* vaccines are licensed for use in humans.

Table 7.4 Methods used to introduce DNA vaccines

Intramuscular injection
Microprojectile bombardment
Electroporation
Microparticles
Inside a bacterium
Inside a virus
Cationic liposomes
Nanoparticles
Together with cytokines
Together with adjuvants
Using a MIDGE vector

DNA vaccines that are designed for delivery to mucosal surfaces are similar in principle to those used for intramuscular or intradermal delivery. To increase plasmid uptake and decrease its subsequent degradation, various methods of formulating DNA have been tried (Table 7.4). For example, cationic (positively charged) liposomes have been used to deliver DNA (which has a negatively charged phosphate backbone) to the respiratory tract, and DNA entrapment in biodegradable microparticles has been used for the oral delivery of foreign DNA. Moreover, to improve the potency of DNA vaccines for humans, a number of strategies have been devised, including using plasmids that, in addition to encoding a target gene, also express a cytokine(s), such as interleukin-2 (IL-2), IL-10, IL-12 (which can act as an intercellular mediator, modulating the activity of white blood cells in the generation of an immune response), granulocyte macrophage colony-stimulating factor (which is part of the immune system that leads to a proliferation of macrophages), or (as discussed above) unmethylated CpG regions in bacterial DNA that can function as adjuvants and stimulate an immune response to DNA vaccines. In fact, the inclusion of CpG motifs in plasmid DNA vaccines induces the expression of proinflammatory cytokines, such as IL-12.

Electroporation

A range of systems, including liposomes, live vectors (bacteria and viruses), and a wide range of adjuvants that increase the immune response (bacterial toxins, carboxymethylcellulose, lipid derivatives, aluminum salts, and saponins), have been tried for delivery of DNA to different cell types. Of necessity, various optimization strategies are tested in mice before they are tried on larger animals and then on humans, with no guarantee that an approach that works well in mice will also be a successful strategy in humans. Nevertheless, given the many perceived advantages of genetic immunization over the use of conventional vaccines (Table 7.5), this has become a very active area of research. For example, in the development of anticancer vaccines, electroporation has been used to increase the transfection of DNA encoding target antigens (Fig. 7.30). With this approach, DNA is injected intramuscularly, and the skeletal muscle is immediately electrically stimulated with a (millisecond) pulse generator. The electrical pulse causes temporary pores or gaps in the cell membranes to open, enabling the DNA

vaccine to enter the cells. Once inside the cells, the DNA vaccine directs the synthesis of the target antigen. Despite the fact that this procedure, which causes some patient discomfort, results in local tissue injury and inflammation, it is tolerated by patients without the need for any anesthetic, and there do not appear to be any long-term negative side effects to delivering DNA in this way. The electrical pulse increases the transfection efficiency of the vaccine DNA by ~1,000-fold, and it is becoming a method of choice for clinical administration of certain DNA vaccines.

Shigella

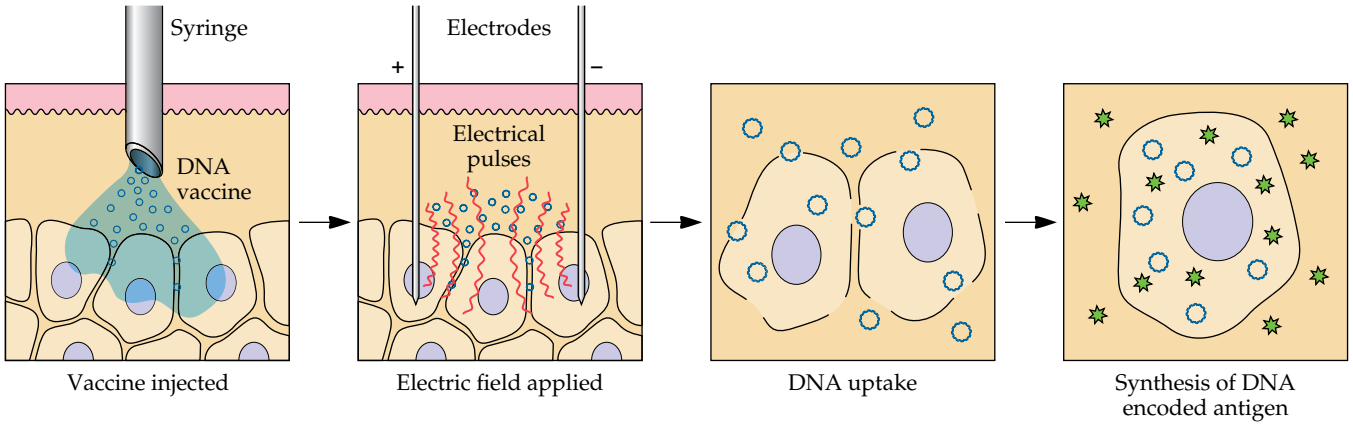
A modified strain of the invasive intestinal bacterium *Shigella flexneri* has been developed to facilitate the delivery of foreign DNA into animal cells for genetic immunization (Fig. 7.31). *Shigella* can enter animal epithelial cells, escaping the phagocytic vacuole, and the bacterium can direct plasmid DNA to the nucleus of the host cell, where its transcription is directed from a eukaryotic promoter included on the plasmid. *Shigella* is normally a human pathogen and, as such, would not be an acceptable DNA delivery system. Therefore, to use *Shigella*, it was first necessary to construct a nonpathogenic version of the wild-type organism. This was done by engineering the bacterium to be (i) toxin deficient (by deleting the toxin gene) and (ii) unable to proliferate in host cells. The latter was accomplished by making a deletion mutation in the *Shigella asd* gene, which encodes the enzyme aspartate β-semialdehyde dehydrogenase. This enzyme is normally involved in the synthesis of diaminopimelic acid, a bacterial cell wall constituent; therefore, the mutant strain cannot grow unless diaminopimelic acid is added to the growth medium. *Shigella* strains with the *asd* mutation can invade animal epithelial cells and deliver their plasmid DNA; however, the *Shigella* cells are unable to proliferate in host cells.

Determination of the safety of using *Shigella* as a vector for the delivery of DNA to animal cells must await the results of human trials; however, the results of experiments with guinea pigs are promising. The greatest potential advantage of this approach is that because *Shigella* is an intestinal bacterium, DNA for vaccination may be delivered orally, greatly simplifying the delivery of a variety of vaccines. In addition to *Shigella*, scientists

Table 7.5 Use of DNA vaccines in clinical trials of cancer vaccines

Cancer type	Phase	Trial start date
Breast cancer	I	2015
Breast cancer	I	2015
Breast cancer	I	2015
Breast cancer	I	2016
Breast cancer	I	2018
Prostate cancer	II	2011
Prostate cancer	II	2013
Prostate cancer	I	2015
Prostate cancer	II	2017
Pancreatic cancer	I	2018
Melanoma	I	2018
Glioblastoma	I/II	2018
Renal cell carcinoma	II	2019
Anal neoplasm	II	2018
Urothelial carcinoma	I/II	2018
Cervical cancer	I/II	2015
Cervical cancer	II	2015
Cervical cancer	I/II	2017
Cervical cancer	I	2017
Cervical cancer	II	2017
Cervical cancer	I/II	2018
Cervical cancer	II	2018
Cervical cancer	II	2018
Cervical cancer	III	2019
Ovarian cancer	II	2017

Figure 7.30 Schematic representation of the transduction of human cells with vaccine plasmid-encoded DNA by electroporation.



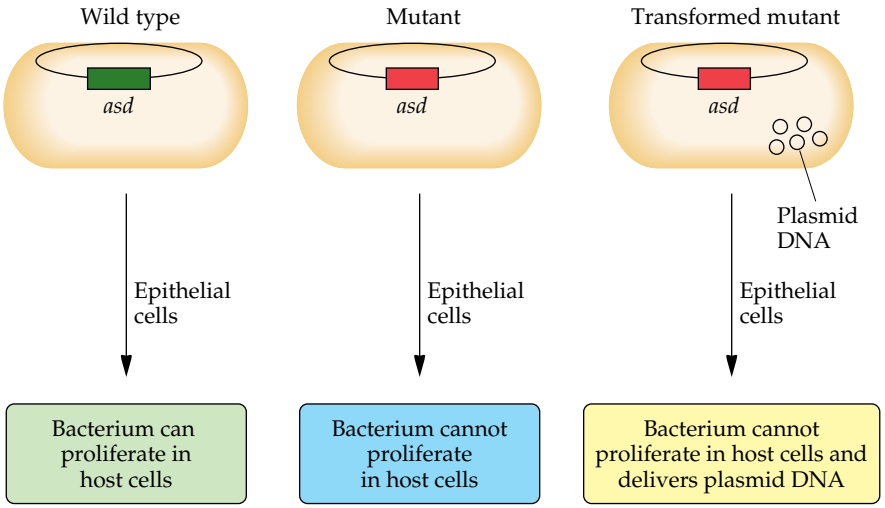
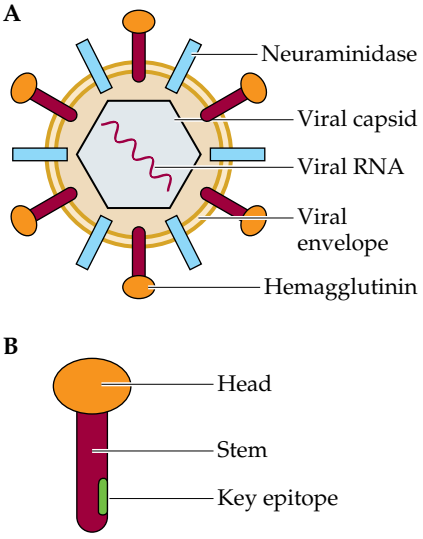


Figure 7.31 Use of a nonpathogenic strain of *S. flexneri* to deliver foreign DNA to mammalian epithelial cells. While the wild-type strain can proliferate in mammalian epithelial cells, a strain of *Shigella* with a deletion mutation in the *asd* gene, which encodes the enzyme β-semialdehyde dehydrogenase, is unable to proliferate in these cells and can be used as a live vector for foreign DNA.

are currently working to modify lactic acid bacteria, which are already being used for the delivery of subunit vaccines, as a means of delivering DNA vaccines to their target site.

Figure 7.32 Schematic representation of influenza virus. **(A)** The major parts of the virus are shown. The viral capsid consists of multiple copies of the viral M-protein. Although only one RNA strand is shown, the influenza viral genome consists of eight separate single-stranded RNAs, each complexed with proteins. **(B)** Details of the hemagglutinin protein. In the virus, this protein exists as a trimer.



Influenza

The hemagglutinin protein on the surface of the influenza viral envelope is a key determinant of the binding of the virus to the sialic acid residues of target cells (Fig. 7.32). Unfortunately, the DNA encoding the hemagglutinin protein rapidly acquires mutations so that the protein is not a useful target in the development of anti-influenza vaccines. However, a portion of the stem domain of the hemagglutinin protein is relatively invariant and is a target in the development of DNA vaccines against multiple influenza strains. Other relatively invariant proteins, including influenza nucleoprotein, matrix proteins, and the RNA-directed RNA polymerase catalytic subunit, are also being tested as possible universal influenza DNA vaccines. Vaccination of mice with DNA encoding each one of these proteins has been found to decrease the viral load and enhance survival against different strains of influenza virus in mice. In addition, by altering the codon usage in the DNA vaccine to more closely resemble that used by mice, the observed mouse antibody titer was increased up to four-fold. It remains to be seen whether this novel approach is sufficiently effective in humans to be the basis of a universal influenza vaccine.

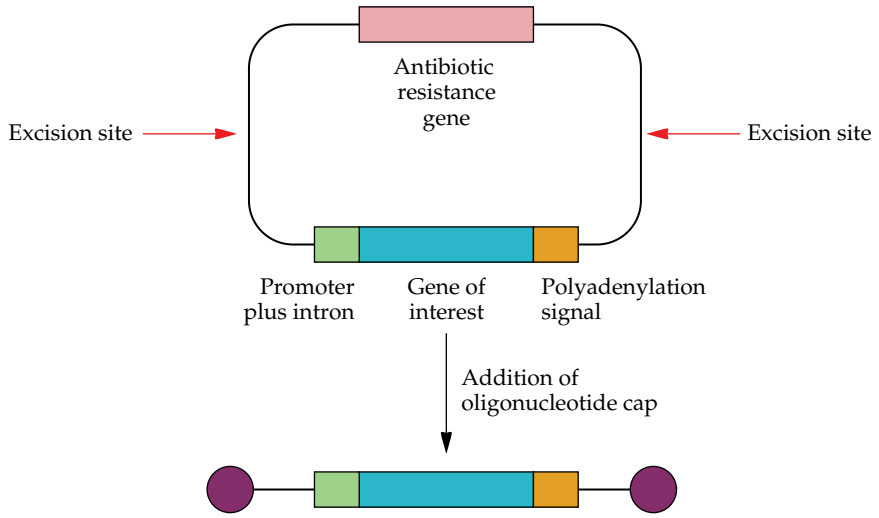
One key question in the use of influenza DNA vaccines is how to increase the immune response to the vaccine. On the one hand, mineral salts such as alum have been observed to result in up to a 5-fold increase in hemagglutinin DNA vaccine-induced antibody titers. Coadministration of cytokines such as interleukin-6 also increases the effectiveness of hemagglutinin DNA vaccines. In addition, only a very small fraction of naked plasmid DNA is able to reach the nuclei of target cells; however, both cationic lipids that form liposomes and nanoparticle-forming polymers have been observed to significantly enhance the uptake of plasmid DNA into target cell nuclei.

MIDGE Vectors

Some limitations of plasmid-based vaccines are (i) the necessity for strong promoters that function *in vivo* to selectively transcribe the introduced DNA, (ii) low levels of foreign-gene expression resulting from differences in codon usage between the introduced gene (often of viral, bacterial, or parasitic origin) and the animal being inoculated, and (iii) the presence of antibiotic resistance marker genes on the plasmid vector. To avoid the use of antibiotic resistance genes, researchers have developed a series of minimalistic immunogenically defined gene expression (MIDGE) vectors (Fig. 7.33). Following insertion of the gene of interest into a MIDGE vector, the antibiotic resistance gene that is initially employed as a selectable marker is excised from the vector, and an oligonucleotide cap is added specifically to both ends of the linearized DNA that contains only the gene of interest with a promoter/intron and a polyadenylation signal. The excised DNA fragment containing the antibiotic resistance gene, which is not protected by the oligonucleotide caps, is degraded by exonucleases. The capped ends are resistant to exonuclease digestion because most exonucleases cannot cleave DNA strands unless they contain a terminal 3′-OH group. The purified, capped linear MIDGE vector is then used directly for transfection. These vectors have been used successfully as a substitute for plasmid vectors.

At the present time, there is a commercial recombinant subunit vaccine (ENGERIX-B) against hepatitis B virus (HBV) that is based on the small hepatitis B surface antigen adsorbed onto aluminum hydroxide. This vaccine is administered by intramuscular injection. Unfortunately, the vaccine does not induce a protective immune response in about 10% of the people who are vaccinated. To remedy this problem, scientists developed DNA vaccines encoding the small or the large hepatitis B surface antigen. In both instances, the antigen genes were spliced into MIDGE vectors (Fig. 7.34). The immunogenicity of these vectors was then assessed in mice and pigs. In both animal models, vectors encoding the small protein induced stronger responses than vectors encoding the large protein. This difference in immune response was thought to be due to the fact that the

Figure 7.33 Use of a MIDGE vector to produce a capped linear DNA sequence containing the gene of interest, a promoter, an intron (which facilitates expression of the gene of interest), and a polyadenylation signal.



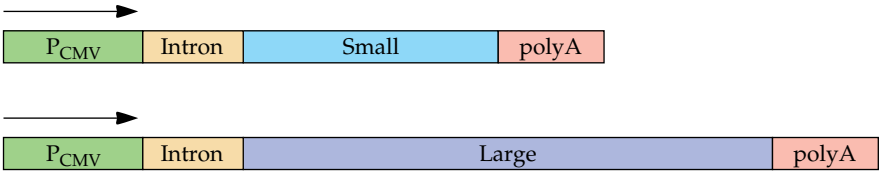


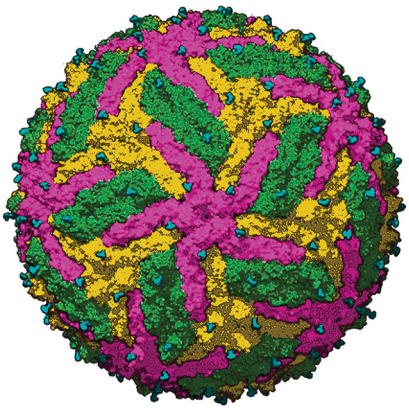
Figure 7.34 Schematic representation of the expression vectors encoding the small and large antigens from hepatitis B virus. The promoter (P) is from human cytomegalovirus (CMV). The arrow indicates the direction of transcription. One vector encodes the hepatitis B virus small surface antigen, while the other vector encodes the hepatitis B virus large surface antigen.

large hepatitis B surface antigen was not secreted from transfected cells, despite the presence of a heterologous secretion signal. Moreover, in pigs, the MIDGE vector that encoded the small protein (delivered in a cationic lipid) elicited an immune response that was approximately the same magnitude as the commercialized protein vaccine (which included an aluminum hydroxide adjuvant). The small protein-specific antibody levels were higher than antibody levels that would be considered to be protective in humans. Although the antibody titer in immunized animals declined by six months after the final immunization, the decline was less pronounced when the DNA vaccine was used. This DNA vaccine remains to be tested in humans.

Cancer

Since tumor antigens are typically weakly immunogenic, they often induce only a low level of immunity and in some cases result in tolerance. To increase the immunogenicity of a DNA vaccine encoding a tumor antigen, the use of adjuvants and immunomodulatory molecules has been extensively employed in clinical trials. Some of the immunomodulatory molecules that have been used include cytokines such as human interleukin-2 (IL-2), IL-7, IL-12, IL-15, and granulocyte-macrophage colony-stimulating factor (GM-CSF). Clinical trials conducted over the last few years have shown promising results, particularly when DNA tumor vaccines were used in combination with other forms of vaccines. Delivery of therapeutic gene-based vaccines by physical methods (electroporation and gene gun) has amplified the immune responses induced by therapeutic vaccines against cancer. Several recent clinical trials of different types of cancer in which DNA vaccines have been used are summarized in Table 7.5. In this table, several cancers are listed more than once, indicating that multiple DNA vaccines have been tested.

Figure 7.35 Schematic representation of the surface proteins of the Zika virus capsid (notwithstanding that this is an enveloped virus).



Zika Virus

Zika virus (Fig. 7.35), a *Flavivirus* related to the dengue, yellow fever, Japanese encephalitis, and West Nile viruses, is an enveloped virus that is spread by mosquitos. Zika virus usually causes only mild cold-like symptoms that are typically treated with acetaminophen (Tylenol, paracetamol). However, Zika virus can also cause miscarriages and stillbirths. In 2015 and 2016, more than 500,000 suspected cases of Zika virus infections (and several thousand congenital birth defects) were reported in South and Central America and the Caribbean. It has been suggested that the rapid

emergence of this more virulent form of Zika virus was caused by a mutation in the virus. By late 2016, there were 18 different Zika virus vaccines in various stages of development (either in preclinical or phase I clinical trials). One group of researchers reported complete protection of mice following a single immunization with a DNA vaccine, a plasmid encoding Zika virus premembrane and envelope proteins, on a plasmid. This vaccine was directed against a strain of Zika virus that was endemic in northeast Brazil, although other equally successful anti-Zika virus DNA vaccines have also been reported. These results suggested that a human Zika virus vaccine should be readily attainable. Strangely, by the summer of 2017, the number of reported Zika virus cases had dropped dramatically, possibly as a consequence of populations in many affected areas acquiring herd immunity. Nevertheless, Zika virus remains a problem, and several Zika virus vaccines are currently in phase II clinical trials.

Dental Caries

The Gram-positive, facultatively anaerobic bacteria *Streptococcus mutans* and *Streptococcus sobrinus* are considered to be the primary causative agents of dental caries (tooth decay). These organisms colonize tooth surfaces and metabolize sucrose to produce lactic acid, which causes the tooth enamel to become vulnerable to decay. Sucrose is also used by the bacteria to produce a sticky, extracellular, dextran-based polysaccharide (glucan) that facilitates *Streptococcus* cells' adherence to one another and to tooth surfaces, forming plaque. It is the combination of plaque and acid that leads to tooth decay.

Two regions of one of the adhesion proteins found on the surfaces of *S. mutans* and *S. sobrinus* cells are important in the initial adherence of these bacteria to tooth surfaces: one sequence is rich in alanine residues, while the other is rich in proline residues. Another important component of the mechanism of tooth decay is the enzyme glucosyltransferase, which is responsible for the synthesis of glucan, an insoluble extracellular polymer of glucose moieties. A DNA vaccine designed to prevent dental caries included the coding sequences for the alanine- and proline-rich peptide, as well as the C-terminal domain of the *Streptococcus* glucosyltransferase. This C-terminal domain is necessary to anchor the extracellular glucosyltransferase to the bacterial cell surface. This DNA vaccine therefore encoded two separate peptides, both derived from surface proteins that facilitate the binding of *Streptococcus* cells to the tooth surface. In an attempt to overcome the tendency of many DNA vaccines to induce only a weak immune response, the DNA vaccine construct encoded two additional elements that bind to proteins on antigen-presenting cells. First, DNA encoding the extracellular domain of cytotoxic T-lymphocyte antigen 4 (CTLA4), which binds to the B7 protein on the surfaces of antigen-presenting cells, was included in the construct (Fig. 7.36). Second, DNA encoding the Fc region of an immunoglobulin G molecule, which can bind to the Fc receptor on the antigen-presenting cell, was included in this construct. The use of both of these peptides was designed to specifically target the above-mentioned multidomain fusion protein (Fig. 7.36) to antigen-presenting cells and thereby amplify the immune response and enhance the efficacy of the vaccine. In fact, rabbits immunized with this vaccine, either intranasally or intramuscularly, displayed a significantly enhanced, specific systemic and

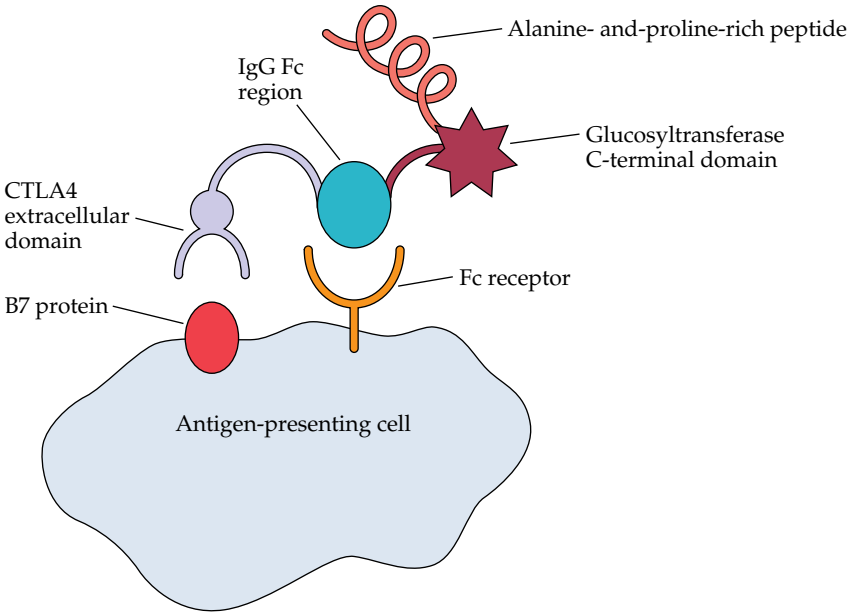


Figure 7.36 Schematic representation of a multidomain protein encoded by a DNA vaccine. Two domains, the CTLA4 extracellular domain and the immunoglobulin G (IgG) Fc region, are designed to target the other two portions of the molecule, i.e., the glucosyltransferase C-terminal domain and the alanine-proline-rich peptide, to antigen-presenting cells.

mucosal immune response compared to immunization with only the alanine- and proline-rich peptide fused to the C-terminal domain of glucosyltransferase. Moreover, it was subsequently shown that this DNA vaccine could provide significant protection against dental caries in rats that were challenged with *S. mutans* and *S. sobrinus*. Although the problem of delivering the DNA vaccine to humans still needs to be addressed so that its clinical efficacy can be tested, this is a very promising strategy that could be enormously beneficial to human populations. However, technical difficulties in developing a vaccine for human dental caries, coupled with a lack of strong economic interests, have so far thwarted the further development of such a vaccine.

Engineered Attenuated Vaccines

Genetic manipulation may be used to construct modified organisms (bacteria or viruses) that are used as live recombinant vaccines. These vaccines are either nonpathogenic organisms that have been engineered to carry and express antigenic determinants from a target pathogenic agent or engineered strains of pathogenic organisms in which the virulence genes have been modified or deleted. In these instances, as part of a bacterium or a virus, the important antigenic determinants are presented to the immune system with a conformation that is very similar to the form of the antigen in the disease-causing organism. Although successful in some cases, purified antigen alone often lacks the native conformation and elicits a weak immunological response.

Herpes Simplex Virus

There is currently no approved vaccine for herpes, and given the failure of subunit vaccines directed against glycoprotein D (the major antigenic determinant of this virus) to completely protect humans against HSV, scientists tried another approach to develop a vaccine against this virus. More than 3.7 billion people under the age of 50 around the world are infected with HSV-1, while over 500 million have HSV-2. In addition, HSV-2 can contribute significantly to the spread of HIV. A mutant of HSV was generated by deleting the glycoprotein D gene (Fig. 7.37). Since glycoprotein D is required for HSV to enter into cells, deleting the gene for this protein prevented the cell-to-cell spread of the virus, making the attenuated strain completely safe and ensuring that it elicited antibodies against other, intact, surface proteins, including glycoproteins B, C, and H,L (since a vaccine based on glycoprotein D had already been shown to not be effective in humans). In order to make an attenuated vaccine against HSV-2, researchers grew mutant HSV-2 (with the glycoprotein D gene deleted from its genome) in a cell line that expresses the HSV-1 version of glycoprotein D. The mutant HSV-2 virus assembled the HSV-1 glycoprotein D produced by the cell into its envelope membrane so that when it was introduced into a mouse, the mutant HSV-2 with HSV-1 glycoprotein D was able to enter the mouse's cells. However, upon replication of the mutant HSV-2, the progeny viruses were not able to infect new cells because the gene encoding glycoprotein D was not present. The HSV-2 surface proteins activated the mouse's immune system to produce antibodies against HSV-2. In this case, the mutant HSV-2 evoked the immune system's antibody-dependent cell-mediated cytotoxicity (ADCC) response (Fig. 7.38). In the ADCC response, antibodies recognize HSV-2 infected cells that present the viral antigens on their surface. Natural killer (NK) cells bind to the Fc portion of the antibodies on the surface of the infected cells and mediate the killing of the infected cells. Subsequent infection by wild-type HSV-2 activates this defense response. Thus, when mice that were first inoculated with the attenuated strain of HSV-2 that lacked glycoprotein D were challenged with 1,000 times the amount of HSV-2 generally needed to cause an infection, the mice survived and did not develop

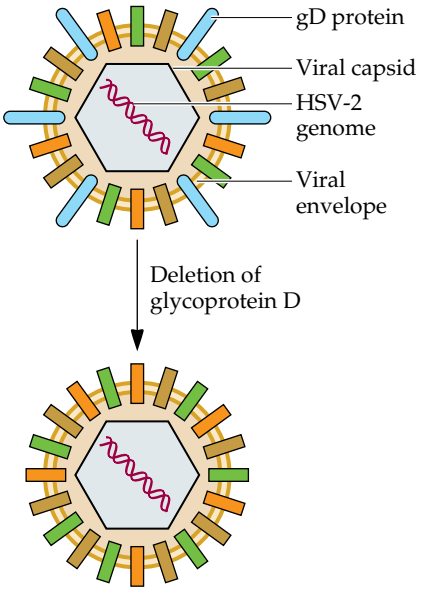
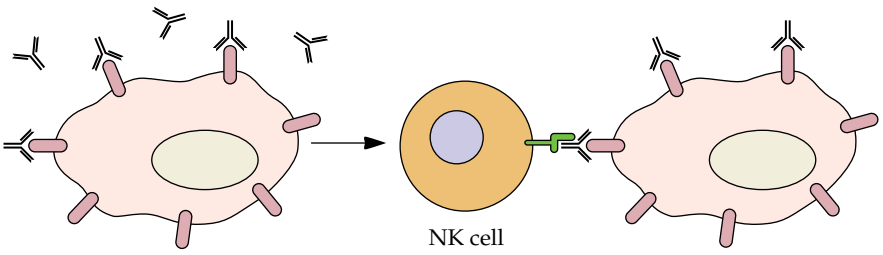


Figure 7.37 Deletion of viral glycoprotein D from HSV-2 with the result that the mutant virus can act as an attenuated vaccine. The orange, green, and brown shapes represent viral envelope proteins.

Figure 7.38 Schematic representation of antibody-dependent cell-mediated cytotoxicity. Following binding of antibodies to surface antigens of infected cells, the CD16 proteins of the natural killer cells recognize the Fc region of the antibodies bound to the infected cells. Eventually, the CD16 proteins become cross-linked and the infected cells are destroyed.



herpes. Moreover, this vaccine is also effective against HSV-1. Given the highly successful results of this vaccine with mice, human clinical trials are expected to begin soon.

Cholera

It is usually advantageous to develop a live vaccine, because they are generally much more effective than killed or subunit vaccines. The major requirement for a live vaccine is that no virulent forms be present in the inoculation material. With this objective in mind, a live attenuated cholera vaccine has been developed. Cholera, caused by the bacterium *V. cholerae,* is a fast-acting intestinal disease characterized by fever, dehydration, abdominal pain, and diarrhea. It is transmitted by contaminated drinking water. In many developing countries, the threat of cholera is a real and significant health concern whenever water purification and sewage disposal systems are inadequate.

The burden of cholera is greatest in sub-Saharan Africa and South Asia. Among the 1.4 billion people at risk of cholera, 2 to 5 million cholera cases are estimated to occur each year. The current vaccine against cholera (Dukoral) generates only moderate protection, typically lasting from about 3 to 6 months. Moreover, this vaccine requires refrigeration, and it is not effective against all strains of *V. cholerae*. Thus, there is a high public health need to vaccinate against cholera to help eradicate the global cholera burden.

Cholera toxin is a multimeric protein complex that consists of six protein subunits: one copy of the A subunit and five copies of the B subunit (Fig. 7.7). The B subunits bind specifically to an intestinal mucosal cell receptor; the A_1 portion of the A subunit has ADP-ribosylation (toxin) activity, while the A_2 portion of the A subunit noncovalently links the A subunit to the B subunits. Since *V. cholerae* colonizes the surface of the intestinal mucosa, it was reasoned that an effective cholera vaccine should be administered orally and directed to these tissues. With this in mind, a strain of *V. cholerae* was created with part of the coding sequence for the A_1 peptide deleted. This strain cannot produce active enterotoxin; therefore, it is nonpathogenic and is a good candidate for a live vaccine. Initially, a tetracycline resistance gene was incorporated into the A_1 peptide DNA sequence on the *V. cholerae* chromosome. This insertion inactivated the A_1 peptide activity and also made the strain resistant to tetracycline. Although the A_1 peptide sequence was disrupted, this strain is not acceptable as a vaccine because the tetracycline resistance gene can excise spontaneously, thereby restoring enterotoxin activity.

Next, a strain carrying a defective A_1 peptide sequence that could not revert to virulence was constructed (Fig. 7.39). In this strategy, the tetracycline resistance gene was exchanged for a deleted A_1 peptide DNA sequence. To do this, (i) a plasmid containing the cloned DNA segment for the A_1 peptide was digested with the restriction enzymes ClaI and XbaI to remove a portion of the A_1 peptide-coding sequence. (ii) To recircularize the plasmid, an XbaI linker was added to the ClaI site and then cut with XbaI. (iii) T4 DNA ligase was used to join the plasmid at the XbaI sites, thereby deleting a 550-base pair segment from the middle of the A_1 peptide-coding region. This deletion removed 183 of the 194 amino acids of the A_1 peptide. (iv) Then, by conjugation, the plasmid containing the deleted A_1 peptide-coding sequence was transferred into the *V. cholerae*

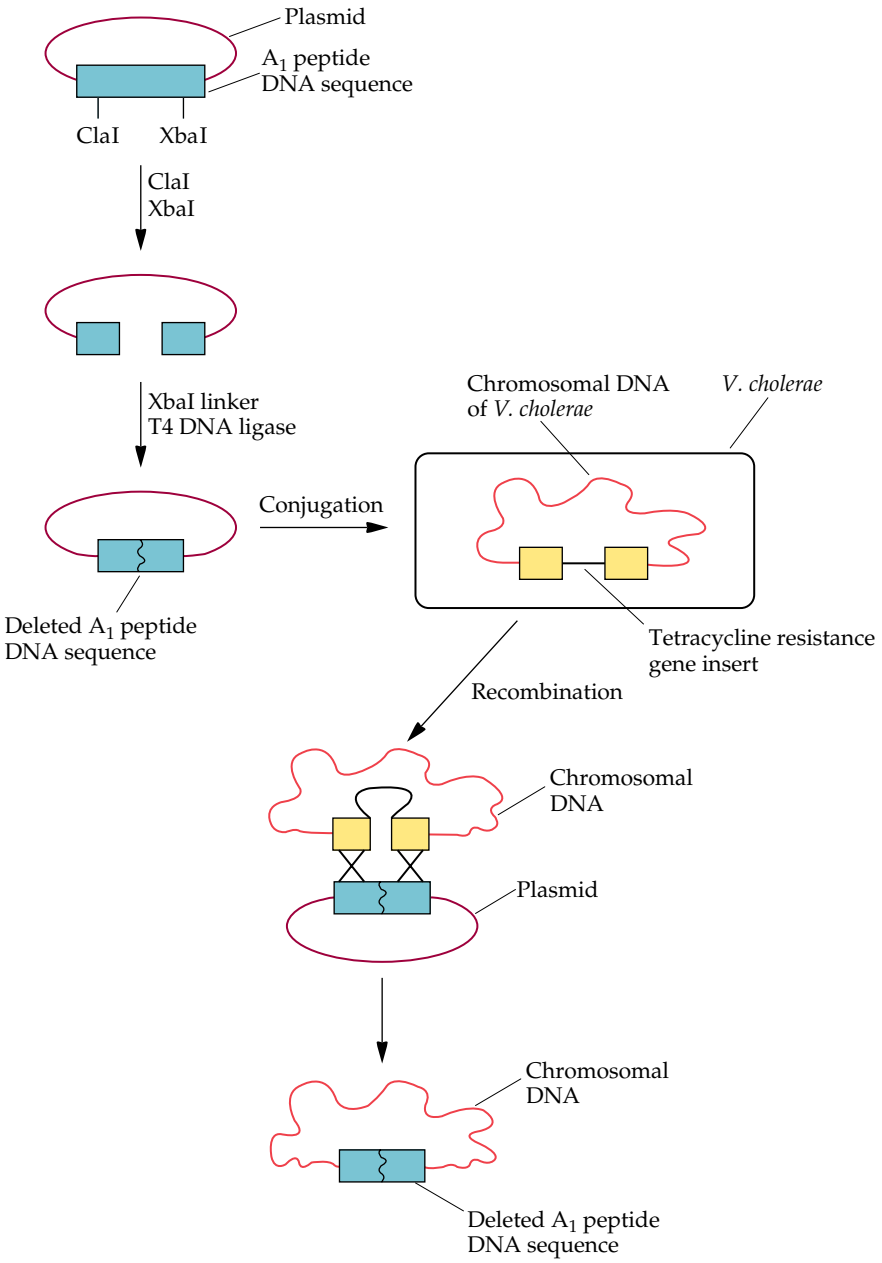


Figure 7.39 Strategy for deleting part of the cholera toxin A_1 peptide DNA sequence from a strain of *V. cholerae*. Note that the tetracycline resistance gene is introduced into the gene for the A1 peptide as part of a transposon. This construct no longer makes A1 peptide but cannot be used as a vaccine because it is possible for the transposon to be excised, with the result that A1 synthesis and pathogenicity are restored.

strain carrying the tetracycline resistance gene within its A_1 peptide DNA sequence. (v) Recombination (a double crossover) between the remaining A_1 coding sequence on the plasmid and the tetracycline resistance gene-disrupted A_1 peptide gene on the chromosome replaced the chromosomal A_1 peptide-coding sequence with the homologous segment on the plasmid

carrying the deletion. (vi) After growth for a number of generations, the extrachromosomal plasmid, which is unstable in *V. cholerae*, was spontaneously lost. (vii) Cells with an integrated defective A_1 peptide coding sequence were selected on the basis of their tetracycline sensitivity. The desired cells no longer had the tetracycline resistance gene but carried the A_1 peptide sequence with the deletion.

A stable attenuated strain with a portion of the A_1 peptide sequence deleted was selected in this way. This strain did not produce active enterotoxin but nevertheless retained all the other features of the pathogenic form of *V. cholerae*; that is, *V. cholerae* with an A_1 peptide containing a deletion is a good vaccine candidate because the bacterium that synthesizes only the A_2 and B peptides is as immunogenic as the native bacterium. Unfortunately, when this strain was evaluated in clinical trials to test its effectiveness as a cholera vaccine, the results were equivocal. While the vaccine conferred nearly 90% protection against diarrheal disease in volunteers, it induced side effects (which were not observed with the subunit/traditional vaccine against cholera, Dukoral, described earlier) in some of those who were tested. In the meantime, a similar live attenuated anticholera vaccine completed clinical trials in February 2020 and has now been approved by the U.S. FDA. This vaccine, VAXCHORA, is a live, attenuated bacterial vaccine that is taken orally; it contains a suspension of the *V. cholerae* strain CVD 103-HgR. This strain was constructed by deleting a portion of both copies of the *ctxA* gene encoding the cholera A toxin subunit. It is unable to synthesize an active cholera toxin but is still able to synthesize the immunogenic but nontoxic B subunit encoded by the *ctxB* gene. In addition, a genetic marker was inserted into the *V. cholerae* hemolysin gene so that the vaccine strain could be distinguished from wild-type *V. cholerae* O1. While the recently approved vaccine is not effective against all strains of *V. cholerae*, similar vaccines that are effective against other strains of this bacterium are in clinical trials.

Salmonella Species

Other attempts to engineer attenuated strains of pathogenic bacteria that could be used as live vaccines have involved deletions in chromosomal regions that code for various essential functions. At least two deletions are preferred, because the probability that both sets of functions can be simultaneously reacquired is very small. It is assumed that a "doubly deleted" strain would have a very limited ability to proliferate (depending upon the two genes that are deleted) when it is used as a vaccine, thereby curtailing its pathogenicity while allowing it to stimulate an immunological response.

Strains of the genus *Salmonella* (a Gram-negative bacterium) cause enteric fever, infant death, typhoid fever, and food poisoning. Therefore, an effective vaccine against these organisms is needed. A number of different genes have been deleted to attenuate various *Salmonella* strains (Table 7.6). These mutations can be grouped into three basic categories: mutations in (i) biosynthetic genes, (ii) regulatory genes, and (iii) genes involved in virulence. In addition, strains with more than one deletion have been constructed. For example, one double-deletion strain has deletions in an *aro* gene, which encodes an enzyme involved in the biosynthesis of aromatic amino acids, and in a *pur* gene, which encodes an enzyme involved in purine metabolism, both essential functions. Double-deletion

Table 7.6 Some deleted genes and their functions in the development of attenuated strains of *Salmonella* spp.[a]

Deleted gene	Gene function
galE	Synthesis of lipopolysaccharide; decreases toxicity from galactose
aroA, *aroC*, or *aroD*	Synthesis of chorismate, an aromatic amino acid precursor and a PABA precursor; PABA is involved in the synthesis of iron chelators
purA or *purE*	Synthesis of purines
asd	Peptidoglycan and lysine biosynthesis
phoP and *phoQ*	Regulation of acid phosphatases and genes necessary for survival in the microphage
cya	Encodes adenylate cyclase, which is involved in cAMP synthesis
crp	Encodes cAMP receptor; regulates expression of proteins involved in transport and breakdown of carbohydrates and amino acids
cdt	Involved in tissue colonization by the bacterium
dam	Encodes DNA methylase; appears to be a master switch for 20–40 different virulence genes
htrA	Encloses a stress-induced polypeptide; results in significantly reduced persistence in human tissues
fur	Iron uptake regulatory protein; needed to maintain adequate iron concentration
rpoS	Sigma factor that regulates stationary-phase growth and sensitivity to environmental stressors
ssaV	Component of the secretion apparatus; mutation reduces replication and survival in host and reduces spreading
galE	Encodes UDP-galactose epimerase needed to synthesize LPS outer core and O-antigen and thus necessary for virulence
pmi	Encodes phosphomannose isomerase needed for synthesis of GDP-mannose for LPS O-antigen (necessary for virulence)

[a]cAMP, cyclic AMP; PABA, *p*-aminobenzoic acid; LPS, lipopolysaccharide.

strains, which can be grown on a complete and enriched medium that supplies the missing nutrients, generally establish only low-level infections, since their host cells contain only a very low level of the metabolites that they require for growth. Typically, the virulence of these mutants is reduced by 100-fold or more from the level of the wild-type strain. These attenuated *Salmonella* strains are effective oral vaccines for mice, sheep, cattle, chickens, and humans. Several engineered attenuated *Salmonella* strains have been tested as live oral vaccines in phase I/phase II clinical trials (Table 7.7). These include strains that have had one to four chromosomal genes deleted. Many of these strains include a deletion in the gene for the stationary-phase sigma factor, *rpoS*, which is important in regulating which proteins are expressed under environmentally stressful conditions such as the presence of acid. Despite the fact that several of these attenuated strains are clinically safe and immunogenic in humans, in all cases, there is an unacceptable shedding of viable vaccine organisms for several weeks following treatment with these strains. Thus, additional modification of these strains is required before they can be used as a vaccine.

Deletion of the *dam* gene, which encodes DNA methylase, may also be an effective approach to produce avirulent *Salmonella* strains. The *dam* gene is a master switch that regulates the expression of 20 to 40 different *Salmonella* regulatory proteins. Mice immunized with Dam-negative strains of *Salmonella* tolerated up to 10,000 times the normally lethal dose of the bacterium. Generally, pathogenic bacteria turn on many of their genes as briefly as possible to avoid detection and attack by the host's immune system. However, with Dam-negative strains, these genes are expressed for much longer periods, making it easier for the host immune system to detect and destroy the invading bacteria. Because many other gut-colonizing bacteria have *dam* genes, if this approach with *Salmonella* turns out to be as effective as hoped, it may be possible to utilize a similar protocol with a range of pathogenic bacteria.

Table 7.7 Genotypes of some genetically engineered attenuated *Salmonella* strains that have been tested as live oral vaccines in phase I/II clinical trials[a]

Δ*rpoS*
Δ*phoP*, Δ*phoQ*
Δ*rpoS*, Δ*phoP*, Δ*phoQ*
Δ*aroC*, Δ*ssaV*
Δ*rpoS*, Δ*aroC*, Δ*ssaV*
Δ*rpoS*, Δ*aroC*, Δ*aroD*
Δ*rpoS*, Δ*aroC*, Δ*aroD*, Δ*htrA*

[a]The genes are described in Table 7.6. The symbol Δ indicates that the gene has been deleted in this particular strain.

Leishmania Species

Leishmaniasis is a parasitic disease that is common in parts of the tropics and subtropics and that may be caused by any one of more than 20 different species of *Leishmania*. The parasites that are responsible for this disease are carried by certain types of sandflies, causing around 2 million new cases and 20 to 30 thousand deaths every year. Although the human immune system can respond to infections by protozoan parasites of the genus *Leishmania*, it has been difficult to develop an effective vaccine against these organisms. Traditional attenuated strains of *Leishmania* are sometimes effective as vaccines; however, they often revert to virulence. Also, the attenuated parasite can persist for long periods in an infected but apparently asymptomatic individual. Such individuals can act as reservoirs for the parasite, which can be transferred to other people by an intermediate host. To overcome these problems, an attenuated strain of *Leishmania* that is unable to revert to virulence was created by targeted deletion of an essential metabolic gene, such as the one encoding dihydrofolate reductase–thymidylate synthase. In one of these attenuated strains, *Leishmania major* E10-5A3, the two dihydrofolate reductase–thymidylate synthase genes that are present in wild-type strains were replaced with the genes encoding resistance to the antibiotics G-418 and hygromycin, which act as selectable markers (Fig. 7.40A). For growth in culture, it is necessary to add thymidine to the medium that is required to propagate the attenuated (but not the wild-type) strain. In addition, unlike the wild type, the attenuated strain is unable to replicate in macrophages in tissue culture unless thymidine is added to the growth medium (Fig. 7.40B). Importantly, the attenuated strain survives for only a few days when inoculated into mice; in that time, it does not cause any disease. Moreover, this period is sufficient to induce substantial immunity against *Leishmania* in BALB/c mice (a widely used inbred strain of mice that is often used in animal experimentation) after administration of the wild-type parasite (Fig. 7.41). Since the attenuated parasite did not establish a persistent infection or cause disease, even in the most susceptible strains of mice tested, it is considered to be a strong candidate vaccine. In mid-2018, researchers developed several additional attenuated *Leishmania* vaccines with targeted deletions in the genes encoding other proteins, such as KHARON1. In this case, the parasite's cell division is impaired, and it is unable to sustain an infection in mice. Following additional experiments with animals, it should be possible to test whether these attenuated parasites are effective as a vaccine in humans.

Vector Vaccines

Vaccines Directed against Viruses

The use of a live vaccinia virus vaccine helped to eradicate smallpox globally. The vaccinia poxvirus consists of a double-stranded DNA genome that contains 187 kilobase pairs (kb) and encodes approximately 200 different proteins. Vaccinia virus DNA replicates within the cytoplasm of infected cells. Cytoplasmic, rather than nuclear, replication and transcription are possible because vaccinia virus DNA contains genes for DNA polymerase, RNA polymerase, and the enzymes to cap, methylate, and polyadenylate messenger RNA (mRNA). Thus, if a foreign gene is inserted into the vaccinia virus genome under the control of a vaccinia virus promoter, it will be expressed independently of host regulatory and enzymatic

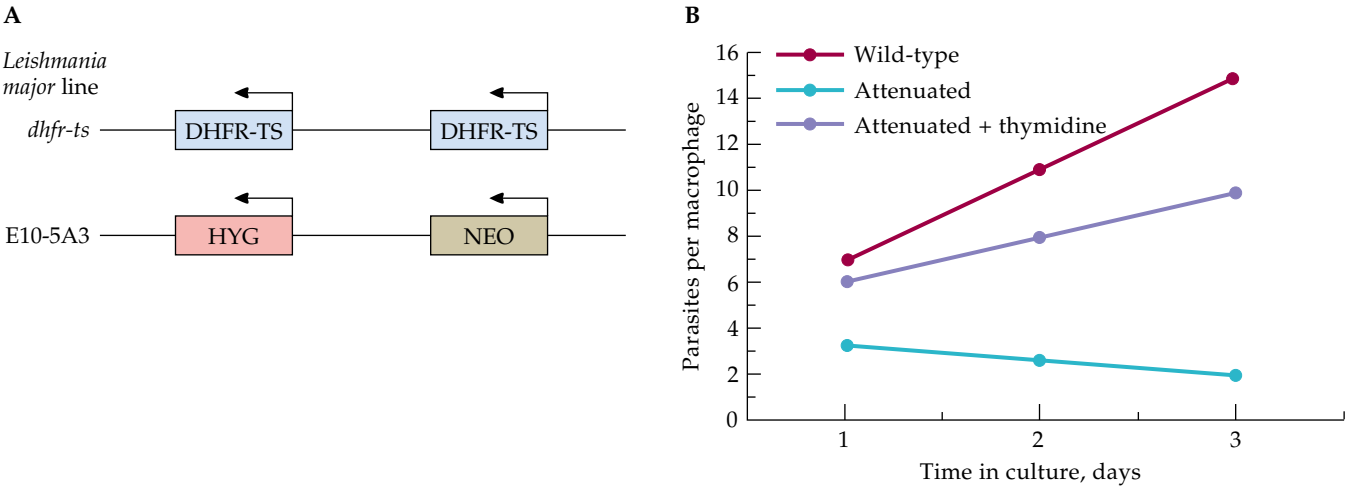
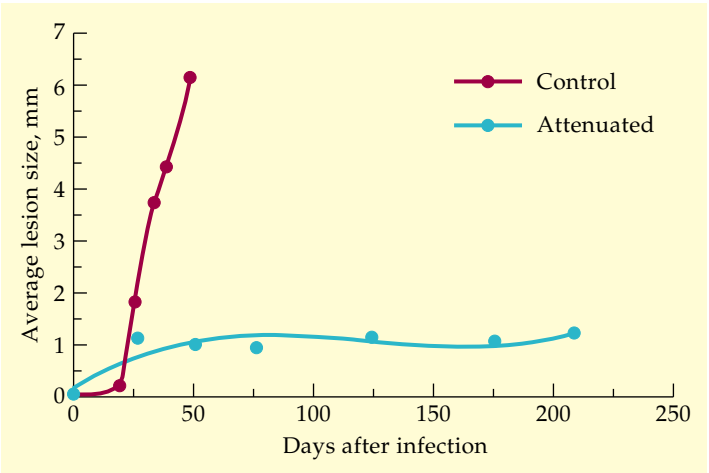


Figure 7.40 **(A)** Schematic representation of the *Leishmania major* DHFR-TS E10-5A3 auxotrophic cell line. The DHFR-TS E10-5A3 line was produced by targeted deletion of two copies of an essential metabolic gene, *dhfr-ts* (which encodes the bifunctional dihydrofolate reductase–thymidylate synthase), in *L. major*. This line does not cause disease when injected into highly susceptible mouse strains, and vaccination confers significant and specific protective immunity. In line E10-5A3, the two *dhfr-ts* genes were replaced by the *hph* (hygromycin phosphotransferase [HYG]) gene, which confers resistance to hygromycin B, and the *aph* (neomycin phosphotransferase [NEO]) gene, which confers resistance to aminoglycosides such as G-418. **(B)** Proliferation of wild-type and attenuated *L. major* in mouse macrophages. At time zero, macrophages were infected with the same amount of stationary-phase *L. major*. The wild-type parasite and the attenuated parasite in the presence of thymidine were able to proliferate, while the attenuated strain did not proliferate in the absence of thymidine in the medium. Data from Titus et al., *Proc. Natl. Acad. Sci. USA* **92:**10267–10271, 1995.

Figure 7.41 Immunity to virulent *L. major* induced in BALB/c mice inoculated with attenuated *L. major*. At time zero, mice that were previously inoculated with attenuated *L. major* were challenged with virulent *L. major*, and the sizes of the parasite-induced lesions were measured at various times. Control mice were not vaccinated with attenuated *L. major*. Data from Titus et al., *Proc. Natl. Acad. Sci. USA* **92:**10267–10271, 1995.



functions. The virus can infect humans and many other vertebrates, as well as invertebrates.

In addition to having a broad host range, vaccinia virus is well characterized at the molecular level, is stable for years after lyophilization (freeze-drying), and is usually nonpathogenic. For these reasons, it is a strong candidate for use as a vector vaccine. A vector vaccine delivers and expresses cloned genes encoding antigens that elicit neutralizing antibodies against pathogenic agents but is not itself a pathogen. However, the very large vaccinia virus genome means that it lacks any unique restriction sites so that it is not possible to insert additional DNA directly into the viral genome.

Of necessity, the genes for specific antigens must be introduced into the viral genome by *in vivo* homologous recombination. This is accomplished as follows: (i) The DNA sequence coding for a specific antigen is inserted into a plasmid vector immediately downstream of a cloned vaccinia virus promoter and in the middle of a nonessential vaccinia virus gene, such as the gene for the enzyme thymidine kinase (Fig. 7.42A). (ii) This plasmid is used to transfect thymidine kinase-negative animal cells in culture, usually in chicken embryo fibroblasts, that have previously been infected with wild-type vaccinia virus, which produces a functional thymidine kinase. (iii) Recombination between DNA sequences that flank the promoter and the neutralizing antigen gene on the plasmid and the homologous sequences on the viral genome results in the incorporation of the cloned gene into the viral DNA (Fig. 7.42B). Although the recombination event is rare, the absence of thymidine kinase activity in the recombined virus enriches for cell lines that carry a recombinant vaccinia virus. (iv) The definitive selection of cells with a recombinant vaccinia virus may be made by PCR.

Since thymidine kinase-negative mutants of vaccinia virus arise spontaneously at a relatively high frequency, a selectable marker is often cotransferred with the target gene. This makes it much easier to distinguish a spontaneous thymidine kinase mutant from a mutant deliberately generated by homologous recombination. In other words, a virus with a spontaneous mutation would not carry the selectable marker, whereas a virus that underwent homologous recombination would. The *neo* gene, which encodes the enzyme neomycin phosphotransferase II and confers resistance to the kanamycin analogue G-418, is often used as the selectable marker. The *neo* gene is quite stable after insertion into the vaccinia virus genome. However, regulatory agencies prefer vaccines to be free of antibiotic resistance genes.

To avoid disrupting any vaccinia virus genes or the necessity of using antibiotic resistance genes as selectable markers, a novel system has been devised in which every recombinant virus that can form a plaque will contain and express the target gene. Wild-type vaccinia virus contains a gene, *vp37*, that is responsible for the formation of plaques when the virus is grown on an animal cell monolayer (Fig. 7.43A). Deleting the *vp37* gene and replacing it with a marker gene (Fig. 7.43B) creates a vaccinia virus mutant that does not form plaques after 2 to 3 days of growth in cell culture. Target genes are introduced into the mutant vaccinia virus by homologous recombination with a transfer vector that carries the *vp37* gene, as well as the target gene (Fig. 7.43C). If homologous recombination between the genome of the non-plaque-forming mutant and the transfer vector occurs, the viruses that can form plaques have acquired the *vp37* gene. Also, the target gene is inserted into the vaccinia virus genome, and

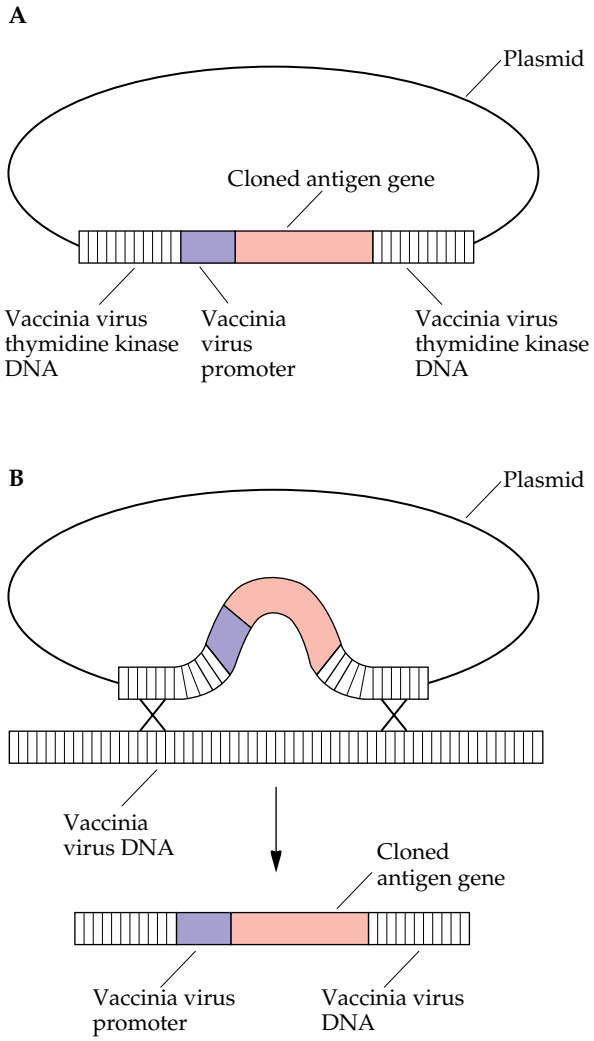


Figure 7.42 Method for the integration into vaccinia virus of a gene whose protein product, generally a viral antigen, elicits an immunological response. **(A)** Plasmid carrying a cloned expressible antigen gene. **(B)** A double-crossover event results in the integration of the antigen gene into vaccinia virus DNA.

the selectable marker gene is lost. Since the *vp37* gene has been deleted in the mutant vaccinia virus, it is impossible for this mutation to revert to the wild type. Therefore, every virus that forms a plaque carries the desired construct. This simple procedure is applicable to the cloning and expression of any target gene, does not require any extra marker genes, and does not result in the disruption of any vaccinia virus genes.

Several antigen genes have been successfully inserted into the vaccinia virus genome and subsequently expressed in animal cells in culture. For example, a recombinant vaccinia virus that expresses the HSV-1 gD (glycoprotein D) gene prevents herpes infections in mice. Another recombinant vaccinia virus that expresses the rabies virus surface antigen gene was able to elicit neutralizing antibodies in foxes, which are major carriers of rabies in Europe. The vaccinia–rabies virus glycoprotein recombinant virus vaccine that has been approved and on the market (as Raboral) since 1987

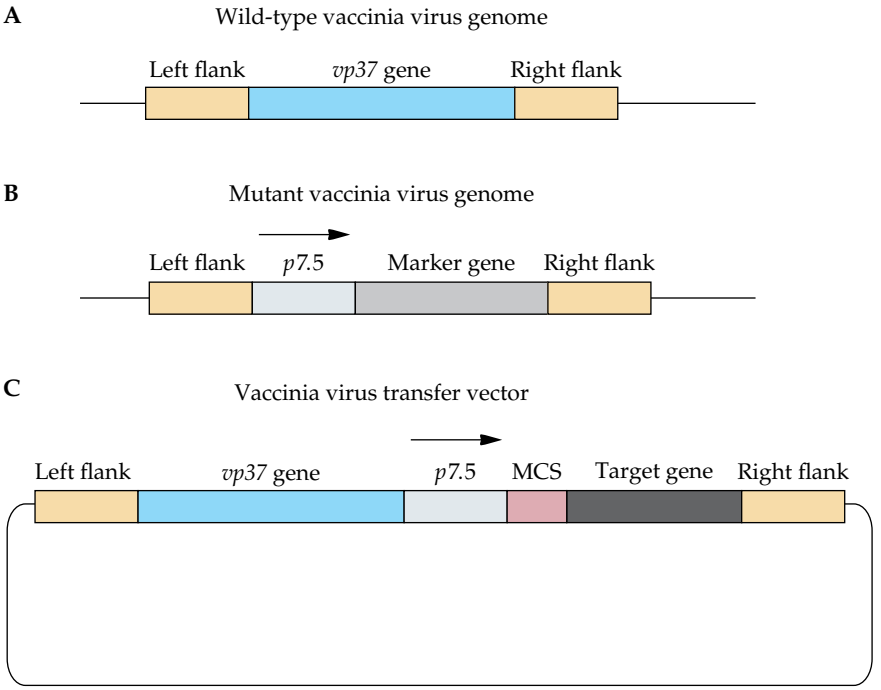


Figure 7.43 **(A)** Portion of a wild-type vaccinia virus genome that contains the *vp37* gene that is responsible for plaque formation in host cells. **(B)** Portion of a mutant vaccinia virus genome in which the *vp37* gene has been replaced by a marker gene. **(C)** Portion of a vaccinia virus transfer vector. "Left flank" and "right flank" refer to the DNA sequences that immediately precede and follow the *vp37* gene in the wild-type vaccinia virus genome. The native *vp37* promoter is part of the *vp37* gene sequence (not shown). MCS is a multiple cloning site with seven unique restriction enzyme sites. *p*7.5 is a strong early/late vaccinia virus promoter. The target gene is inserted into the multiple cloning site. Subsequently, homologous recombination between the transfer vector **(C)** and the genomic DNA of the mutant virus **(B)** results in the replacement of the marker gene with the *vp37* gene, together with a target gene.

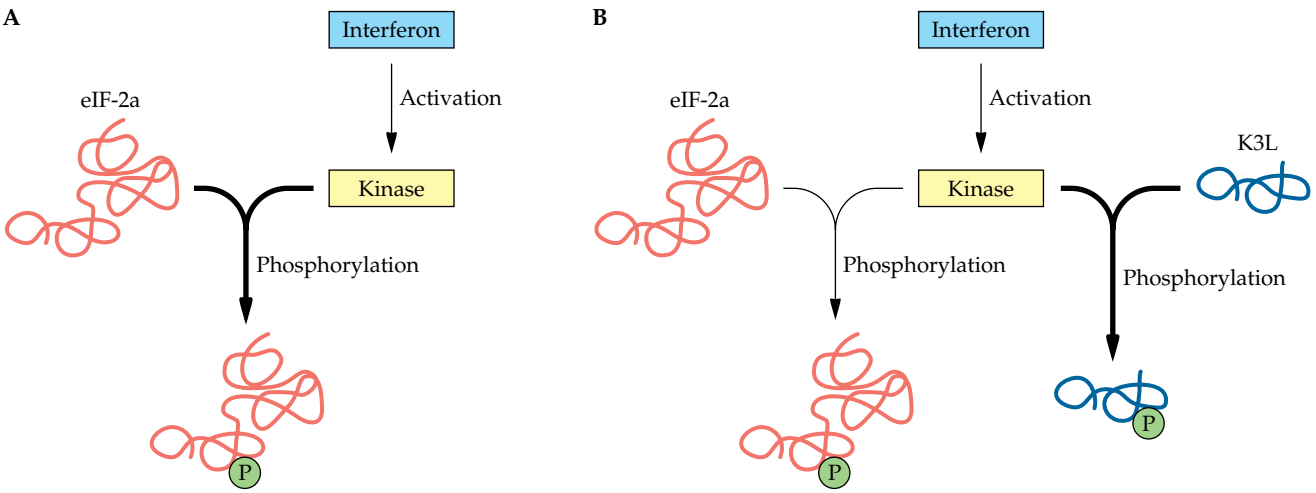
contains 10^8 plaque-forming units (PFU), or live viral particles, per dose. The vaccine is distributed to animals by airplanes or helicopters, or it is placed in bait stations where packages containing the vaccine are coated with a fishmeal attractant. In the United States, this vaccine is only sold to government agencies conducting rabies control programs. The vaccine is constructed by insertion of the gene encoding the surface glycoprotein G of a rabies virus (actually a DNA copy of the gene encoded in the rabies virus RNA genome) into the thymidine kinase gene of a strain of vaccinia virus. Upon ingestion by a fox, the vaccinia virus begins to replicate and express rabies glycoprotein G, which stimulates the development of immune responses to the rabies glycoprotein. This process results in the production of neutralizing antibodies against the rabies virus in immunized foxes and typically lasts about 12 months in cubs and 18 months in adult animals.

The use of vector vaccines constructed from vaccinia virus also offers the possibility of vaccinating individuals against several different diseases with one treatment. This may be achieved by using a recombinant vaccinia virus carrying cloned genes encoding a number of different antigens. When genes encoding several different foreign proteins are inserted into one vaccinia virus, each is placed under the control of a different vaccinia virus

promoter to avoid the possibility of homologous recombination between different portions of the virus genome that might cause the cloned genes to be lost. For example, researchers have created a vaccinia-based vaccine that carries genes encoding both interleukin-15 and the protective antigen from the pathogen *Bacillus anthracis*. This engineered virus provides effective protection against both smallpox (due to a vaccinia virus antigen) and anthrax (due to production of the *B. anthracis* antigen), both potential bioterrorism agents. The inclusion of the gene for interleukin-15 increases the effectiveness and the immunogenicity of the vaccine. To date, this vaccine has only been tested in rabbits.

A live recombinant viral vaccine has several advantages over killed virus or subunit vaccines. First, the virus can express the authentic antigen(s) in a manner that closely resembles a natural infection. Second, the virus can replicate within the host, thereby amplifying the amount of antigen that activates the release of antibodies from B cells (humoral response) and stimulates the production of T cells (cell-mediated immune response). A disadvantage of using a live recombinant viral vaccine is that vaccination of an immunosuppressed host, such as an individual with AIDS, can lead to a serious viral infection. To avoid this problem, it would be helpful to kill or inhibit the vaccinia virus after vaccination. One way to do this would be to insert the gene encoding human IL-2 into the viral vector. IL-2 enhances the response of activated T cells, enabling the recipient to limit the proliferation of the viral vector, and thereby decreases the possibility of an unwanted infection. Alternatively, an interferon-sensitive vaccinia virus (wild-type vaccinia virus is relatively resistant to interferon) whose proliferation is limited may be employed. Such a virus vector would be susceptible to drug intervention if complications from vaccination with vaccinia virus vectors arose.

Figure 7.44 Competitive inhibition of the interferon-stimulated phosphorylation (inhibition) of eIF-2a by protein K3L, which is encoded by vaccinia virus and is nearly identical to a portion of eIF-2a. **(A)** In the presence of interferon, a kinase is activated that phosphorylates eIF-2a molecules and thereby prevents them from functioning. **(B)** When vaccinia virus protein K3L is also present, it is phosphorylated instead of the eIF-2a, so the eIF-2a remains active. The thickness of the arrows represents the relative flux through each pathway.



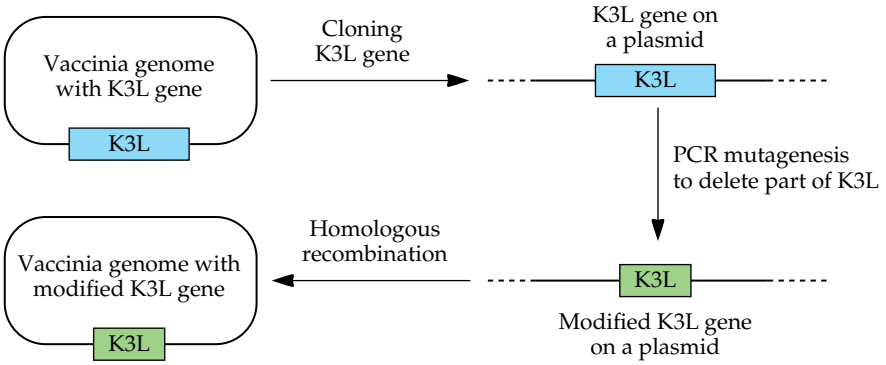


Figure 7.45 Construction of a K3L-negative mutant of vaccinia virus.

The basis of the resistance of vaccinia virus to interferon was not known until a vaccinia virus open reading frame (K3L) was found to encode a 10.5-kDa protein with an amino acid sequence that is very similar to that of a portion of the 36.1-kDa host cell eukaryotic initiation factor 2a (eIF-2a). The N-terminal regions of both of these proteins contain 87 amino acids that are nearly identical. Moreover, this shared sequence contains a serine residue, amino acid 51, which in eIF-2a is normally phosphorylated by interferon-activated P1 kinase. When this serine residue in eIF-2a is phosphorylated in interferon-treated cells, protein synthesis, and therefore viral replication, is inhibited. Vaccinia virus may avoid inhibition by interferon because the K3L protein acts as a competitive inhibitor of eIF-2a phosphorylation (Fig. 7.44). Therefore, deletion of all or a portion of the K3L gene from vaccinia virus should make the virus sensitive to interferon. A K3L-negative mutant of vaccinia virus was constructed by PCR mutagenesis of the K3L gene cloned in a plasmid, followed by homologous recombination to replace the wild-type K3L sequence with the modified version (Fig. 7.45). When the wild-type and mutant versions of vaccinia virus were tested for sensitivity to interferon, the mutant was 10 to 15 times more sensitive to interferon than was the wild-type version (Fig. 7.46). Reinsertion of the wild-type K3L sequence into the mutant virus restored the level of interferon resistance found in the wild type. This indicates that K3L is indeed involved in the interferon resistance phenotype of vaccinia virus. This work is an important step in the development of safer vaccinia virus vectors. Moreover, other interferon-resistant viruses may contain sequences comparable to K3L, and they may also be amenable to the construction of interferon-sensitive deletion mutants. In addition to deleting the K3L sequence, one group of researchers developed a vaccinia virus vector that contained an interferon-γ gene under the transcriptional control of a tetracycline-inducible promoter. In this case, immunodeficient mice were protected from a lethal vaccinia infection when they were treated with tetracycline.

Currently, several veterinary vaccinia virus-based vaccines have been licensed, and clinical studies to test their efficacies in preventing a number of human infectious diseases are under way. This technology is based in part on the development of an attenuated version of the vaccinia virus strain that had previously been used in the eradication of smallpox. To avoid any risk of the vector itself becoming a source of disease, some genetic information was removed from the viral genome so that the viral vector was highly attenuated. This attenuated virus has been used to express a

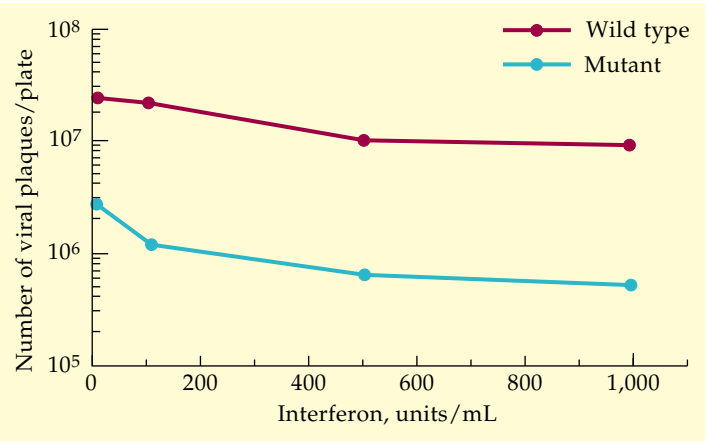


Figure 7.46 Sensitivity of wild-type ($K3L^+$) and mutant ($K3L^-$) vaccinia virus to interferon. Mouse L929 cells were pretreated with a mixture of mouse alpha and beta interferon for 24 hours before the virus was added. Plates without interferon had approximately 3×10^7 viral plaques per plate. Data from Paoletti and Tartaglia, U.S. patent 5,378,457, 1995.

number of viral antigens with the expectation that the recombinant virus would be an effective live vaccine. Protection has been achieved by cloning glycoproteins from porcine pseudorabies virus, hemagglutinin glycoproteins from equine influenza virus, a spike protein from the SARS virus, and a polyprotein of Japanese swine encephalitis virus and then vaccinating humans to prevent transmission of these viruses. Based on the success of these attenuated vaccinia virus vaccines, it has been proposed that this virus be considered a general delivery system for a wide range of proteins.

For mass vaccination campaigns in developing countries, it would be advantageous to be able to deliver live vaccines in a simple, expeditious, and cost-effective manner. In addition, with mucosally transmitted pathogens, such as HIV, traditional vaccination routes may not induce mucosal immune responses sufficient to provide protective immunity. One possible alternative to traditional vaccination is aerosol immunization, which is potentially safer, easier, and less expensive to administer. To this end, researchers tested the abilities of two attenuated vaccinia virus-based vectors to be delivered effectively by aerosol immunization. In fact, it was found that aerosol delivery was both safe and effective, yielding long-lasting systemic and mucosal immune responses when delivered to rhesus macaques (monkeys). This approach still needs to be tested with humans; however, it could offer an effective means of inoculating large numbers of individuals in the future.

Although much of the work on the development of live viral vaccines has been done with vaccinia virus, other viruses, such as adenovirus, poliovirus, and varicella-zoster virus, are also being tested as potential vaccine vectors. Live attenuated poliovirus can be delivered orally, and such a mucosal vaccine, which is directed to receptors in the lungs or gastrointestinal tract, might also be useful against a range of diseases, including cholera, typhoid fever, influenza, pneumonia, mononucleosis, and rabies. However, the safety and efficacy of any apparently benign virus as a gene delivery and expression system must be firmly established before any clinical trials are undertaken.

There is also considerable interest in modified vaccinia Ankara virus (MVA) as a vector system, which consists of a highly attenuated strain of vaccinia virus produced by frequent passage of the virus in chicken cells. MVA has lost about 10% of the vaccinia virus genome so that it cannot replicate efficiently in primate cells and is considered the vaccinia virus strain of choice for clinical investigation because of its high safety profile (in monkeys, mice, swine, sheep, cattle, elephants, and humans). Studies with mice and nonhuman primates have further demonstrated the safety of MVA under conditions of immunosuppression. Recently, scientists developed and tested a recombinant MVA-pandemic influenza A/H5N1 vaccine. In these experiments, mice were immunized with MVA vectors expressing influenza virus proteins, either H5N1-derived nucleoprotein, the stem region of hemagglutinin, matrix proteins 1 and 2, viral polymerase basic protein 1, or the hemagglutinin stem fused to a quadrivalent matrix protein 2 extracellular domain. The immunized mice were subsequently challenged with lethal doses of influenza virus H5N1, H7N1, or H9N2; they were then monitored for disease symptoms. The result of this experiment was that only MVA vectors that expressed nucleoprotein, alone or with other influenza proteins, protected mice from lethal challenge with the above-mentioned viruses. This experiment clearly demonstrates the usefulness of MVA as a live vector vaccine platform.

Dengue Virus

Dengue is one of the most important and widespread mosquito-borne viral diseases of humans, with about half the world's population at risk. The World Health Organization estimates that 50 to 100 million dengue virus infections occur each year in more than 100 countries and that up to half a million people develop severe dengue hemorrhagic fever necessitating hospital admission. A specific treatment is not available, and in the absence of a vaccine, prevention relies on individual protection against mosquitoes and vector control strategies.

The major challenges facing the development of a dengue vaccine include the existence of four separate pathogenic dengue virus serotypes (serotypes 1 to 4) which make it necessary to develop a vaccine that provides immunity against all four serotypes. Additional problems are associated with the lack of suitable animal models in which to test a vaccine. One dengue virus vaccine that was first licenced for use in December 2015 and is now licenced in over 20 countries is CYD-TDV (Dengvaxia), a recombinant, live, attenuated, tetravalent dengue vaccine based on the yellow fever 17D vaccine strain (which is an attenuated virus), and is produced in Vero cells (monkey kidney epithelial cells are used as host cells for growing the virus) (Fig. 7.47). Phase I, II, and III trials were conducted in five countries in southeast Asia and five in Latin America with cohorts of adults and children. Following a three-dose regimen given over 12 months, it was found that the vaccine elicited neutralizing antibody responses against the four serotypes with an overall efficacy of ~60%. Protection was limited to three of the four serotypes of dengue virus; however, the vaccine efficacy with serotype 2 was only ~34%. Nonetheless, the dengue vaccine was well tolerated, with no adverse side effects after several years of follow-up after the first dose.

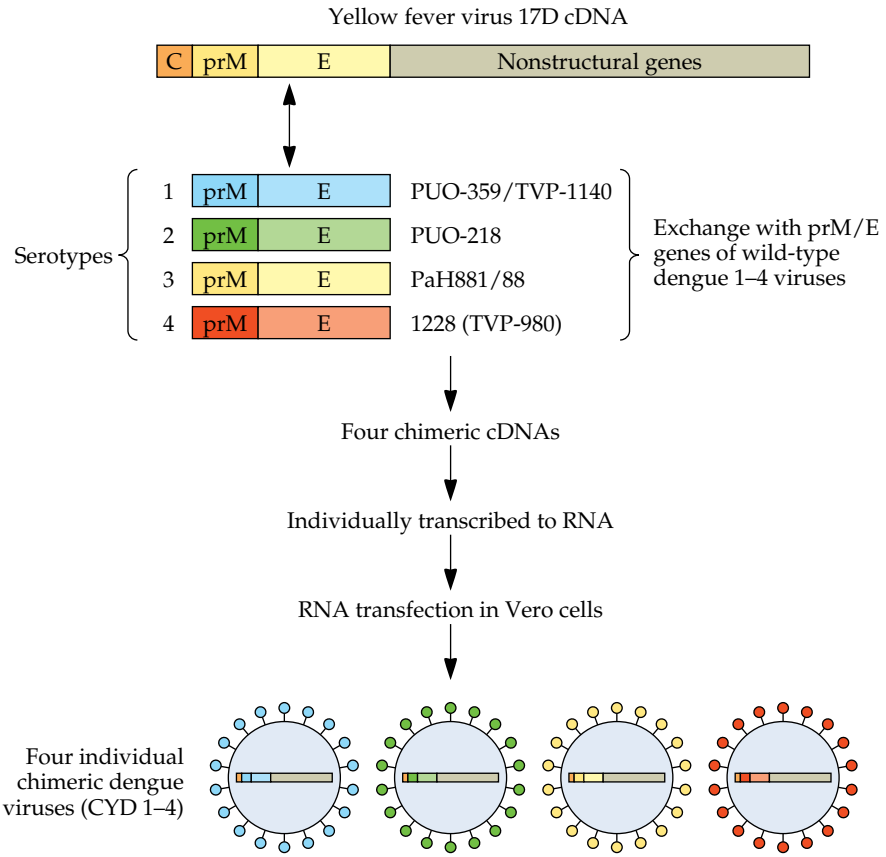


Figure 7.47 Schematic diagram of the CYD-TDV dengue virus vaccine. In 2010, the first phase III clinical trial to investigate the CYD tetravalent dengue vaccine (TDV) was initiated. The CYD-TDV candidate is composed of four recombinant, live, attenuated vaccines (CYD 1 to 4) based on a yellow fever virus vaccine 17D (YFV 17D) backbone, each expressing the premembrane (prM) and envelope (E) genes of one of the four wild-type dengue virus serotypes. The letter C denotes the gene encoding the capsid protein. These wild-type viruses were the PUO-359/TVP-1140 Thai strain for serotype 1, PUO-218 Thai strain for serotype 2, PaH881/88 Thai strain for serotype 3, and 1228 (TVP-980) Indonesian strain for serotype 4. The CYD-TDV vaccine is produced by combination of the four CYD viruses into a single preparation. Preclinical studies have demonstrated that CYD-TDV induces controlled stimulation of human dendritic cells and significant immune responses in monkeys. Adapted from Guy et al., *Vaccine* **29:**7229–7241, © 2011, with permission from Elsevier.

Ebola Virus

An Ebola outbreak in West Africa (Sudan, Democratic Republic of the Congo, Sierra Leone, Guinea, and Liberia) began in 2013, and in less than two years more than 27,000 cases and 11,000 deaths were reported. Another more recent outbreak of Ebola occurred in the Democratic Republic of the Congo beginning in August of 2018 and officially ending in June 2020, infecting ~4,000 people and resulting in ~2,500 deaths. Ebola is caused by the Ebola virus, and although it is not highly contagious, it is quite lethal. Once the Ebola virus has crossed the physical barrier of the skin, its initial targets are the dendritic cells (a type of antigen-presenting cell)

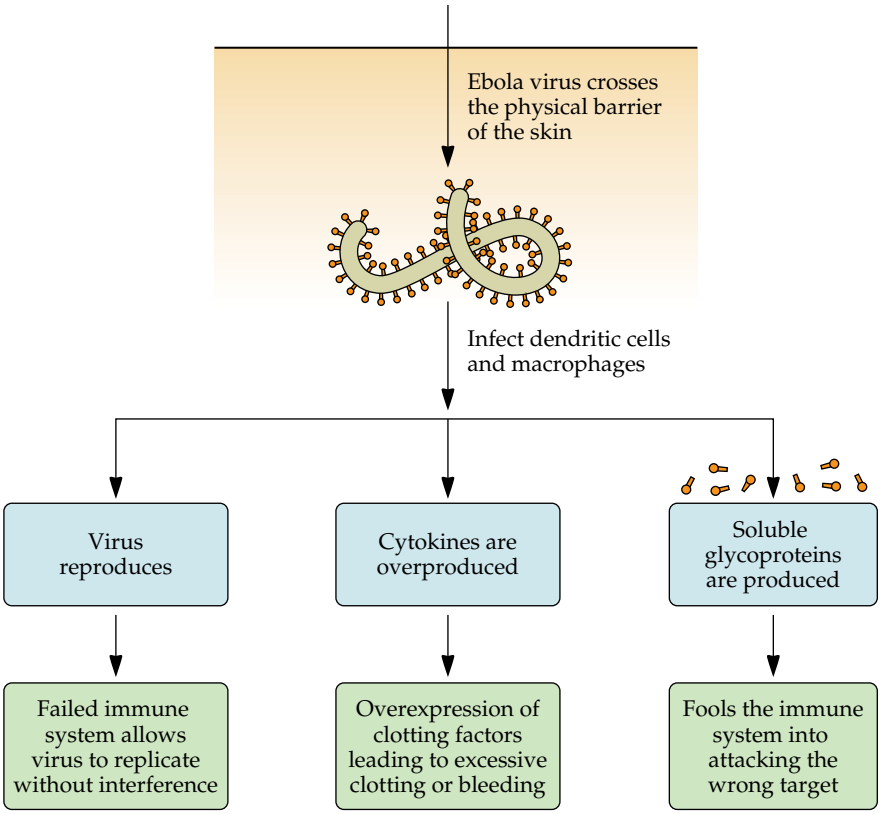


Figure 7.48 Flow chart depicting how the Ebola virus causes disease.

and macrophages (large phagocytic cells) (Fig. 7.48). Rather than trying to avoid these immune system responders, Ebola viruses seek them out and begin reproducing inside them. The attack of these front-line immune cells allows the viruses to disrupt the cells' normal ability to activate the rest of the immune system and also facilitates the rapid spread of the virus throughout the body as the virus travels inside these cells to the lymph nodes, liver, spleen, and other areas of the body. The virus programs the cells that it infects to manufacture and release large amounts of soluble glycoprotein, which is very similar to the glycoprotein molecule that projects from the viruses' outer surface. The soluble glycoprotein is a decoy to avoid the immune system, which would normally target the glycoprotein on the virus' surface and subsequently mount an effective defense against the infection.

Considerable effort has been directed toward developing an Ebola vaccine. One promising Ebola vaccine includes a recombinant vesicular stomatitis virus expressing the Ebola virus glycoprotein (Fig. 7.49). Vesicular stomatitis virus is a nonsegmented, negative-stranded RNA virus that is primarily an animal pathogen that does not cause severe disease in humans. This live-attenuated recombinant vaccine was found to be safe in phase I clinical trials. Moreover, in preclinical trials, this vaccine completely protected macaques against lethal challenge with one strain of Ebola virus. The mechanism of protection against the Ebola virus is the development of antibodies directed against the Ebola virus glycoprotein encoded within the vesicular stomatitis virus carrier. As a consequence of the urgency to rapidly develop an effective Ebola vaccine, in March 2015,

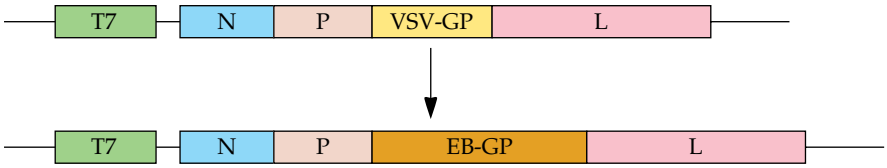


Figure 7.49 Schematic representation of the development of a vesicular stomatitis virus (VSV)-based vaccine against Ebola. The VSV genome was cloned into a plasmid under the control of the bacteriophage T7 polymerase promoter (T7); then, the VSV gene for its own glycoprotein (VSV-GP) was excised and replaced by the Ebola virus glycoprotein gene (EB-GP). The VSV glycoprotein is one of this virus' major virulence factors so that its deletion further attenuates this vector. Other VSV proteins include N, nucleoprotein; P, phosphoprotein; and L, polymerase.

phase II and phase III clinical trials started in Guinea and the Democratic Republic of the Congo at the same time. Approximately 300,000 people in West Africa were vaccinated against Ebola; 80% did not contract the disease, while the other 20% developed only mild cases. This vaccine was approved by the European Medicines Agency in November 2019 and by the U.S. FDA in December 2019. Health authorities are optimistic/hopeful that this vaccine will provide a safe and efficacious means of preventing any future Ebola epidemics. In addition, seven other vaccines against Ebola are in development.

Vaccines Directed against Bacteria

Since the discovery and subsequent widespread dissemination of antibiotics, only a modest amount of research has been directed toward the development of vaccines for bacterial diseases. However, there are good reasons for developing bacterial vaccines. First, not all bacterial diseases are readily treated with antibiotics. Second, the use of antibiotics over the last 40 years has resulted in the proliferation of bacterial strains that are resistant to numerous antibiotics. Third, reliable refrigeration facilities for the storage of antibiotics are not commonly available in many tropical countries. Fourth, it is often difficult to ensure that individuals receiving antibiotic therapy undergo the full course of treatment.

Given the need to produce vaccines that will be effective against bacterial diseases, an important question is, which strategies are likely to be most effective? In instances where the disease-causing bacterium does not grow well in culture, the development of an attenuated strain is not feasible. For these bacteria, alternative approaches must be used. For example, *Rickettsia rickettsii*, a Gram-negative obligately intracellular bacterium that causes Rocky Mountain spotted fever and is carried by ticks, does not grow in culture. In this case, a cloned 155-kDa protein that is a major surface antigen of *R. rickettsii* was used as a subunit vaccine and protected immunized mice against infection by this disease-causing bacterium. However, as recently as 2019, no vaccines had been approved to help to prevent this disease.

Tuberculosis

Today, along with HIV and malaria, tuberculosis remains one of the "big three" infectious diseases globally. This disease is caused by the bacterium *Mycobacterium tuberculosis*, which can form lesions that lead to cell death in any tissue or organ. The lungs are most commonly affected. Patients suffer fever and loss of body weight, and without treatment, tuberculosis is often

fatal. It is estimated that approximately 10 to 12 million people worldwide currently have active cases of tuberculosis and that approximately 1.5 to 2 million deaths a year result from these infections. Over the past 50 years, antibiotics have been used to treat patients infected with *M. tuberculosis*. However, numerous multidrug-resistant strains of *M. tuberculosis* are now prevalent. In the United States, among HIV patients infected with an antibiotic-resistant strain of *M. tuberculosis,* there is a 50% mortality rate within 60 days. Consequently, a bacterial disease that was thought to be under control has become a serious public health problem in many parts of the world.

Currently, in some countries, bacillus Calmette-Guérin (BCG), a traditionally attenuated strain of *Mycobacterium bovis* (the cause of tuberculosis in cattle) that was developed between 1906 and 1919, is used as a vaccine against tuberculosis. However, the use of this vaccine has some drawbacks. While the BCG vaccine can prevent disseminated disease and mortality in newborns and children, it cannot prevent chronic infection or protect against pulmonary tuberculosis in adults. Consequently, *M. tuberculosis* establishes a latent chronic infection that reactivates when there are diminished immune responses, e.g., in aged people, in individuals with genetic immune defects, and in those whose medication reduces their immune responses, such as patients treated with antibodies against tumor necrosis factor alpha. Immunosuppression caused by HIV is also an important factor in the reactivation of tuberculosis, and in many people coinfected by HIV and *M. tuberculosis*, it is the major cause of mortality in this population. About 2 billion people worldwide carry a latent *M. tuberculosis* infection, and approximately 10% progress to active disease at some time. Additional limitations to the use of BCG as a vaccine are that individuals immunized with BCG respond positively to a common tuberculosis diagnostic test, which makes it impossible to distinguish between individuals infected with *M. tuberculosis* and those inoculated with BCG cells. In addition, various strains of BCG have undergone significant genetic change over the years so that not all strains of BCG are equally effective against tuberculosis for all populations. For these reasons, the BCG strain is not approved for use in a number of countries, including the United States.

To determine whether a safer and more effective vaccine against tuberculosis might be developed, the extent of the immunoprotection elicited by purified *M. tuberculosis* extracellular proteins was examined. Following growth of the bacterium in liquid culture, 6 of the most abundant of the approximately 100 secreted proteins (Fig. 7.50) were purified. Each of these proteins was used separately and then in combination to immunize guinea pigs. The immunized animals were then challenged with an aerosol containing approximately 200 cells of live *M. tuberculosis*—a large dose for these animals. The animals were observed for 9 to 10 weeks before their lungs and spleens were examined for the presence of disease-causing organisms. In these experiments, some of the purified protein combinations provided a slightly lower level of protection against weight loss, death, and infection of lungs and spleen than did the live BCG vaccine. A major protein that provided protection was the *M. tuberculosis* major secretory protein, a 30-kDa mycolyltransferase, commonly known as α-antigen. However, a DNA vaccine encoding this protein was less effective than the purified secreted protein.

While this and possibly other *M. tuberculosis*-secreted proteins might eventually be part of a safe and efficacious vaccine for the prevention of

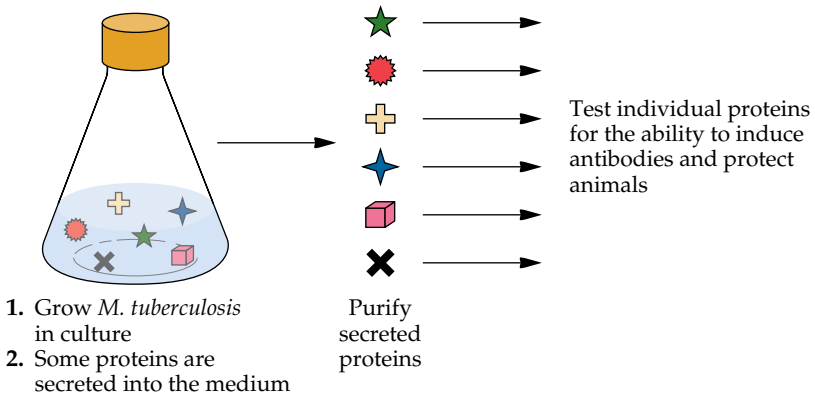


Figure 7.50 Schematic representation of the development of a multiprotein subunit vaccine for tuberculosis. The six most abundant secreted proteins from *M. tuberculosis* are purified from the growth medium and then tested for the ability to induce antibodies in guinea pigs. The immunized animals are subsequently challenged with *M. tuberculosis*.

tuberculosis in humans, it is necessary to develop a suitable delivery system for them. In theory, the optimal delivery system for an antigen that provides protection against tuberculosis should be (i) able to multiply in the mammalian host, (ii) nonpathogenic, and (iii) able to express and secrete the protective antigen. All of these requirements are satisfied by the available BCG strain. Therefore, an *E. coli*–mycobacterium shuttle vector that contained the gene for the 30-kDa protein (α-antigen) under the control of its own promoter was introduced into two different BCG strains (Fig. 7.51A). Transformed cells produced 2.0- to 5.4-fold more 30-kDa protein than did nontransformed cells. In addition, despite the fact that the introduced genes were plasmid encoded and therefore potentially unstable, transformed cells continued to express a high level of 30-kDa protein after the vaccination of a test animal. In agreement with the hypothesis that the extracellular proteins of intracellular organisms are key immunoprotective molecules, guinea pigs immunized with the recombinant BCG strains had significantly fewer bacilli in their lungs and spleens. In addition, there were smaller and fewer lesions in their lungs, spleens, and livers, and the survival of the animals was significantly increased compared with animals vaccinated with a nontransformed BCG strain (Fig. 7.51B). This is the first report of a vaccine against tuberculosis that is more potent than the currently available commercial vaccine.

There are currently a number different vaccines against tuberculosis in clinical trials. Several include the α-antigen. One is a recombinant BCG strain engineered to overexpress the BCG α-antigen to make it more immunogenic. Clinical trials of this recombinant BCG were found to induce better $CD4^+$ T-cell responses against α-antigen than wild-type BCG. Another vaccine in clinical trials consists of modified vaccinia Ankara (as a live carrier) that produces BCG α-antigen; this is administered together with wild-type BCG.

Unfortunately, the α-antigen is primarily expressed during the initial stage of infection, when *M. tuberculosis* divides rapidly and consequently requires the synthesis and assembly of cell wall components. As the bacterial growth rate decreases, there is a decreased need for mycolyl synthesis, a process in which α-antigen is involved and the expression of α-antigen is

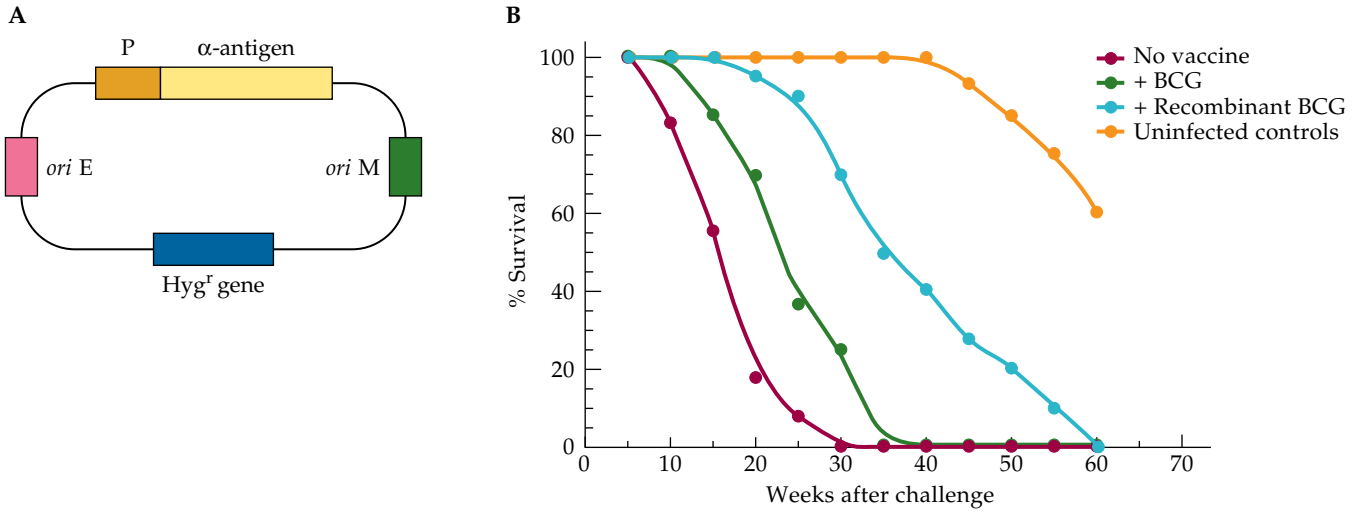
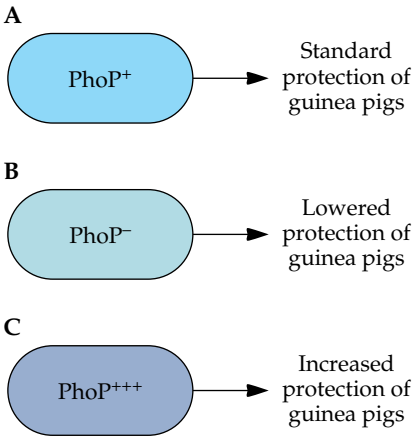


Figure 7.51 **(A)** The plasmid construct used to transform BCG to make it a more effective vaccine (recombinant BCG). After construction of the plasmid, it was isolated from *E. coli* cells and then introduced into BCG by electroporation. *ori* E, *E. coli* origin of replication; *ori* M, *Mycobacterium* origin of replication; Hyg^r gene, hygromycin resistance gene (and its promoter) selectable marker; P and α-antigen, the promoter and the coding region of the 30-kDa secreted protein. **(B)** Survival of guinea pigs infected (challenged) with a pathogenic strain of *M. tuberculosis*. BCG is the traditional live bacterial vaccine strain.

reduced to a low level. Thus, while antibodies directed against this antigen are effective during the initial stages of infection, they are much less effective in chronically infected individuals. It is necessary to identify vaccine antigens that are constitutively expressed. In fact, researchers have very recently reported positive results with a subunit vaccine consisting of a fusion protein that included three separate *M. tuberculosis* secreted proteins. A subunit vaccine with purified fusion protein was reported to provide protection from tuberculosis in mice that was equivalent to the protection provided by the BCG vaccine strain. It will be interesting to see whether the expression of this fusion protein in a recombinant BCG vaccine will provide added protection against *M. tuberculosis* in mice and whether the vaccine will be effective in humans.

One group of researchers observed that a strain of BCG that had no PhoP activity did not effectively protect guinea pigs against infection by *M. tuberculosis* (Fig. 7.52). PhoP is a response regulator of the PhoP-PhoR two-component system that positively regulates the expression of more than 40 BCG genes, including α-antigen. Moreover, strains of BCG that were engineered to overproduce PhoP better protected guinea pigs against challenge with BCG. This result suggests a possible strategy to engineer vaccine strains of BCG to increase its efficacy as a tuberculosis vaccine in humans. In addition, several other laboratories have reported similar positive, albeit preliminary, results in genetically manipulating BCG to be a more effective vaccine.

Figure 7.52 Schematic representation of the effectiveness of three different genetic variants of BCG as vaccines to protect guinea pigs against *M. tuberculosis* infection. **(A)** Wild-type BCG that produces a normal amount of protein PhoP; **(B)** a mutant BCG strain that does not synthesize PhoP; **(C)** a genetically engineered variant of BCG that overproduces PhoP.



Bacteria as Antigen Delivery Systems

One strategy for the development of highly effective vaccines involves the use of attenuated strains of intestinal bacteria such as *Salmonella* spp.

to create live vectors that synthesize protective antigens from unrelated pathogens. The attenuated *Salmonella* then delivers these antigens to the immune system, eliciting protective immune responses. Advantages of using *Salmonella* include the fact that these vaccines (i) can be delivered orally, (ii) target M cells (specialized epithelial cells in the gut that transport antigens to underlying immune system cells), resulting in the induction of an immune response, (iii) are internalized by dendritic cells and macrophages which then present the antigen to lymphocytes, and (iv) stimulate broad immune responses. Macrophages help to initiate immune defense mechanisms by recruiting other immune cells such as lymphocytes. Using *S. enterica* serovar Typhimurium as the live vector and mice as the model, protective vaccines to many foreign antigens of other bacteria, viruses, protozoa, and helminths (parasitic worms) have been generated. Interestingly, the expression of foreign antigens may overattenuate the *Salmonella* strain and compromise the ability of the vector to persist in the animal host and generate foreign antigen-specific immunity. This phenomenon may be a consequence of the metabolic load that is caused by the expression of the foreign antigen. To overcome this problem, researchers are developing *Salmonella* spp. strains that are less attenuated so that sufficient levels of foreign antigens may be expressed (in order to induce a strong immune response against the foreign antigen) while avoiding overattenuation (and thereby debilitation) of the strain.

Cholera

Antigens that are located on the outer surface of a bacterial cell are generally more immunogenic than those in the cytoplasm. Thus, localization of a neutralizing antigen from a pathogenic bacterium on the surface of a live nonpathogenic bacterium is expected to increase the immunogenicity of the antigen. Flagella consist of filaments of a single protein called flagellin, which microscopically appear as conspicuous threadlike structures on the outer surfaces of some bacteria. If the flagella of a nonpathogenic organism could be configured to carry a specific epitope from a pathogenic bacterium, protective immunogenicity might be easily achieved.

This strategy was used to engineer a cholera vaccine (Fig. 7.53). A synthetic oligonucleotide specifying an epitope of the cholera toxin B subunit was inserted into a portion of the *Salmonella* flagellin gene that varies considerably from one strain to another (hypervariable segment). The construct was then introduced into a flagellin-negative strain of *Salmonella*. The epitope, which consisted of amino acid residues 50 to 64 of the cholera toxin B subunit, elicits antibodies directed against intact cholera toxin. The chimeric flagellin functioned normally. Furthermore, the epitope was present at the flagellum surface. Immunization of mice by intraperitoneal injections of approximately 5×10^6 live or formalin-killed "flagellum-engineered" bacteria elicited high levels of antibodies directed against both the peptide, i.e., amino acids 50 to 64, and the intact cholera toxin B subunit. Two or three different epitopes can be inserted into a single *Salmonella* flagellin gene, thereby creating a multivalent bacterial vaccine.

Tetanus

Attenuated *Salmonella* strains can be administered orally, which would enable them to deliver a range of bacterial, viral, and parasite antigens to the mucosal immune system. For this purpose, the choice of the promoter that drives the transcription of the foreign antigen is important. If a strong

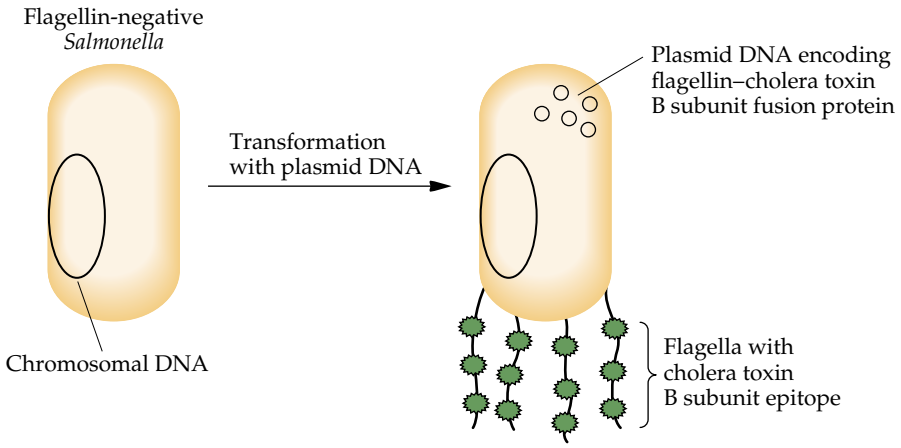


Figure 7.53 Using *Salmonella* as an antigen delivery system and a flagellin–antigen fusion protein for presenting the antigen to the host immune system. A flagellin-negative strain of *Salmonella* was transformed with a plasmid containing a synthetic oligonucleotide specifying an epitope of the cholera toxin B subunit inserted into a hypervariable region of a *Salmonella* flagellin gene.

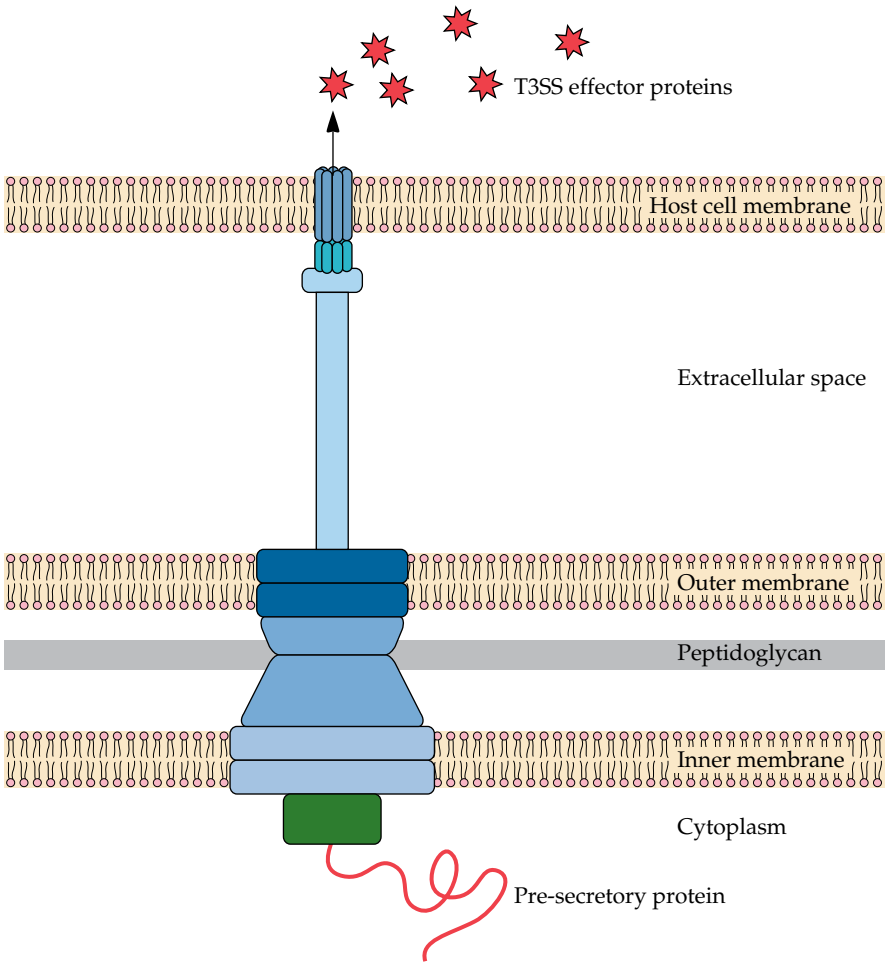
promoter is used, the metabolic load of antigen overexpression might constrain bacterial proliferation. Moreover, promoters that are activated by shifting the temperature or adding specific metabolites to induce foreign-gene expression cannot be activated when the bacterial vector is added to a host animal, as this would be a problem for the animal. On the other hand, promoters that respond to environmental signals may provide effective means of controlling the expression of the foreign antigen gene. For example, the *E. coli nirB* promoter, which is regulated by both nitrite and the oxygen tension of the environment, becomes active under anoxic conditions that are found in the animal gut. In one series of experiments, the *nirB* promoter was used to direct the expression of an immunogenic peptide derived from the *Clostridium tetani* toxin (tetanus toxin) in an attenuated strain of *Salmonella*. It is estimated that more than 1 million deaths per year in the developing world are the result of *C. tetani* infections. When the engineered *Salmonella* strain was grown aerobically in culture, the tetanus toxin peptide was not synthesized. However, following oral administration of the bacterium to test mice, the peptide was produced, and the animals generated antibodies against it. Thus, the engineered *Salmonella* strain has potential as a live oral tetanus vaccine.

Staphylococcus aureus

The Gram-positive bacterium *Staphylococcus aureus* is commonly a part of the human microbiota, typically found in the upper respiratory tract and on the skin. This facultative anaerobe can become an opportunistic pathogen causing skin diseases, respiratory infections, and food poisoning. Many strains of *S. aureus* are antibiotic resistant, and these strains are estimated to be responsible for ~50,000 deaths per year in the United States. Several vaccines were developed in an effort to immunize people against *S. aureus* infections; however, all of them ended in failure in clinical trials. Knowing that this bacterial pathogen was becoming a major cause of hospital-acquired infections, researchers tried a novel approach

of developing an effective vaccine against *S. aureus*. The genes coding for two *S. aureus* virulence factors (SaEsxA and SaEsxB) were separately fused with a *Salmonella* protein so that the fusion proteins would be secreted by the *Salmonella* type III secretion system (Fig. 7.54). When mice were orally treated (on day 1, day 8, and day 22) with either of the engineered *Salmonella* strains, the bacteria were taken up by mouse macrophage cells which presented the fusion protein, including the *S. aureus* virulence factor. Antigen-specific immune responses in the treated mice were examined approximately 1 week after a second vaccine booster, and the effectiveness of the vaccine was evaluated by a lethal challenge with two different strains of *S. aureus* clinical isolates. The results of this test indicated that both vaccines induced an expected immune response in treated mice and also increased their survival when they were challenged with the *S. aureus* pathogen strains. Unfortunately, the protection conferred by these vaccines was not complete, and it was suggested that a multivalent vaccine that

Figure 7.54 Schematic representation of the type III secretion system (T3SS) found on the surface of some Gram-negative bacteria (mostly pathogens). Bacterial proteins with specific secretion signal sequences are secreted through the needle-like projection into the cytoplasm of eukaryotic cells. This system helps the pathogen escape the eukaryotic immune system.



included both SaEsxA and SaEsxB as well as other antigens might be more effective. Also, the vaccine efficacy might be improved by increasing the level of expression of the introduced antigens.

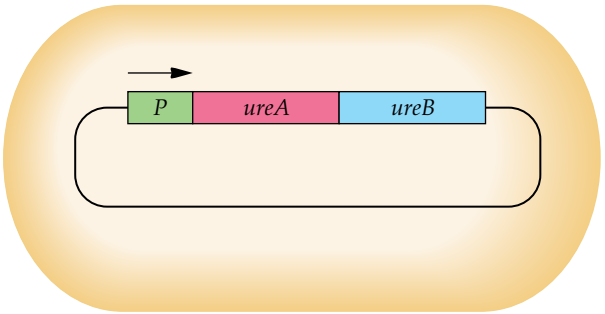
Helicobacter pylori

The Gram-negative bacterium *Helicobacter pylori* is a gastrointestinal, microaerophilic (requires a less-than-atmospheric level of oxygen to survive) organism that is widely distributed among human populations. It is the causative agent of a number of gastrointestinal diseases, including chronic gastritis, peptic ulcers, gastric lymphoma, and gastric cancer. Among infected individuals, which includes more than half of the world's population, about 10% are at risk of developing peptic ulcers. In recent years, the medical treatment for peptic ulcers has changed from antacids to two different antibiotics and a proton pump inhibitor. The antibiotics eradicate the *H. pylori* infection, while the proton-pump inhibitor blocks the enzyme hydrogen–potassium ATPase, preventing the production of acid from the parietal cells at the gastric mucosa, which facilitates the healing of the mucosa.

Unfortunately, *H. pylori* is resistant to a number of commonly used antibiotics, including metronidazole, amoxicillin, erythromycin, and clarithromycin. Treatment of *H. pylori* requires multidrug regimens because the organism resides in a layer of mucus that acts as a barrier to antibiotic penetration. In addition, the necessary course of antibiotic treatment is too expensive for the populations of many less developed countries.

Colonization of the gastrointestinal tract by *H. pylori* is facilitated by the action of an *H. pylori*-encoded urease. This enzyme hydrolyzes urea to carbon dioxide and ammonia, thereby neutralizing stomach acid, making it possible for the bacterium to survive, bind, and function in the host. Urease is a cytosolic and surface-exposed nickel metalloenzyme and is one of the most abundant proteins expressed in *H. pylori*. The enzyme comprises two subunits, A and B, that assemble into a complex [(αβ)3]4 supramolecular structure. Subunit B is more antigenic, making it a possible vaccine candidate. To develop a vaccine that protects individuals against *H. pylori* infections, the genes encoding *H. pylori* urease subunits A and B were constitutively expressed under the control of a *Salmonella* promoter in a genetically deleted (attenuated) strain of *S. enterica* serovar Typhi (Fig. 7.55). Neither immunization with urease-expressing *S. enterica*

Figure 7.55 Schematic representation of an attenuated strain of *S. enterica* serovar Typhi transformed with a plasmid encoding *H. pylori* urease subunits A and B under the transcriptional control of a *Salmonella* promoter (*P*). The arrow indicates the direction of transcription.



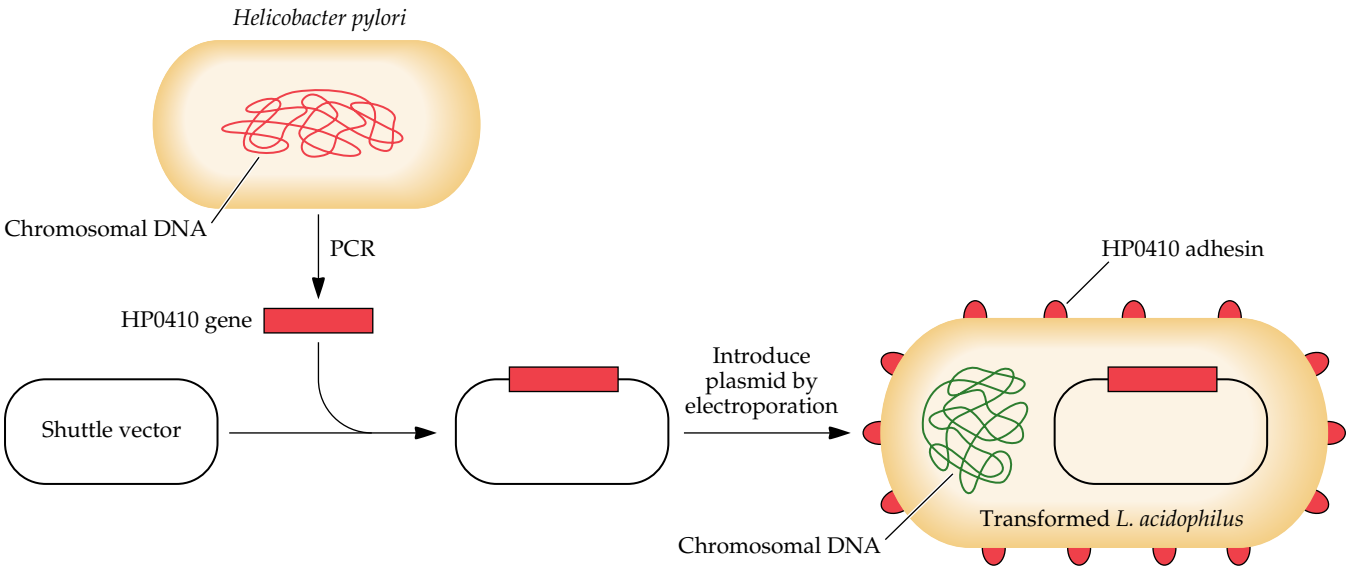
serovar Typhi alone nor immunization with the purified urease enzyme plus an adjuvant conferred protection against challenge with a mouse-adapted strain of *H. pylori*. On the other hand, a vaccination protocol that combined both urease-expressing *S. enterica* serovar Typhi and urease plus an adjuvant was protective. While the success of this approach remains to be established in humans, these initial results are nevertheless encouraging and give hope that a human vaccine against *H. pylori* will be developed in the near future.

Several researchers have attempted to develop live bacterial vaccines that express other key *H. pylori* proteins. For example, one group of researchers PCR-amplified the *H. pylori* gene encoding the HP0410 adhesion protein, inserted this gene onto an *E. coli-Lactobacillus acidophilus* plasmid shuttle vector, and then introduced the plasmid into *L. acidophilus* (Fig. 7.56). The transformed *L. acidophilus* expressed the adhesion protein on its surface and was used as a live vaccine to protect mice from *H. pylori*. The results of this experiment indicated that IgA antibodies were elicited in the mucosal tissue of the mice that were immunized with the recombinant bacteria. While these results are still preliminary, recombinant *L. acidophilus* expressing HP0410 is considered to be a promising vaccine candidate against *H. pylori*.

Monoclonal Antibody Passive Immunity

It has been known for many years that individuals can be protected from a variety of infectious diseases by passive immunity, the transfer of immunity from one individual or animal to another. Passive immunization occurs naturally and most commonly when maternal antibodies are

Figure 7.56 Development of a recombinant *L. acidophilus* vaccine against *H. pylori*. The *H. pylori* gene encoding the HP0410 adhesion protein was amplified by the polymerase chain reaction, cloned onto an *L. acidophilus–E. coli* shuttle vector, analyzed in *E. coli*, and then introduced into *L. acidophilus* by electroporation, where the HP0410 protein is expressed on the bacterial surface.



transferred to a fetus through the placenta. Passive immunization can also be artificial. For example, around 1890, researchers began inoculating animals with the bacterium *Corynebacterium diphtheriae,* which causes diphtheria in humans. The resulting crude antiserum containing antibodies against *C. diphtheriae* (then called diphtheria antitoxin) was used to treat this often-fatal childhood disease. In those days, mortality sometimes reached 45%. *C. diphtheriae* infects the throat or tonsils and produces an exotoxin that is lethal to human cells. Since an individual's immune system often makes antibodies against the administered foreign antibodies, artificial passive immunization can usually be administered only once. However, the use of human antiserum is less problematic, and the FDA has approved several products for passive immunization for diseases for which there is no approved vaccine. Unfortunately, the production of human antibodies in a laboratory is expensive and difficult to do and also requires a large number of human blood donors. Nevertheless, it is currently possible to produce humanized or human monoclonal antibodies on a large scale against a variety of pathogenic agents (see chapter 3). Moreover, in addition to treating infected individuals, these antibodies may be used to prevent various infections.

Influenza Virus

During the past 150 years, six separate human influenza pandemics have been directly responsible for well over 100 million deaths worldwide. Each pandemic virus was derived, in part, from an avian influenza virus by reassortment (mixing) of genetic material from avian and human viruses. In each case, an avian hemagglutinin envelope glycoprotein was acquired that was antigenically distinct from the hemagglutinins of the human viruses in circulation at that time.

The viral replication cycle cannot initiate without the binding of hemagglutinin to the sialic acid residues that are found on the surface of host target cells. Thus, hemagglutinin influences the host range and pathogenicity of influenza. Hemagglutinin is the primary target of neutralizing antibodies against influenza virus; however, this protein rapidly and continuously accumulates mutations that allow it to escape recognition by the immune system (Fig. 7.32A). In addition, although hemagglutinin from one strain of influenza induces strong neutralizing antibodies against that viral strain, it is rarely effective against other strains of influenza. In an effort to overcome this shortcoming, researchers selected monoclonal antibodies against conserved regions of hemagglutinin. Hemagglutinin consists of two major domains: a stem domain that does not change very much and a globular head domain that often undergoes structural changes. Until recently, all of the strongly neutralizing antibodies were directed against the head region. However, one of the monoclonal antibodies recognized a highly conserved region on the stem portion of the hemagglutinin molecule near the region where it interacts with the viral envelope membrane (Fig. 7.32B). When tested in cultured cells, this monoclonal antibody prevented the fusion between viral and host cell membranes that is required for entry of the virus into the host cell. Moreover, this monoclonal antibody interacted with a broad range of influenza virus strains and subtypes. It is thought that this and similar humanized or human monoclonal antibodies may eventually provide effective protection against a wide range of influenza viruses.

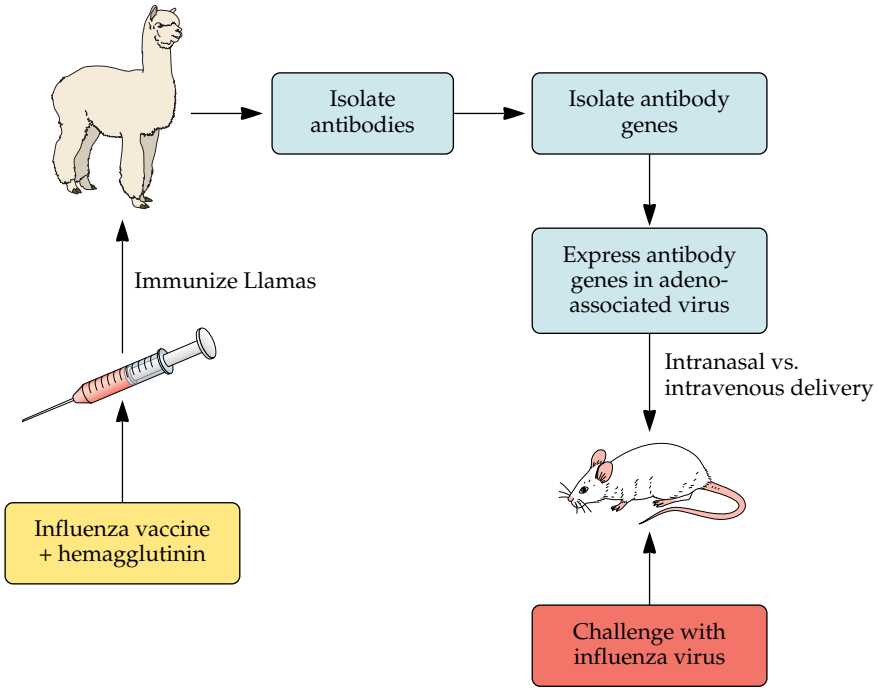
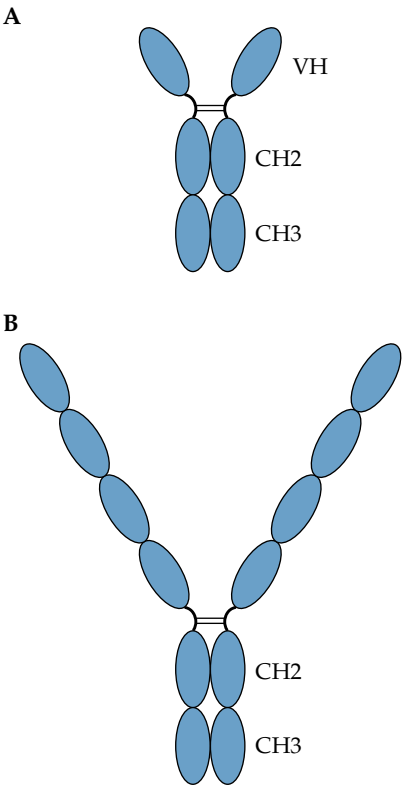


Figure 7.57 Schematic representation of the development of llama antibodies against influenza virus hemagglutinin and the use of those antibodies to protect mice against influenza virus.

One group of researchers immunized llamas with a commercial (traditional) influenza vaccine plus purified hemagglutinin protein isolated from several influenza A and B strains along with purified recombinant hemagglutinin proteins from other influenza viruses (Fig. 7.57). Subsequently, the llama antibodies directed against highly conserved epitopes on the hemagglutinin protein were isolated, and then the genes encoding these antibodies were isolated and characterized. The genes encoding the variable regions from four different llama antihemagglutinin antibody genes were combined into a single gene encoding a broadly neutralizing antibody (Fig. 7.58) that was then inserted into an adeno-associated virus vector (see chapter 6) which was used to deliver this unique llama antibody (fusion protein) directed against the conserved region of hemagglutinin into mice, either intranasally or intravenously. When it was tested *in vitro*, the fusion protein was able to neutralize 60 different influenza A and B viruses. Importantly, the intranasally treated mice were protected against challenges with lethal doses of influenza virus as efficiently as when the fusion protein was delivered intravenously. It is thought that intranasal delivery may provide human populations with a greater level of protection against influenza virus, as this is the natural route of the virus infection. The rationale for using llama antibodies in this experiment is that llama IgG molecules contain only heavy chains with a single variable region (Fig. 7.58), so the llama IgG molecules are smaller and therefore easier to manipulate than conventional mouse or human antibodies.

A recently discovered problem with currently existing influenza vaccines is that the immunity provided by these vaccines is not long-lasting.

Figure 7.58 Schematic representation of **(A)** llama IgG molecule and **(B)** engineered llama IgG molecule. In panel B each of the variable segments is derived from a different llama anti-influenza monoclonal antibody.



It was observed that individuals contain high levels of serum antibody against the influenza strains in the inoculum at 4 weeks postinoculation. However, these serum antibody levels decrease to preinoculation levels 1 year after the inoculation. Thus, in addition to finding a more universal influenza vaccine, it is necessary to significantly enhance the persistence of both current and future vaccines.

summary

Traditionally, vaccines have been either inactivated or attenuated infectious agents (bacteria, parasites, or viruses) that are injected into an antibody-producing organism to produce immunity. There are a number of drawbacks to these vaccines. For example, not all pathogenic organisms can be grown in the large volumes needed to make a vaccine, there are safety concerns when large volumes of pathogenic organisms are being handled, attenuated strains may revert to the infectious state, inactivation may be incomplete, and shelf life is often dependent on refrigeration.

A number of different strategies have been used to create effective vaccines, and often the most effective vaccine strategy must be determined by trying several different approaches. Traditional vaccines include either killed or naturally attenuated pathogenic agents. Recombinant DNA technology has been used in various ways to create more reliable vaccines. Recombinant vaccines include the use of virus-like particles, cloned antigenic pathogen proteins and their fragments, peptide vaccines, DNA constructs encoding the antigenic protein or peptide, or recombinant viral or bacterial vectors encoding pathogen antigens. Immunologically active, noninfectious agents are produced by deleting the genes that cause virulence; with this deletion, a live vaccine would never be able to revert to the infectious form. A gene(s) that encodes the major antigenic determinant(s) from a pathogenic organism can be cloned into the genome of a benign carrier organism (usually a virus or bacterium), which can be used as a vaccine without concern that any pathogenic organisms are present. The genes or segments of genes that encode the major antigenic determinants of pathogenic organisms can be cloned into expression vectors, and large amounts of the product can be harvested, purified, and used as a vaccine. Complete genes produce subunit vaccines, and cloned domains of the major antigenic determinants produce peptide vaccines. Peptide vaccines that represent only a portion of a key antigenic determinant may also be produced by chemical peptide synthesis. As an alternative to using cloned antigenic proteins or peptides for inoculation, DNA constructs encoding the antigenic protein or peptide may be utilized. These DNA constructs may be delivered directly to animals or humans. There are currently a large number of safer and more effective vaccines being developed using these techniques, potentially greatly improving human health and saving millions of lives.

REFERENCES

Agarwal K, Agarwal S. 2008. *Helicobacter pylori* vaccine: from past to future. *Mayo Clin Proc* **83:**169–175.

Ahn SK, Tran V, Leung A, Ng M, Li M, Liu J. 2018. Recombinant BCG overexpressing *phoP-phoR* confers enhanced protection against tuberculosis. *Mol Ther* **26:**2863–2874.

Amara RR, Villinger F, Altman JD, Lydy SL, O'Neil SP, Staprans SI, Montefiori DC, Xu Y, Herndon JG, Wyatt LS, Candido MA, Kozyr NL, Earl PL, Smith JM, Ma H-L, Grimm BD, Hulsey ML, Miller J, McClure HM, McNicholl JM, Moss B, Robinson HL. 2001. Control of a mucosal challenge and prevention of AIDS by a multiprotein DNA/MVA vaccine. *Science* **292:**69–74.

Audran R, Cachat M, Lurati F, Soe S, Leroy O, Corradin G, Druilhe P, Spertini F. 2005. Phase I malaria vaccine trial with a long synthetic peptide derived from the merozoite surface protein 3 antigen. *Infect Immun* **73:**8017–8026.

Bagnoli F, Fontana MR, Soldaini E, Mishra RPN, Fiaschi L, Cartocci E, Nardi-Dei V, Ruggiero P, Nosari S, De Falco MG, Lofano G, Marchi S, Galletti B, Mariotti P, Bacconi M, Torre A, Maccari S, Scarselli M, Rinaudo CD, Inoshima N, Savino S, Mori E, Rossi-Paccani S, Baudner B, Pallaoro M, Swennen E, Petracca R, Brettoni C, Liberatori S, Norais N, Monaci E, Bubeck Wardenburg J, Schneewind O, O'Hagan DT, Valiante NM, Bensi G, Bertholet S, De Gregorio E, Rappuoli R, Grandi G. 2015. Vaccine composition formulated with a novel TLR7-dependent adjuvant induces high and broad protection against *Staphylococcus aureus*. *Proc Natl Acad Sci USA* **112:**3680–3685.

Ball P. 2020. Anti-vaccine movement might undermine pandemic efforts. *Nature* **581:**251.

Baum A, Fulton BO, Wloga E, Copin R, Pascal KE, Russo V, Giordano S, Lanza K, Negron N, Ni M, Wei Y, Atwal GS, Murphy AJ, Stahl N, Yancopoulos GD, Kyratsous CA. 2020. Antibody cocktail to SARS-CoV-2 spike protein prevents rapid

mutational escape seen with individual antibodies. *Science* **369:**1014–1018.

Blancou J, Kieny MP, Lathe R, Lecocq JP, Pastoret PP, Soulebot JP, Desmettre P. 1986. Oral vaccination of the fox against rabies using a live recombinant vaccinia virus. *Nature* **322:**373–375.

Blasco R, Moss B. 1995. Selection of recombinant vaccinia viruses on the basis of plaque formation. *Gene* **158:**157–162.

Bustamante LY, Powell GT, Lin Y-C, Macklin MD, Cross N, Kemp A, Cawkill P, Sanderson T, Crosnier C, Muller-Sienerth N, Doumbo OK, Traore B, Crompton PD, Cicuta P, Tran TM, Wright GJ, Rayner JC. 2017. Synergistic malaria vaccine combinations identified by systematic antigen screening. *Proc Natl Acad Sci USA* **114:**12045–12050.

Charles I, Dougan G. 1990. Gene expression and the development of live enteric vaccines. *Trends Biotechnol* **8:**117–121.

Chatfield SN, Charles IG, Makoff AJ, Oxer MD, Dougan G, Pickard D, Slater D, Fairweather NF. 1992. Use of the *nirB* promoter to direct the stable expression of heterologous antigens in *Salmonella* oral vaccine strains: development of a single-dose oral tetanus vaccine. *Biotechnology (N Y)* **10:**888–892.

Cichutek K. 2000. DNA vaccines: development, standardization and regulation. *Intervirology* **43:**331–338.

Clarke BE, Newton SE, Carroll AR, Francis MJ, Appleyard G, Syred AD, Highfield PE, Rowlands DJ, Brown F. 1987. Improved immunogenicity of a peptide epitope after fusion to hepatitis B core protein. *Nature* **330:**381–384.

Corbett M, Bogers WM, Heeney JL, Gerber S, Genin C, Didierlaurent A, Oostermeijer H, Dubbes R, Braskamp G, Lerondel S, Gomez CE, Esteban M, Wagner R, Kondova I, Mooij P, Balla-Jhagjhoorsingh S, Beenhakker N, Koopman G, van der Burg S, Kraehenbuhl J-P, Le Pape A. 2008. Aerosol immunization with NYVAC and MVA vectored vaccines is safe, simple, and immunogenic. *Proc Natl Acad Sci USA* **105:**2046–2051.

Corum J, Zimmer C. 2021. How nine Covid-19 vaccines work. *New York Times*. 11 January 2021.

Cremer KJ, Mackett M, Wohlenberg C, Notkins AL, Moss B. 1985. Vaccinia virus recombinant expressing herpes simplex virus type 1 glycoprotein D prevents latent herpes in mice. *Science* **228:**737–740.

Da Costa XJ, Jones CA, Knipe DM. 1999. Immunization against genital herpes with a vaccine virus that has defects in productive and latent infection. *Proc Natl Acad Sci USA* **96:**6994–6998.

Davies MR, McIntyre L, Mutreja A, Lacey JA, Lees JA, Towers RJ, Duchêne S, Smeesters PR, Frost HR, Price DJ, Holden MTG, David S, Giffard PM, Worthing KA, Seale AC, Berkley JA, Harris SR, Rivera-Hernandez T, Berking O, Cork AJ, Torres RSLA, Lithgow T, Strugnell RA, Bergmann R, Nitsche-Schmitz P, Chhatwal GS, Bentley SD, Fraser JD, Moreland NJ, Carapetis JR, Steer AC, Parkhill J, Saul A, Williamson DA, Currie BJ, Tong SYC, Dougan G, Walker MJ. 2019. Atlas of group A streptococcal vaccine candidates compiled using large-scale comparative genomics. *Nat Genet* **51:**1035–1043.

Davis CW, Jackson KJL, McCausland MM, Darce J, Chang C, Linderman SL, Chennareddy C, Gerkin R, Brown SJ, Wrammert J, Mehta AK, Cheung WC, Boyd SD, Waller EK, Ahmed R. 2020. Influenza vaccine-induced human bone marrow plasma cells decline within a year after vaccination. *Science* **370:**237–241.

Del Campo J, Pizzorno A, Djebali S, Bouley J, Haller M, Pérez-Vargas J, Lina B, Boivin G, Hamelin M-E, Nicolas F, Le Vert A, Leverrier Y, Rosa-Calatrava M, Marvel J, Hill F. 2018. OVX836 a recombinant nucleoprotein vaccine inducing cellular responses and proactive efficacy against multiple influenza A subtypes. *Vaccine* **4:**4.

DiMarchi R, Brooke G, Gale C, Cracknell V, Doel T, Mowat N. 1986. Protection of cattle against foot-and-mouth disease by a synthetic peptide. *Science* **232:**639–641.

Dowd KA, Ko S-Y, Morabito KM, Yang ES, Pelc RS, DeMaso CR, Castilho LR, Abbink P, Boyd M, Nityanandam R, Gordon DN, Gallagher JR, Chen X, Todd J-P, Tsybovsky Y, Harris A, Huang YS, Higgs S, Vanlandingham DL, Andersen H, Lewis MG, De La Barrera R, Eckels KH, Jarman RG, Nason MC, Barouch DH, Roederer M, Kong W-P, Mascola JR, Pierson TC, Graham BS. 2016. Rapid development of a DNA vaccine for Zika virus. *Science* **354:**237–240.

Du L, He Y, Zhou Y, Liu S, Zheng B-J, Jiang S. 2009. The spike protein of SARS-CoV: a target for vaccine and therapeutic development. *Nat Rev Microbiol* **7:**226–236.

Ekiert DC, Friesen RHE, Bhabha G, Kwaks T, Jongeneelen M, Yu W, Ophorst C, Cox F, Korse HJ, Brandenburg B, Vogels R, Brakenhoff JP, Kompier R, Koldijk MH, Cornelissen LA, Poon LL, Peiris M, Koudstaal W, Wilson IA, Goudsmit J. 2011. A highly conserved neutralizing epitope on group 2 influenza A viruses. *Science* **333:**843–850.

Endmann A, Klünder K, Kapp K, Riede O, Oswald D, Talman EG, Schroff M, Kleuss C, Ruiters MHJ, Juhls C. 2014. Cationic lipid-formulated DNA vaccine against hepatitis B virus: immunogenicity of MIDGE-Th1 vectors encoding small and large surface antigen in comparison to a licensed protein vaccine. *PLoS One* **9:**e101715.

Fernando GJP, Chen X, Prow TW, Crichton ML, Fairmaid EJ, Roberts MS, Frazer IH, Brown LE, Kendall MAF. 2010. Potent immunity to low doses of influenza vaccine by probabilistic guided micro-targeted skin delivery in a mouse model. *PLoS One* **5:**e10266.

Flingai S, Czerwonko M, Goodman J, Kudchodkar SB, Muthumani K, Weiner DB. 2013. Synthetic DNA vaccines: improved vaccine potency by electroporation and co-delivered genetic adjuvants. *Front Immunol* **4:**354.

Galen JE, Pasetti MF, Tennant S, Ruiz-Olvera P, Sztein MB, Levine MM. 2009. *Salmonella enterica* serovar Typhi live vector vaccines finally come of age. *Immunol Cell Biol* **87:**400–412.

Galen JE, Buskirk AD, Tennant SM, Pasetti MF. 2016. Live attenuated human *Salmonella* vaccine candidates: tracking the pathogen in natural infection and stimulation of host immunity. *EcoSal Plus* **7:**10.

Gómez-Duarte OG, Bumann D, Meyer TF. 1999. The attenuated *Salmonella* vaccine approach for the control of *Helicobacter pylori*-related diseases. *Vaccine* **17:**1667–1673.

Grigg P, Titong A, Jones LA, Yilma TD, Verardi PH. 2013. Safety mechanism assisted by the repressor of tetracycline (SMART) vaccinia virus vectors for vaccines and therapeutics. *Proc Natl Acad Sci USA* **110:**15407–15412.

Guy B, Barrere B, Malinowski C, Saville M, Teyssou R, Lang J. 2011. From research to phase III: preclinical, industrial and clinical development of the Sanofi Pasteur tetravalent dengue vaccine. *Vaccine* **29:**7229–7241.

Hessel A, Savidis-Dacho H, Coulibaly S, Portsmouth D, Kreil TR, Crowe BA, Schwendinger MG, Pilz A, Barrett PN, Falkner FG, Schäfer B. 2014. MVA vectors expressing conserved influenza proteins protect mice against lethal challenge with H5N1, H9N2 and H7N1 viruses. *PLoS One* **9:**e88340.

Hongying F, Xianbo W, Fang Y, Yang B, Beiguo L. 2014. Oral immunization with recombinant *Lactobacillus acidophilus* expressing the adhesin Hp0410 of *Helicobacter pylori* induces mucosal and systemic immune responses. *Clin Vaccine Immunol* **21:**126–132.

Horwitz MA, Lee B-WE, Dillon BJ, Harth G. 1995. Protective immunity against tuberculosis induced by vaccination with major extracellular proteins of *Mycobacterium tuberculosis*. *Proc Natl Acad Sci USA* **92:**1530–1534.

Horwitz MA, Harth G, Dillon BJ, Maslesa-Galic′ S. 2000. Recombinant bacillus Calmette-Guerin (BCG) vaccines expressing the *Mycobacterium tuberculosis* 30-kDa major secretory protein induce greater protective immunity against tuberculosis than conventional BCG vaccines in a highly susceptible animal model. *Proc Natl Acad Sci USA* **97:**13853–13858.

Horwitz MA, Harth G. 2003. A new vaccine against tuberculosis affords greater survival after challenge than the current vaccine in the guinea pig model of pulmonary tuberculosis. *Infect Immun* **71:**1672–1679.

Jia R, Guo JH, Fan MW, Bian Z, Chen Z, Fan B, Yu F, Xu QA. 2006. Immunogenicity of CTLA4 fusion anti-caries DNA vaccine in rabbits and monkeys. *Vaccine* **24:**5192–5200.

Kaper JB, Morris JG Jr, Levine MM. 1995. Cholera. *Clin Microbiol Rev* **8:**48–86.

Kaslow DC, Isaacs SN, Quakyi IA, Gwadz RW, Moss B, Keister DB. 1991. Induction of *Plasmodium falciparum* transmission-blocking antibodies by recombinant vaccinia virus. *Science* **252:**1310–1313.

Kim E, Erdos G, Huang S, Kenniston TW, Balmert SC, Carey CD, Raj VS, Epperly MW, Klimstra WB, Haagmans BL, Korkmaz E, Falo LD Jr, Gambotto A. 2020. Microneedle array delivered recombinant coronavirus vaccines: immunogenicity and rapid translational development. *EBioMedicine* **55:**102743.

Kirnbauer R, Booy F, Cheng N, Lowy DR, Schiller JT. 1992. Papillomavirus L1 major capsid protein self-assembles into virus-like particles that are highly immunogenic. *Proc Natl Acad Sci USA* **89:**12180–12184.

Knudsen NPH, Nørskov-Lauritsen S, Dolganov GM, Schoolnik GK, Lindenstrøm T, Andersen P, Agger EM, Aagaard C. 2014. Tuberculosis vaccine with high predicted population coverage and compatibility with modern diagnostics. *Proc Natl Acad Sci USA* **111:**1096–1101.

Koutsky LA, Ault KA, Wheeler CM, Brown DR, Barr E, Alvarez FB, Chiacchierini LM, Jansen KU, Proof of Principle Study Investigators. 2002. A controlled trial of a human papillomavirus type 16 vaccine. *N Engl J Med* **347:**1645–1651.

Krammer F. 2020. SARS-CoV-2 vaccines in development. *Nature* **586:**516–527.

Larocca RA, Abbink P, Peron JPS, Zanotto PM, Iampietro MJ, Badamchi-Zadeh A, Boyd M, Ng'ang'a D, Kirilova M, Nityanandam R, Mercado NB, Li Z, Moseley ET, Bricault CA, Borducchi EN, Giglio PB, Jetton D, Neubauer G, Nkolola JP, Maxfield LF, De La Barrera RA, Jarman RG, Eckels KH, Michael NL, Thomas SJ, Barouch DH. 2016. Vaccine protection against Zika virus from Brazil. *Nature* **536:**474–478.

Lasky LA, Dowbenko D, Simonsen CC, Berman PW. 1984. Protection of mice from lethal herpes simplex virus infection by vaccination with a secreted form of cloned glycoprotein D. *Bio/Technology* **2:**527–532.

Laurens MB. 2020. RTS,S/AS01 vaccine (Mosquirix™): an overview. *Hum Vaccin Immunother* **16:**480–489.

Laurens MB. 2018. The promise of a malaria vaccine: are we closer? *Annu Rev Microbiol* **72:**273–292.

Laursen NS, Friesen RHE, Zhu X, Jongeneelen M, Blokland S, Vermond J, van **Eijgen A, Tang C,** van **Diepen H, Obmolova G,** van der **Neut Kolfschoten M, Zuijdgeest D, Straetemans R, Hoffman RMB, Nieusma T, Pallesen J, Turner HL, Bernard SM, Ward AB, Luo J, Poon LLM, Tretiakova AP, Wilson JM, Limberis MP, Vogels R, Brandenburg B, Kolkman JA, Wilson IA.** 2018. Universal protection against influenza infection by a multidomain antibody to influenza hemagglutinin. *Science* **362:**598–602.

Leitner WW, Ying H, Restifo NP. 1999. DNA and RNA-based vaccines: principles, progress and prospects. *Vaccine* **18:**765–777.

Liao H-Y, Wang S-C, Ko YA, Lin KI, Ma C, Cheng TR, Wong CH. 2020. Chimeric hemagglutinin vaccine elicits broadly protective CD4 and CD8 T cell responses against multiple influenza strains and subtypes. *Proc Natl Acad Sci USA* **117:**17757–17763.

Londoño-Arcila P, Freeman D, Kleanthous H, O'Dowd AM, Lewis S, Turner AK, Rees EL, Tibbitts TJ, Greenwood J, Monath TP, Darsley MJ. 2002. Attenuated *Salmonella enterica* serovar Typhi expressing urease effectively immunizes mice against *Helicobacter pylori* challenge as part of a heterologous mucosal priming-parenteral boosting vaccination regimen. *Infect Immun* **70:**5096–5106.

Lopes A, Vandermeulen G, Préat V. 2019. Cancer DNA vaccines: current preclinical and clinical developments and future perspectives. *J Exp Clin Cancer Res* **38:**146.

Maki J, Guiot AL, Aubert M, Brochier B, Cliquet F, Hanlon CA, King R, Oertli EH, Rupprecht CE, Schumacher C, Slate D, Yakobson B, Wohlers A, Lankau EW. 2017. Oral vaccination of wildlife using a vaccinia-rabies-glycoprotein recombinant virus vaccine (RABORAL V-RG®): a global review. *Vet Res (Faisalabad)* **48:**57.

Manoj S, Babiuk LA, van Drunen Littel-van den Hurk S. 2004. Approaches to enhance the efficacy of DNA vaccines. *Crit Rev Clin Lab Sci* **41:**1–39.

Martinez-Alonso S, Martinez-Lopez A, Estepa A, Cuesta A, Tafalla C. 2011. The introduction of multi-copy CpG motifs into an antiviral DNA vaccine strongly up-regulates its immunogenicity in fish. *Vaccine* **29:**1289–1296.

Marzi A, Feldmann H, Geisbert TW, Falzarano D. 2011. Vesicular stomatitis virus-based vaccines for prophylaxis and treatment of filovirus infections. *J Bioterror Biodef* **S1**(Suppl 1)**:**4.

Marzi A, Robertson SJ, Haddock E, Feldmann F, Hanley PW, Scott DP, Strong JE, Kobinger G, Best SM, Feldmann H. 2015. Ebola vaccine. VSV-EBOV rapidly protects macaques against infection with the 2014/15 Ebola virus outbreak strain. *Science* **349:**739–742.

McCluskie MJ, Davis HL. 1999. Mucosal immunization with DNA vaccines. *Microbes Infect* **1:**685–698.

Mekalanos JJ, Sadoff JC. 1994. Cholera vaccines: fighting an ancient scourge. *Science* **265:**1387–1389.

de **Camargo TM,** de **Freitas EO, Gimenez AM, Lima LC,** de **Almeida Caramico K, Françoso KS, Bruna-Romero O, Andolina C, Nosten F, Rénia L, Ertl HCJ, Nussenzweig RS, Nussenzweig V, Rodrigues MM, Reyes-Sandoval A, Soares IS.** 2018. Prime-boost vaccination with recombinant protein and adenovirus-vector expressing *Plasmodium vivax* circumsporozoite protein (CSP) partially protects mice against Pb/Pv sporozoite challenge. *Sci Rep* **8:**1118.

Michel M-L, Davis HL, Schleef M, Mancini M, Tiollais P, Whalen RG. 1995. DNA-mediated immunization to the hepatitis B surface antigen in mice: aspects of the humoral response mimic hepatitis B viral infection in humans. *Proc Natl Acad Sci USA* **92:**5307–5311.

Mina MJ, Kula T, Leng Y, Li M, de **Vries RD, Knip M, Siljander H, Rewers M, Choy DF, Wilson MS, Larman HB, Nelson AN, Griffin DE, de Swart RL, Elledge SJ.** 2019. Measles virus infection diminishes preexisting antibodies that offer protection from other pathogens. *Science* **366:**599–606.

Moon JJ, Suh H, Li AV, Ockenhouse CF, Yadava A, Irvine DJ. 2012. Enhancing humoral responses to a malaria antigen with nanoparticle vaccines that expand T_{fh} cells and promote germinal center induction. *Proc Natl Acad Sci USA* **109:**1080–1085.

Moss B. 1991. Vaccinia virus: a tool for research and vaccine development. *Science* **252:**1662–1667.

Newton SMC, Jacob CO, Stocker BAD. 1989. Immune response to cholera toxin epitope inserted in *Salmonella* flagellin. *Science* **244:**70–72.

Nussenzweig RS, Long CA. 1994. Malaria vaccines: multiple targets. *Science* **265:**1381–1383.

Nuccitelli A, Cozzi R, Gourlay LJ, Donnarumma D, Necchi F, Norais N, Telford JL, Rappuoli R, Bolognesi M, Maione D, Grandi G, Rinaudo CD. 2011. Structure-based approach to rationally design a chimeric protein for an effective vaccine against group B *Streptococcus* infections. *Proc Natl Acad Sci USA* **108:**10278–10283.

Odevall L, Hong D, Digilio L, Sahastrabuddhe S, Mogasale V, Baik Y, Choi S, Kim JH, Lynch J. 2018. The Euvichol story: development and licensure of a safe, effective and affordable oral cholera vaccine through global public private partnerships. *Vaccine* **36:**6606–6614.

Ottenhoff THM, Kaufmann SHE. 2012. Vaccines against tuberculosis: where are we and where do we need to go? *PLoS Pathog* **8:**e1002607.

Paoletti E. 1996. Applications of pox virus vectors to vaccination: an update. *Proc Natl Acad Sci USA* **93:**11349–11353.

Paoletti E, Tartaglia J. January 1995. Interferon sensitive recombinant poxvirus vaccine. U.S. patent 5,378,457.

Patel M. 2020. Dental caries vaccine: are we there yet? *Lett Appl Microbiol* **70:**2–12.

Petro C, González PA, Cheshenko N, Jandl T, Khajoueinejad N, Bénard A, Sengupta M, Herold BC, Jacobs WR Jr. 2015. Herpes simplex type 2 virus deleted in glycoprotein D protects against vaginal, skin and neural disease. *eLife* **4:**e06054.

Porta C, Kotecha A, Burman A, Jackson T, Ren J, Loureiro S, Jones IM, Fry EE, Stuart DI, Charleston B. 2013. Rational engineering of recombinant picornavirus capsids to produce safe, protective vaccine antigen. *PLoS Pathog* **9:**e1003255.

Raj DK, Das Mohapatra A, Jnawali A, Zuromski J, Jha A, Cham-Kpu G, Sherman B, Rudlaff RM, Nixon CE, Hilton N, Oleinikov AV, Chesnokov O, Merritt J, Pond-Tor S, Burns L, Jolly G, Ben Mamoun C, Kabyemela E, Muehlenbachs A, Lambert L, Orr-Gonzalez S, Gnädig NF, Fidock DA, Park S, Dvorin JD, Pardi N, Weissman D, Mui BL, Tam YK, Friedman JF, Fried M, Duffy PE, Kurtis JD. 2020. Anti-PfGARP activates programmed cell death of parasites and reduces severe malaria. *Nature* **582:**104–108.

Regules JA, Cummings JF, Ockenhouse CF. 2011. The RTS,S vaccine candidate for malaria. *Expert Rev Vaccines* **10:**589–599.

Rice J, Ottensmeier CH, Stevenson FK. 2008. DNA vaccines: precision tools for activating effective immunity against cancer. *Nat Rev Cancer* **8:**108–120.

Rosa BR, Cunha AJLAD, Medronho RA. 2019. Efficacy, immunogenicity and safety of a recombinant tetravalent dengue vaccine (CYD-TDV) in children aged 2-17 years: systematic review and meta-analysis. *BMJ Open* **9:**e019368.

Santi AMM, Lanza JS, Tunes LG, Fiuza JA, Roy G, Orfanó ADS, de **Carvalho AT, Frézard F, Barros ALB, Murta SMF,** do **Monte-Neto RL.** 2018. Growth arrested live-attenuated *Leishmania infantum* KHARON1 null mutants display cytokinesis defect and protective immunity in mice. *Sci Rep* **8:**11627.

Shaikh H, Lynch J, Kim J, Excler J-L. 2020. Current and future cholera vaccines. *Vaccine* **38**(Suppl 1)**:**A118–A126.

Shaw AR. 2013. Human papillomavirus vaccines six years after approval. *Annu Rev Med* **64:**91–100.

Sheridan C. 2005. The business of making vaccines. *Nat Biotechnol* **23:**1359–1366.

Singh M, Briones M, Ott G, O'Hagan D. 2000. Cationic microparticles: a potent delivery system for DNA vaccines. *Proc Natl Acad Sci USA* **97:**811–816.

Singh S, Soe S, Mejia J-P, Roussilhon C, Theisen M, Corradin G, Druilhe P. 2004. Identification of a conserved region of *Plasmodium falciparum* MSP3 targeted by biologically active antibodies to improve vaccine design. *J Infect Dis* **190:**1010–1018.

Sizemore DR, Branstrom AA, Sadoff JC. 1995. Attenuated *Shigella* as a DNA delivery vehicle for DNA-mediated immunization. *Science* **270:**299–302.

Stover CK, de la Cruz VF, Fuerst TR, Burlein JE, Benson LA, Bennett LT, Bansal GP, Young JF, Lee MH, Hatfull GF, Snapper SB, Barletta RG, Jacobs WR Jr, Bloom BR. 1991. New use of BCG for recombinant vaccines. *Nature* **351:**456–460.

Stranger-Jones YK, Bae T, Schneewind O. 2006. Vaccine assembly from surface proteins of *Staphylococcus aureus*. *Proc Natl Acad Sci USA* **103:**16942–16947.

Takahashi K, Orito N, Tokunoh N, Inoue N. 2019. Current issues regarding the application of recombinant lactic acid bacteria to mucosal vaccine carriers. *Appl Microbiol Biotechnol* **103:**5947–5955.

Titus RG, Gueiros-Filho FJ, de Freitas LAR, Beverley SM. 1995. Development of a safe live *Leishmania* vaccine line by gene replacement. *Proc Natl Acad Sci USA* **92:**10267–10271.

Ulmer JB, Donnelly JJ, Parker SE, Rhodes GH, Felgner PL, Dwarki VJ, Gromkowski SH, Deck RR, DeWitt CM, Friedman A, Hawe LA, Leander KR, Martinez D, Perry HC, Shiver JW, Montgomery DL, Liu MA. 1993. Heterologous protection against influenza by injection of DNA encoding a viral protein. *Science* **259:**1745–1749.

Wang S, Kong Q, Curtiss R III. 2013. New technologies in developing recombinant attenuated *Salmonella* vaccine vectors. *Microb Pathog* **58:**17–28.

Weeratna RD, McCluskie MJ, Comanita L, Wu T, Davis HL. 2000. Optimization strategies for DNA vaccines. *Intervirology* **43:**218–226.

Wells J. 2011. Mucosal vaccination and therapy with genetically modified lactic acid bacteria. *Annu Rev Food Sci Technol* **2:**423–445.

Wen X, Tong X, Wang M, Wang J, Ni H, Ran X. 2020. Protective immunity following vaccination with a recombinant multiple-epitope protein of bovine herpesvirus type I in a rabbit model. *Appl Microbiol Biotechnol* **104:**3011–3023.

Wyszyńska A, Kobierecka P, Bardowski J, Jagusztyn-Krynicka EK. 2015. Lactic acid bacteria: 20 years exploring their potential as live vectors for mucosal vaccination. *Appl Microbiol Biotechnol* **99:**2967–2977.

Xu C, Zhang B-Z, Lin Q, Deng J, Yu B, Arya S, Yuen K-Y, Huang JD. 2018. Live attenuated *Salmonella typhimurium* vaccines delivering SaEsxA and SaEsxB via type III secretion system confer protection against *Staphylococcus aureus* infection. *BMC Infect Dis* **18:**195.

Xu QA, Yu F, Fan MW, Bian Z, Chen Z, Peng B, Jia R, Guo JH. 2007. Protective efficacy of a targeted anti-caries DNA plasmid against cariogenic bacteria infections. *Vaccine* **25:**1191–1195.

Zakhartchouk AN, Sharon C, Satkunarajah M, Auperin T, Viswanathan S, Mutwiri G, Petric M, See RH, Brunham RC, Finlay BB, Cameron C, Kelvin DJ, Cochrane A, Rini JM, Babiuk LA. 2007. Immunogenicity of a receptor-binding domain of SARS coronavirus spike protein in mice: implications for a subunit vaccine. *Vaccine* **25:**136–143.

Zimet GD, Shew ML, Kahn JA. 2008. Appropriate use of cervical cancer vaccine. *Annu Rev Med* **59:**223–236.

review questions

1. Briefly describe a protocol for developing a vaccine against a toxin-producing bacterium.
2. Briefly describe the different types of vaccines that might be generated against a viral pathogen.
3. What factors limit the usefulness of conventional vaccines?
4. Describe the development of a subunit vaccine directed against influenza virus.
5. As part of your work with an international animal health organization, you are given the task of developing a vaccine against a bovine virus that is the cause of tens of thousands of cattle deaths around the world annually. The viral genome consists of a 10-kb linear piece of single-stranded RNA with a poly(A) tail that encodes eight different proteins. The virus does not have a viral envelope, and the major antigenic determinant is the capsid protein viral protein 2. Outline an experimental strategy to develop a vaccine against this virus.
6. Briefly describe an approach that you might take in developing a vaccine against an antibiotic-resistant bacterial pathogen.
7. Discuss the development of peptide vaccines that are directed against malaria.
8. What is vaccinia virus, and how can it be used to produce unique live recombinant vaccines?
9. As an employee of the World Health Organization, you have to decide on the best strategy for eradicating rabies in wild animal populations. Assuming that you must choose between a DNA vaccine and a vaccinia virus-based vaccine, select one type of vaccine and justify your choice.
10. Discuss the advantages of a live recombinant viral vaccine over killed and subunit vaccines.
11. Discuss some of the different strategies that have been used to produce vaccines against cholera.
12. How would you develop a subunit vaccine against herpes simplex virus?
13. How can different bacteria be used as part of a DNA vaccine delivery system?
14. How can vaccinia virus be made more sensitive to interferon? Explain.
15. Suggest several methods that you could use to deliver DNA for genetic immunization to human and animal cells.

16. What are MIDGE vectors, and how can they be used to facilitate genetic immunization?
17. How would you develop a vaccine against human papillomavirus?
18. How would you develop an effective DNA vaccine against dental caries?
19. How would you improve the traditional vaccine against tuberculosis?
20. How would you develop a subunit against SARS?
21. How would develop a vaccine against group B *Streptococcus*?
22. What is a Nanopatch, and why is it effective?
23. What is the modified vaccinia Ankara virus vector system, and why is it useful?
24. How can llama antibodies be used to develop vaccines?
25. What is passive immunity, and how can it be used to develop a vaccine against influenza virus?

Restriction Endonucleases
Small Biological Molecules
L-Ascorbic Acid
Indigo
Amino Acids
Lycopene
Antibiotics
Biopolymers
Solvent Tolerance
Systems Metabolic Engineering to Optimize Product Yield
Microbial Degradation of Xenobiotics
Genetic Engineering of Biodegradative Pathways
Plastics
Utilization of Starch and Sugars
Commercial Production of Fructose and Alcohol
Increasing Alcohol Production
Improving Fructose Production
Utilization of Cellulose and Hemicellulose
Lignocellulosics
Cellulase Genes
Direct Conversion of Biomass to Ethanol
Alcohol Production by *Zymomonas mobilis*
Lipids from Cyanobacteria
Hydrogen Production
SUMMARY
REFERENCES
REVIEW QUESTIONS

8 Industrial and Environmental Uses of Recombinant Microorganisms

Molecular biotechnology research was initially focused on the production of a range of different proteins, including enzymes that are used commercially. However, recombinant DNA techniques can also be used to enhance the production of low-molecular-weight compounds, such as vitamins, amino acids, dyes, precursors of biopolymers, and antibiotics that all have enormous economic value. In addition, researchers are using molecular biotechnology to develop unique strategies to deal with large-scale wastes, such as starch, lignocellulosics, and toxic substances. In this chapter, some of the developing industrial and environmental uses of molecular biotechnology are highlighted and discussed.

Restriction Endonucleases

Recombinant DNA technology would not be possible without a ready supply of different restriction endonucleases. Currently, of the estimated 3,000 known restriction endonucleases recognizing over 230 unique DNA sequences, more than 600 different enzymes are commercially available, with worldwide sales in 2020 of ~$240 million and expected to reach ~$360 million by 2026. These enzymes occur naturally in many different microorganisms, including species that are aerobic, anaerobic, photosynthetic, diazotrophic, mesophilic, thermophilic, psychrophilic, and both slow and fast growing. Initially, for each of the organisms that synthesized a restriction enzyme, a detailed fermentation protocol—specifying the temperature, pH, medium composition, and oxygen tension—had to be developed and optimized to achieve the maximum yield of the target enzyme.

Molecular Biotechnology: Principles and Applications of Recombinant DNA,
Sixth Edition. Bernard R. Glick and Cheryl L. Patten.
© 2022 American Society for Microbiology.

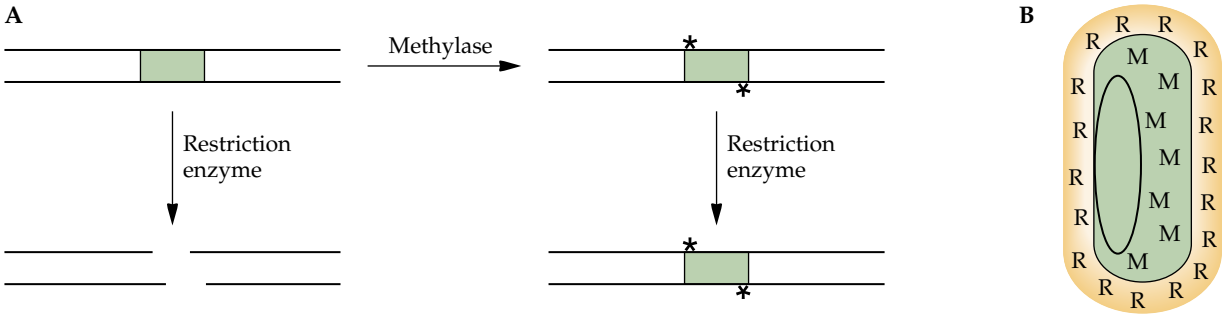
For a company that sold different restriction enzymes to scientists, it was initially necessary for them to maintain a large number of different microorganisms, stock a very wide range of microbial growth medium components, design several different types of fermenters (bioreactors), and spend an inordinate amount of time developing optimal growth conditions for a large number of different organisms (one major supplier of restriction enzymes lists 280 different enzymes in its catalog). To avoid these difficulties, scientists working for companies that were commercial suppliers of restriction endonucleases cloned and expressed the genes for various restriction enzymes in *Escherichia coli*. Exclusive use of *E. coli* allowed scientists to standardize the production conditions for all restriction endonucleases. In addition, *E. coli* cells grow rapidly to high cell densities and can be engineered to significantly overexpress each target restriction enzyme.

Although the technology for isolating and expressing foreign genes in *E. coli* is well established (chapter 2), it should be remembered that *E. coli* is a living entity, and the production or presence of some heterologous proteins may be toxic or even lethal. Restriction endonucleases can digest DNA at sites that are present on all DNA molecules. As a result, an organism that expresses a cloned restriction endonuclease gene is likely to have its own DNA degraded unless a protection mechanism is present.

Microorganisms that naturally make restriction endonucleases have evolved a highly effective self-protecting restriction-modification system. The endonuclease is produced together with a cognate methylase that methylates one or more of the bases of the DNA within the recognition sequence, thereby preventing the homologous restriction endonuclease from cutting the DNA at this site (Fig. 8.1A). Moreover, Gram-negative bacteria have an additional mode of protection that entails localizing the restriction endonuclease within the periplasmic space and the methylation (modification) enzyme in the cell cytoplasm (Fig. 8.1B). This compartmentalization physically separates the restriction endonuclease from the DNA while ensuring that the modification enzyme has ready access to the chromosomal DNA. In addition to protecting the cell from the potentially toxic effects of the restriction enzyme, this segregation provides a cellular defense against attack by any foreign DNA, such as DNA from a bacterial virus that might enter the periplasm.

One way to circumvent the problem of host DNA degradation by heterologous restriction endonucleases is to clone and express the genes for both the restriction and modification enzymes in the host organism. This is

Figure 8.1 **(A)** Protection of DNA from digestion by a restriction endonuclease by prior treatment with methylase, a methylating (modification) enzyme. The asterisks indicate the presence of a methylated base. **(B)** Cytoplasmic localization of modification enzyme (M) and periplasmic localization of restriction endonuclease (R) in Gram-negative bacteria.



facilitated by the common arrangement of the restriction endonuclease and methylase genes adjacent to each other on the bacterial chromosome, usually as part of the same operon.

A strategy to isolate the genes for restriction and modification enzyme systems was developed by a company that eventually became one of the world's leading suppliers of restriction enzymes and consists of the following steps.

1. A genomic DNA library was made from the DNA of a donor bacterium that contained a previously identified restriction endonuclease. The plasmid vector had at least one recognition site for the target restriction endonuclease.
2. The genomic library was introduced into *E. coli* by transformation. This step increased the amount of recombinant plasmid DNA and also allowed the expression of the modification enzyme.
3. Plasmid DNA was isolated from transformed cells that had been grown in liquid medium under conditions that selected for the presence of the plasmid.
4. The plasmid DNA preparation was treated with the target restriction endonuclease.
5. *E. coli* cells were transformed with the restriction endonuclease-treated plasmid DNA preparation.

The rationale for this procedure is that the clones that carry and express the target modification enzyme will produce plasmid DNA that has been methylated at the recognition sites by the modification enzyme and will therefore be resistant to digestion by the target restriction endonuclease. For example, following transformation of *E. coli* with a genomic library of *Desulfovibrio desulfuricans* DNA, plasmid DNA was isolated and digested with the restriction enzyme DdeI (Fig. 8.2). Plasmids that encoded and expressed the DdeI modification enzyme were not digested by DdeI restriction endonuclease because the eight DdeI recognition sites present on the plasmid vector were methylated. The DdeI-treated plasmids were used to transform *E. coli*. Only intact circular plasmids yield transformants, and these carry the gene for a functional DdeI modification enzyme. All other plasmids are degraded by the restriction endonuclease. The resulting transformants must then be assayed for the DdeI restriction enzyme activity to determine which clones have the genes for both the modification enzyme and the restriction endonuclease. In practice, this strategy is effective for any restriction enzyme gene that is physically close to its modification enzyme gene—most restriction enzymes are encoded in the same operon as their cognate modification enzyme—and is cloned into a plasmid vector that has at least one recognition site for the target enzyme.

Small Biological Molecules

With recombinant DNA technology, it is possible to modify metabolic pathways of organisms either by introducing new genes or by altering existing ones. The goal is to create an organism with a novel enzymatic activity that can convert an existing substrate into a commercial compound that with current technology can be produced by a combination of chemical treatments and fermentation steps. Early metabolic-engineering experiments typically modified one or two genes in a biosynthetic pathway. However, with the knowledge of the complete DNA sequences of many bacterial genomes (by mid-2020, the genomes of ~150,000 bacterial strains had been reported) and the regulation of expression of bacterial genes derived from

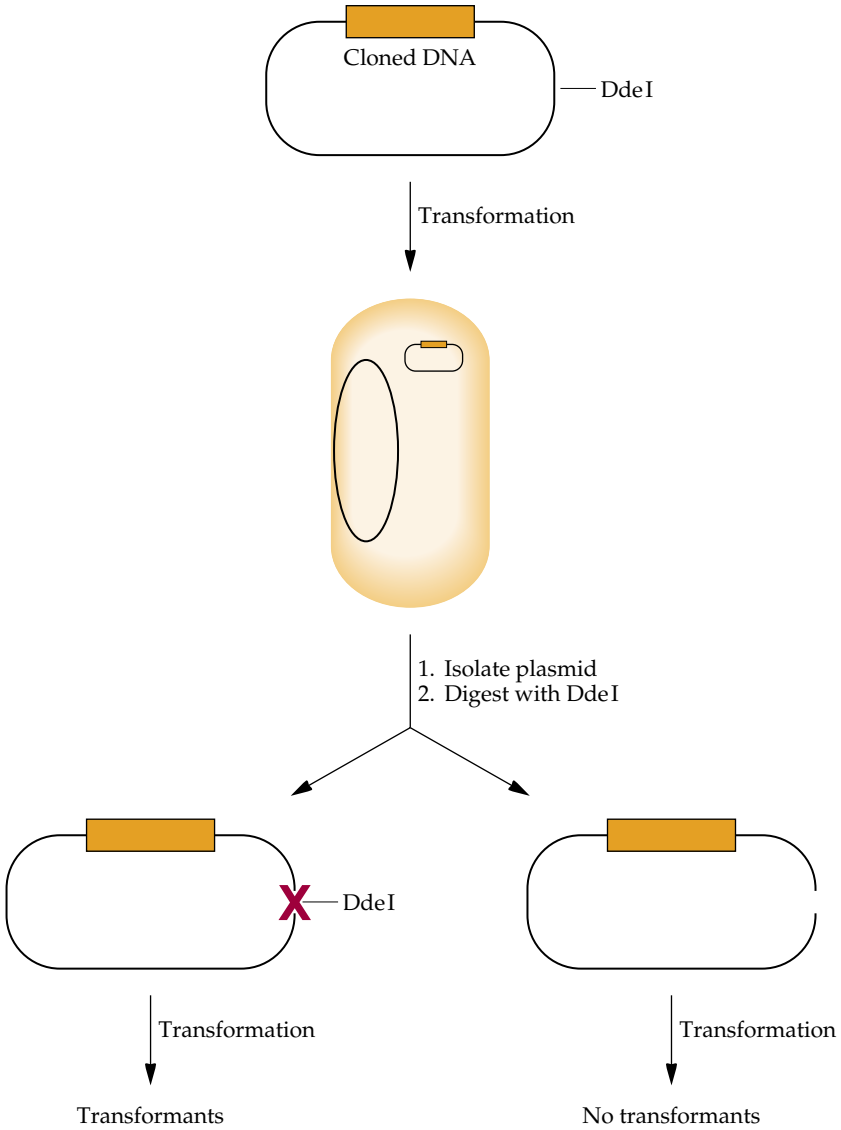


Figure 8.2 Scheme for cloning the DdeI modification enzyme. Plasmids containing cloned *D. desulfuricans* genomic DNA are used to transform *E. coli* cells. After the cells are grown in liquid culture, the plasmid DNA is isolated and digested with the restriction enzyme DdeI. Plasmids encoding the DdeI modification enzyme will be methylated at the DdeI restriction enzyme recognition site and will not be cut. The intact plasmid DNA is used to transform *E. coli* cells. The final transformants are assayed for the DdeI restriction enzyme and the corresponding methylase.

transcriptomic and proteomic studies (chapter 2), it is possible to develop strategies to improve the yield of nearly any bacterially produced molecule. This may be achieved by the introduction or modification of an entire panel of related genes, inhibiting the synthesis of individual enzymes, modulating the level of expression of various genes by altering their transcriptional promoters, removing feedback inhibition, and altering the activity of specific enzymes. Below are a few of the very many examples of the successful metabolic engineering of microorganisms to produce industrially useful

chemicals. These compounds include organic acids, alcohols, amino acids, and others (Table 8.1). Metabolic engineering has been used to enhance the synthesis of compounds in both laboratory strains, where productivity may be relatively low (usually <5 g per liter), and industrial strains, where productivity has generally been optimized (>50 g per liter).

L-Ascorbic Acid

L-Ascorbic acid (vitamin C) is naturally found in a wide range of different foods and is sold as a dietary supplement. It is used to prevent and to treat scurvy, is considered to be an essential nutrient in the repair of some tissues and the enzymatic production of some neurotransmitters, and is an essential component (coenzyme) in the functioning of several enzymes. In addition, there is some evidence that vitamin C may reduce the severity and duration of the common cold. In 2020, the worldwide market for vitamin C was estimated to be ~$7.6 billion.

Vitamin C is currently synthesized commercially from D-glucose by a process that includes one microbial fermentation step and a number of chemical steps (Fig. 8.3). The last step in this process is the acid-catalyzed conversion of 2-keto-L-gulonic acid (2-KLG) to L-ascorbic acid. However, some bacteria (*Acetobacter*, *Gluconobacter*, and *Erwinia*) can convert glucose to 2,5-diketo-D-gluconic acid (2,5-DKG), and others (*Corynebacterium*, *Brevibacterium*, and *Arthrobacter*) have the enzyme 2,5-DKG reductase, which converts 2,5-DKG to 2-KLG (Fig. 8.4A).

The current procedure for synthesizing ascorbic acid could be improved by producing 2-KLG from glucose by cofermentation with suitable organisms. Unfortunately, cocultivation has problems of its own. The medium requirements, temperature and pH optima, and growth rates of the two fermenting organisms differ in such a way that the fermentation conditions that are optimal for one organism are suboptimal for the other. This situation leads to the eventual "washout" (depletion or loss) of one of the organisms. Some of these problems may be overcome by utilizing a tandem fermentation process in which the two organisms are cultivated in succession. This approach requires two separate fermentations rather than one, and if the organisms have different growth requirements, it is difficult to run the process on a continuous basis. The best way to convert glucose into 2-KLG would be to engineer a single microorganism that carried all of the required enzymes (Fig. 8.4B). The conversion of D-glucose to 2,5-DKG by *Pantoea agglomerans* (formerly *Erwinia herbicola*) includes several enzymatic steps, whereas the transformation of 2,5-DKG to 2-KLG by a *Corynebacterium* sp. requires only one. Consequently, the simplest strategy for constructing a single organism that is able to convert D-glucose to 2-KLG is to isolate the 2,5-DKG reductase gene from the *Corynebacterium* sp. and express it in *P. agglomerans*.

The first step in cloning the 2,5-DKG reductase gene from the *Corynebacterium* sp. involved purifying the enzyme and determining the sequence of the first 40 amino acids from the N-terminal end of the molecule. On the basis of this amino acid sequence, two 43-nucleotide DNA hybridization probes, each corresponding to a different portion of the protein molecule, were synthesized. Because 71% of the nucleotides in the *Corynebacterium* sp. are either G or C, the probes were designed to include, where possible, a G or C in the third position of all codons, thereby minimizing the extent of

Table 8.1 Some of the compounds whose synthesis has been metabolically engineered in microbes

Type of compound	Compound
Acids	D-Glucuronic acid
	L-Ascorbic acid
	3-Hydroxypropionic acid
	Malic acid
	Muconic acid
	Lactic acid
	Succinic acid
	Itaconic acid
Alcohols	1-Propanol
	2-Butanol
	1-Butanol
	1,4-Butanediol
	1,2-Propanediol
	Ethanol
	Isopropanol
	Isobutanol
	2,3-Butanediol
	1,3-Propanediol
Amino acids	L-Alanine
	L-Arginine
	L-Valine
	L-Lysine
	L-Phenylalanine
	L-Threonine
	L-Tyrosine
	L-Tryptophan
	L-Cysteine
	L-Citrulline
Naturally occurring organic compounds	Monoterpene
	Sesquiterpene
	Diterpene
	Tetraterpene
	Flavonoids
	Amorpha-4,11-diene
	Artemisinic acid
	Indigo
	Lycopene
	β-Carotene
	Zeaxanthin
	Astaxanthin
	Taxadiene
	Geraniol
	Viridiflorol
Others	Biodiesel
	Resveratrol
	Vanillin
	Lipids
	Antibiotics
	Isoprene
	Heme
	Violacein
	Echinomycin
	Farnesene
	Biopolymers

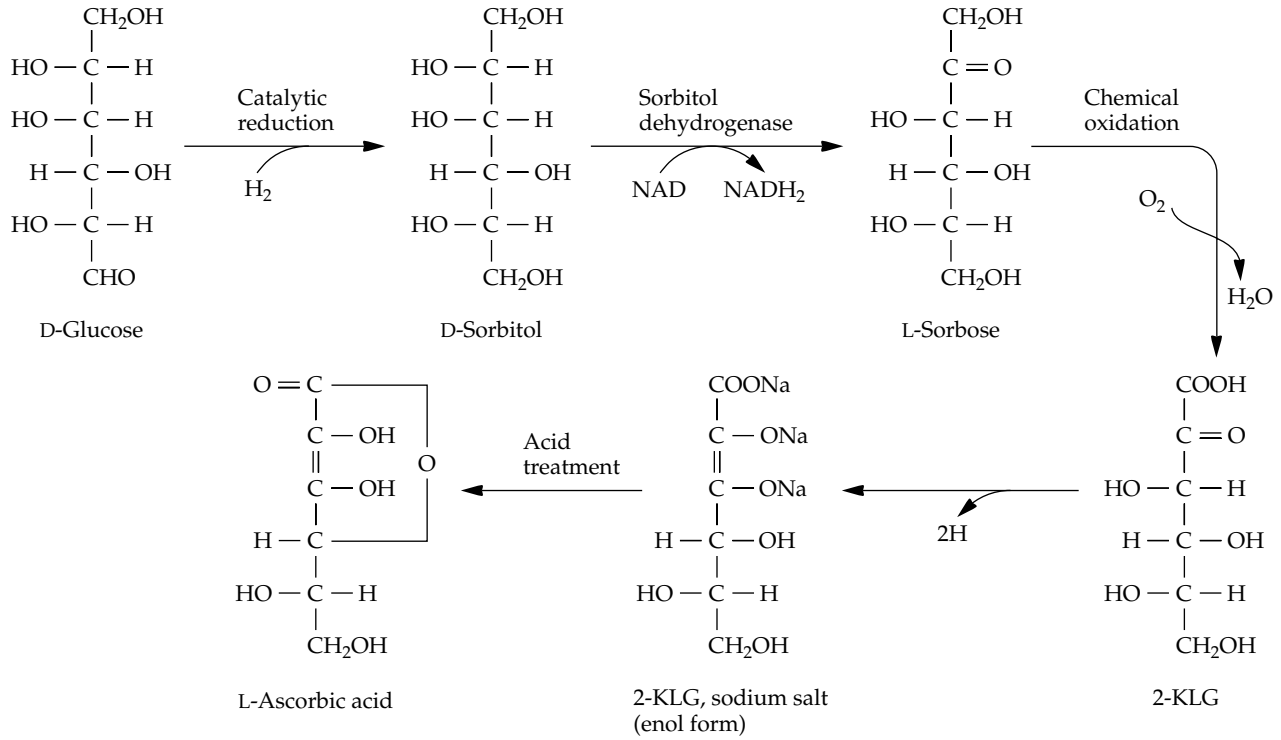


Figure 8.3 Commercial synthesis of L-ascorbic acid. Except for the microbial conversion of D-sorbitol to L-sorbose, all the steps are chemical reactions. The microbial conversion is carried out by *Acetobacter suboxydans*, which produces the enzyme sorbitol dehydrogenase.

the mismatch between the probe and the target DNA. This approach was taken because at the time that this work was done, mixed DNA probes were not readily available, and the DNA sequence of this gene was not known.

A *Corynebacterium* genomic DNA library constructed in *E. coli* was screened with these two probes. Any clones that hybridized with only one of the probes were discarded. It was assumed that any DNA that interacted with only one probe was probably not the target DNA. A clone that hybridized with both probes was isolated and then sequenced; it contained the 2,5-DKG reductase gene. The regulatory transcriptional sequences that were upstream of the ATG translation start signal and the protein coding sequence were deleted and replaced with transcriptional and translational regulatory signals that function in *E. coli*, because the regulatory sequences from Gram-positive microorganisms, such as *Corynebacterium* spp., are not efficiently utilized by Gram-negative *E. coli*. This construct expressed 2,5-DKG reductase activity in *E. coli* and subsequently was subcloned (including the regulatory sequences) onto a broad-host-range vector and introduced into Gram-negative *P. agglomerans*, which is able to use *E. coli* transcriptional and translational signals.

The transformed *Pantoea* cells converted D-glucose directly to 2-KLG. The endogenous *Pantoea* enzymes, localized in the inner membrane of the bacterium, converted glucose to 2,5-DKG, and the cloned 2,5-DKG reductase, produced from the cloned gene and localized in the cytoplasm, catalyzed the conversion of 2,5-DKG to 2-KLG (Fig. 8.4B). Thus, by genetic

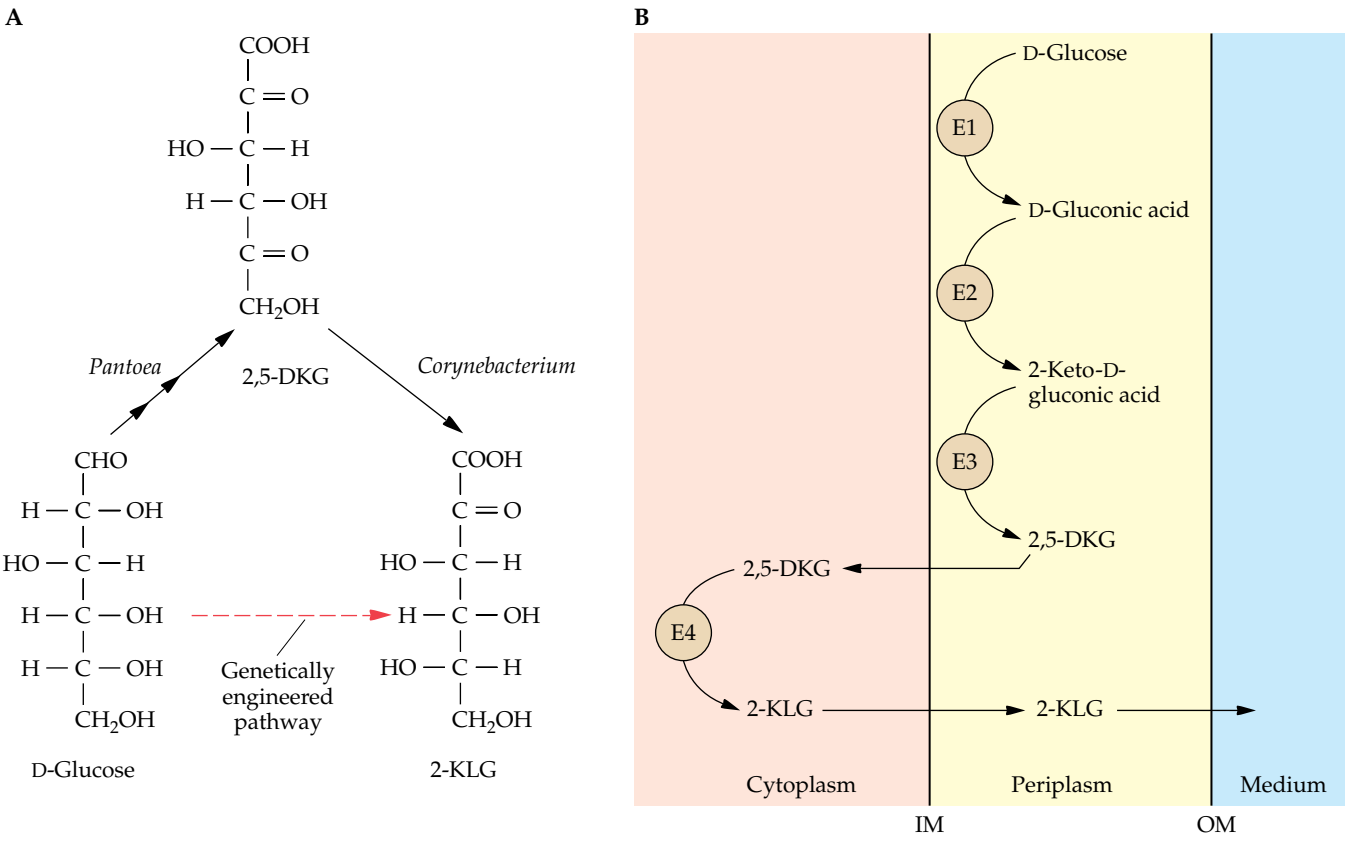


Figure 8.4 **(A)** Biological production of 2-KLG. *Pantoea* has a set of three enzymes that synthesize 2,5-DKG from D-glucose. *Corynebacterium* has an enzyme that converts 2,5-DKG to 2-KLG. Thus, 2-KLG, which is the immediate precursor of L-ascorbic acid, can be produced from D-glucose either by cocultivating these two microorganisms or by genetically engineering *Pantoea* to express the enzyme from *Corynebacterium*, which converts 2,5-DKG to 2-KLG. **(B)** Conversion of D-glucose to 2-KLG by recombinant *P. agglomerans*. The cellular locations of all of the component enzymes are indicated. The enzymes are denoted with the letter E and are numbered consecutively. Enzyme E4 is the introduced 2,5-DKG reductase. The major intermediates in the pathway are named. IM and OM denote the inner and outer membranes, respectively.

manipulation, the metabolic capabilities of two very dissimilar microorganisms were combined into one organism, which was able to produce the end product of the engineered metabolic pathway.

The reaction catalyzed by 2,5-DKG reductase utilizes reduced nicotinamide adenine dinucleotide phosphate (NADPH) as a cofactor. However, the cellular concentration of reduced nicotinamide adenine dinucleotide (NADH) is usually about 10-fold greater than that of NADPH, while the financial cost of NADPH is about 10 times higher than that of NADH. To lower the cost of the bacterial production of ascorbic acid, it would be beneficial to engineer a version of 2,5-DKG reductase that used NADH instead of NADPH. The only structural difference between NADH and NADPH is the presence or absence of a phosphate group attached to the 2′ site of the adenine moiety. From the three-dimensional structure of

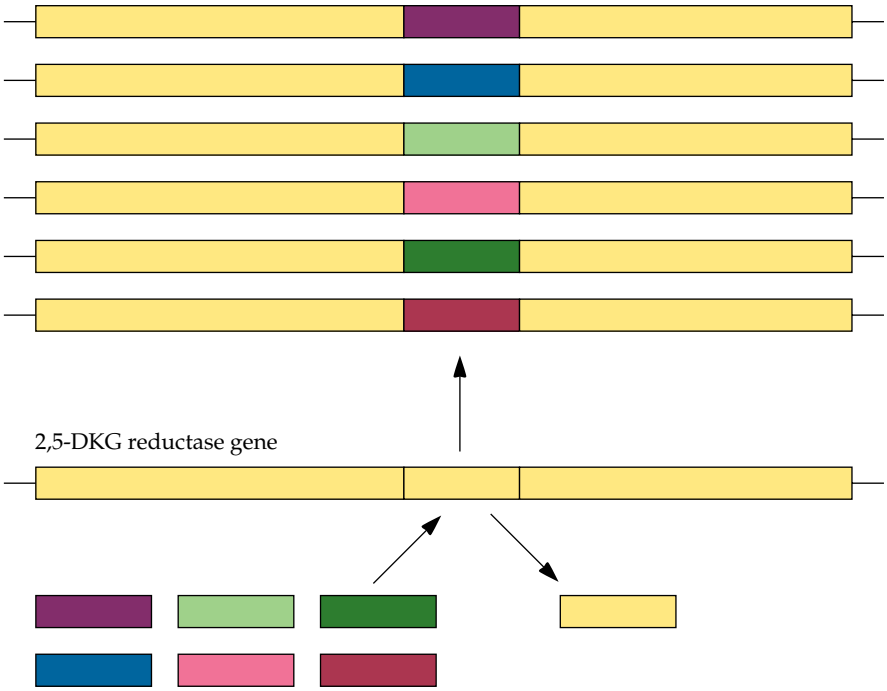


Figure 8.5 Mutagenesis of the *Corynebacterium* sp. gene encoding 2,5-DKG reductase gene to utilize NADH as a cofactor. The isolated gene, on a plasmid, was digested to remove a 57-bp DNA fragment. Then, chemically synthesized 57-bp DNA fragments (mutagenesis cassettes) containing coding sequences for one of the altered amino acids to be changed were spliced into the gene in place of the original fragment. In addition, each 57-bp fragment contained a silent mutation that abolished a PstI site but did not alter the amino acid sequence. The silent mutation facilitated screening against the wild type by digesting the gene with PstI.

2,5-DKG reductase complexed with NADPH, it appeared that 5 amino acid residues interact directly with the 2′ phosphate residue of NADPH. Using cassette mutagenesis (Fig. 8.5), a total of 40 different mutants of this enzyme were constructed; in each constructed mutant, 1 of the 5 amino acid residues that computer models suggested interacted with the 2′ phosphate residue was changed to a different amino acid. Following the expression, purification, and kinetic characterization of the 40 mutants in *E. coli*, it was observed that changing three of the five selected amino acids resulted in increases in 2,5-DKG reductase activity with NADH as the cofactor. In the best case, when the arginine residue at position 238 was changed to histidine, there was a 7-fold improvement in catalytic rate over the wild type with NADH as a cofactor. Moreover, after two amino acid alterations were combined in one protein, an enzyme that showed even more activity with NADH included a change of the lysine residue at position 232 to glycine, as well as the change of arginine at position 238 to histidine. Also, when the best NADH-active mutant was combined with a double mutant that increased the binding of the substrate, additional improvements in activity with NADH were observed. The activity of the enzyme isolated from the final construct was 72 times higher than the activity of the wild-type (NADPH) enzyme. In addition to this modification, researchers continue to genetically manipulate and improve the microbial pathways leading to the synthesis of vitamin C in a single organism.

Indigo

Indigo, a commercially important blue pigment that is used to dye both cotton and wool, was originally isolated from plants but is currently synthesized chemically. At present, approximately 18×10^6 kg of indigo is produced worldwide every year. Indigo, the coloring agent in blue jeans, is the largest-selling dye in the world. Originally, indigo was extracted from the leaves of the plant called *Indigofera tinctoria* (and several other plants as well as some sea shells) and is one of the oldest dyes known to humankind, having been used more than two thousand years ago to dye ancient Egyptian burial shrouds and strings on the corners of ritual Jewish prayer shawls (Numbers 15:38). More recently, indigo was used to make uniforms for Napoleon's army. Beginning in the late 19th century, scientists developed techniques for the chemical synthesis of indigo, largely replacing the plant-extracted dye.

Serendipitously, it was discovered that *E. coli* could be genetically engineered to produce indigo. Some *Pseudomonas* spp. possess the plasmid NAH7, which encodes enzymes for the degradation of the aromatic compound naphthalene, which enables them to grow on naphthalene as the sole carbon source. As the first step toward characterizing these genes, HindIII-digested fragments of NAH7 plasmid DNA were cloned into plasmid pBR322 and introduced into *E. coli* cells. All transformants were then tested by high-performance liquid chromatography (HPLC) for the production of nonvolatile metabolites that might result from the hydrolysis of naphthalene. Surprisingly, during the characterization of one of the transformants that had a 10.5-kb insert and could convert naphthalene to salicylic acid, it was observed that when the minimal growth medium contained tryptophan, it turned blue. A thorough analysis of the blue color revealed that the transformed *E. coli* cells were synthesizing the dye indigo. This synthesis was achieved in four steps (Fig. 8.6):

1. Conversion of tryptophan in the growth medium to indole by the enzyme tryptophanase, which is produced by the *E. coli* host cell
2. Oxidation of indole to *cis*-indole-2,3-dihydrodiol by naphthalene dioxygenase, which is encoded by the DNA that was cloned from the NAH7 plasmid
3. Spontaneous elimination of water
4. Air oxidation to form indigo

Thus, the combination of enzymes from two different pathways and two different organisms resulted in the synthesis of an unexpected compound, the dye indigo. Subsequent studies from another laboratory showed that introduction of the gene for the enzyme xylene oxidase, encoded in the *Pseudomonas* TOL plasmid (as part of a different biodegradative pathway), can convert tryptophan to indoxyl, which then spontaneously oxidizes to indigo (Fig. 8.6).

The ability to produce indigo from bacteria opens the possibility of developing an efficient and economical commercial microbial process for its production. This process would avoid the use of hazardous compounds, such as aniline, formaldehyde, and cyanide, which are needed in the chemical synthesis of indigo.

Despite the environmental benefits of engineering bacteria to make indigo, at the present time, the chemical synthesis of the dye is less expensive, thwarting commercial schemes to produce indigo biologically. One approach to improving indigo production in *E. coli* involves engineering the host strain to overproduce tryptophan, the raw material for the process.

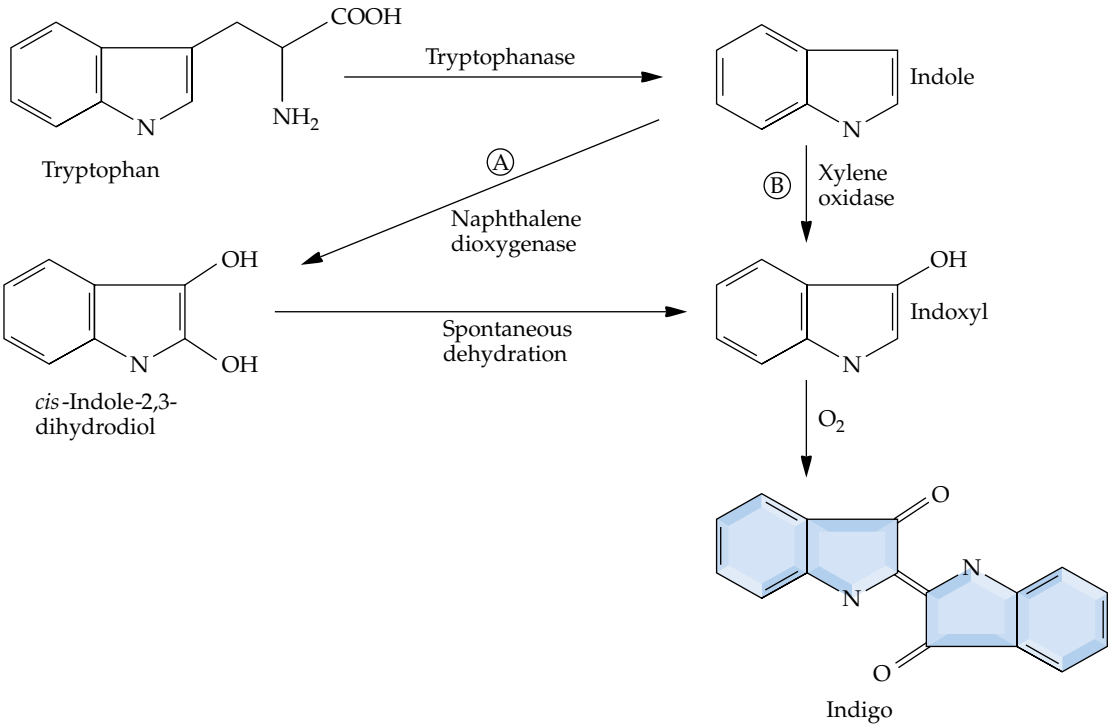


Figure 8.6 Indigo biosynthesis from tryptophan in genetically engineered *E. coli*. Tryptophanase is an *E. coli* enzyme. In pathway A, the naphthalene dioxygenase is derived from the NAH plasmid; in pathway B, the xylene oxidase is from the TOL plasmid. *E. coli* transformants that synthesize indigo contain either pathway A or B, but not both pathways.

However, despite improved efficiency when a tryptophan-overproducing *E. coli* strain is used, the overall process is not yet economically competitive with chemical synthesis. On the other hand, scientists have recently discovered a strain of *Pandoraea* sp. that can produce high levels of indigo when indole is added to the medium. These researchers are hopeful that this bacterial strain might provide the basis of a new process for the cost-effective synthesis of indigo.

Amino Acids

Amino acids are used extensively in the food industry as flavor enhancers, antioxidants, and nutritional supplements; in agriculture as feed additives; in medicine in infusion solutions for postoperative treatment; and in the chemical industry as starting materials for the manufacture of polymers and cosmetics (Table 8.2). It is estimated that ~3.5 million tons of amino acids, worth more than $25 billion (and predicted to reach ~$40 billion by 2025), were produced worldwide in 2019. L-Glutamic acid, which is used in the manufacture of the flavor enhancer monosodium glutamate, makes up nearly half of the total volume.

For the most part, amino acids are commercially produced either by extraction from protein hydrolysates or as fermentation products of either *Corynebacterium* or *Brevibacterium* spp., which are both nonsporulating Gram-positive soil bacteria that naturally secrete large amounts of amino

Table 8.2 Commercial applications of amino acids

Amino acid	Application(s)
Alanine	Flavor enhancer
Arginine	Therapy for liver diseases
Aspartic acid	Flavor enhancer; sweetener synthesis
Asparagine	Diuretic
Cysteine	Bread production; therapy for bronchitis; antioxidant
Glutamic acid	Flavor enhancer
Glutamine	Therapy for ulcers
Glycine	Sweetener synthesis
Histidine	Therapy for ulcers; antioxidant
Isoleucine	Intravenous solutions
Leucine	Intravenous solutions
Lysine	Feed additive; food additive
Methionine	Feed additive
Phenylalanine	Infusions; sweetener synthesis
Proline	Intravenous solutions
Serine	Cosmetics
Threonine	Feed additive
Tryptophan	Intravenous solutions; antioxidant
Tyrosine	Intravenous solutions; precursor for L-DOPA
Valine	Intravenous solutions

acids into the growth medium. Traditionally, the levels of amino acids produced by these bacteria have been improved by bacterial mutagenesis and subsequent screening for strains that overproduce particular amino acids. However, this traditional way of developing new strains is tedious, slow, and inefficient. On the other hand, it is more expeditious to isolate and manipulate the genes encoding the key components of a particular amino acid biosynthesis pathway, although this type of genetic engineering is not a simple matter. For example, the pathway(s) leading to the biosynthesis of certain amino acids contains a number of different enzymes, each of which may be either activated or inhibited by metabolites present in the cell. This often makes it difficult to know which enzyme(s) to manipulate in order to enhance the yield of the end product.

To rationally engineer a bacterium to overproduce a particular amino acid, it is essential to understand how the metabolic pathways of the bacterium are interrelated and regulated. Subsequent systematic reengineering of the metabolism of the organism can then be performed to achieve a much higher yield than would ever be possible by merely modifying one or two genes in the immediate biosynthetic pathway of a particular amino acid. At each stage in the development of the reengineered bacterium, it is possible to monitor the levels of a wide range of transcripts (using microarrays or RNA sequencing) and metabolites (metabolomics) and subsequently construct a detailed computer model that may be used to predict the effects of other possible genetic manipulations on the production of the target amino acid. While this approach is relatively new, such rational genetic engineering may be used to engineer bacterial strains for a variety of purposes, in addition to the commercial production of specific amino acids.

Because most known broad-host-range plasmid vectors replicate only in Gram-negative organisms, to express heterologous genes, it is necessary to

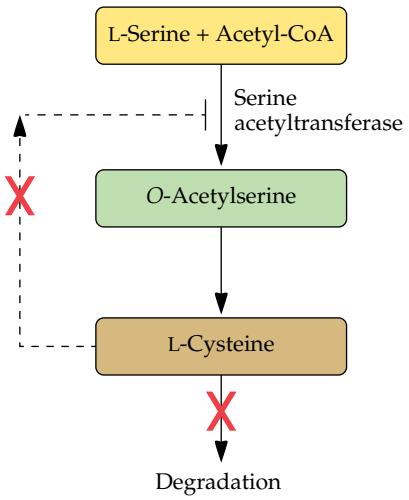
construct expression vectors that are specifically suited for *Corynebacterium* and *Brevibacterium* spp. Such cloning vehicles might take the form of *E. coli–Corynebacterium* shuttle vectors. Because *E. coli* and *Corynebacterium* spp. are both susceptible to the antibiotics tetracycline, chloramphenicol, and kanamycin, they may be used as selectable markers in both organisms. Effective transformation of *Corynebacterium glutamicum*, the species most often used in these experiments, is achieved when foreign DNA is introduced by conjugation or after the formation of protoplasts, i.e., after removal of the cell wall with lysozyme.

An alternative to producing amino acids in *Corynebacterium* and *Brevibacterium* spp. is to produce them in *E. coli*, where metabolic pathways and procedures for genetic manipulation are much more extensively characterized. Its relative ease of manipulation makes *E. coli* an attractive host organism for this sort of metabolic engineering.

L-Cysteine

L-Cysteine, which is one of the most important amino acids in the pharmaceutical, food, and cosmetics industries, has traditionally been obtained by extracting it from acid hydrolysates of human hair and animal feathers. While most microorganisms are able to synthesize L-cysteine, high levels cannot be synthesized from glucose because L-cysteine feedback inhibits the enzyme serine acetyltransferase, which catalyzes one of the key steps in the biosynthesis of L-cysteine (Fig. 8.7). To try to remedy this situation, the methionine residue at position 256 in the *E. coli* serine acetyltransferase amino acid sequence was systematically changed to each of the other 19 amino acids. When *E. coli* was transformed with plasmids carrying *cysE* genes encoding altered serine acetyltransferase, several transformants with altered forms of serine acetyltransferase produced higher levels of L-cysteine than did the wild-type enzyme. Next, plasmids encoding the most effective serine acetyltransferase derivatives were used to transform a mutant strain of *E. coli* that did not degrade L-cysteine, resulting in modest success in overproducing L-cysteine. To improve on this, complementary DNAs (cDNAs) encoding feedback inhibition-insensitive serine acetyltransferases from the plant *Arabidopsis thaliana* were expressed in a serine acetyltransferase-deficient *E. coli* strain that did not utilize L-cysteine. The transformants produced a much higher level of L-cysteine than had previously been possible by merely manipulating the *E. coli* serine acetyltransferase gene. Whether the expression of this protein can be further improved and how this affects the synthesis of L-cysteine remain to be determined.

Figure 8.7 Biosynthesis and overproduction of L-cysteine from L-serine and acetyl-CoA in *E. coli*. The dashed arrow indicates feedback inhibition. The red X leading to the serine acetyltransferase indicates that this enzyme was insensitive to feedback inhibition. The red X from L-cysteine to degradation indicates that, in the appropriate strain of *E. coli*, L-cysteine was not degraded.



L-Valine

To rationally engineer *E. coli* to overproduce L-valine, it was necessary to introduce a large number of mutations into the *E. coli* genome (Fig. 8.8). The flux from glucose to L-valine was first improved by abolishing feedback inhibition by L-valine of the enzyme that converts pyruvate to 2-acetolactate. Then, the subunits of this enzyme were overexpressed by replacing the endogenous promoter with a strong constitutive promoter. Next, the carbon flux toward L-valine was increased. The gene for the enzyme that converts L-threonine to 2-ketobutyrate was knocked out, as were the genes encoding enzymes that convert 2-ketoisovalerate to either pantothenate or L-leucine. In addition, some of the genes that encode enzymes that convert pyruvate to 2-acetolactate were amplified. Subsequently, the expression levels of the genes encoding enzymes

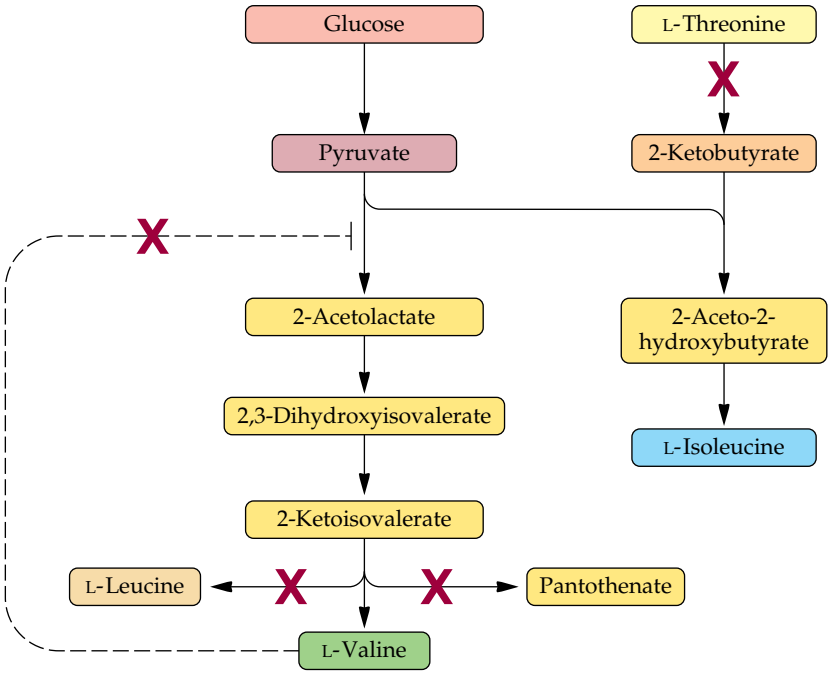


Figure 8.8 Simplified overview of the production of L-valine in rationally engineered *E. coli*. The red Xs indicate pathways that were knocked out, and the dashed line indicates feedback inhibition. Some arrows represent a single enzymatic step, while other arrows represent several enzymatic steps. The interaction of glucose with other pathways, the export of L-valine from the cell, and several more complex regulatory steps are not shown.

that convert 2-acetolactate to L-valine via 2,3-dihydroxyisovalerate and 2-keto-isovalerate were all increased by modifying the regulatory regions of these genes. In addition, the number of copies of the *lrp* gene, encoding the leucine-responsive regulatory protein, was increased, since some genes in the L-valine pathway were under the positive transcriptional control of this protein. Finally, the expression of the two genes that encode proteins responsible for the export of L-valine from the bacterial cell was increased, since it was thought that a low level of these proteins was limiting to L-valine production. Following the extensive metabolic engineering of *E. coli*, the final strain was able to produce 0.378 g of L-valine per g of glucose, which is higher than industrial strains of *C. glutamicum* that have been developed by repeated rounds of random mutagenesis and selection. Moreover, researchers believe that the rationally engineered strain can be further modified with resulting increases in L-valine productivity.

L-Citrulline

It is estimated that the global market for citrulline will be worth ~$40 billion by 2023. The nonprotein amino acid L-citrulline is a naturally occurring precursor of L-arginine (Fig. 8.9); it is sold over the counter in pharmacies and health food stores and is used for a variety of ailments. These include elevated blood pressure, mild erectile dysfunction, Parkinson's disease, certain types of dementia, diabetes, fatigue, and muscle weakness. Significant amounts of L-citrulline are found naturally in some

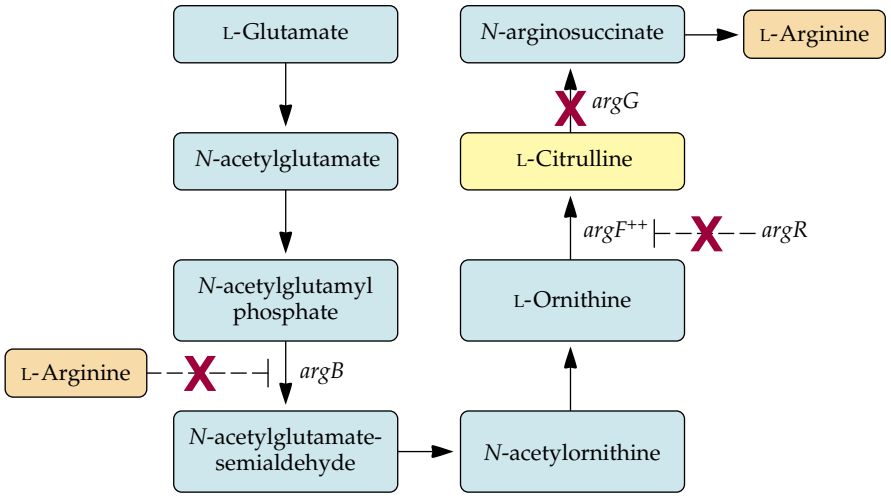


Figure 8.9 Modifying the biosynthesis of L-citrulline, an intermediate in the biosynthesis of L-arginine. The potential synthesis of arginine was derepressed by deleting the arginine repressor gene *argR*; the *argG* gene was deleted so that citrulline could not be converted to *N*-arginosuccinate; feedback inhibition of *N*-acetyl L-glutamate kinase by arginine was prevented by substituting a different version of the *argB* gene, and the product of the *argF* gene, L-ornithine carbamoylphosphate transferase, was overexpressed (*argF*++) by deletion of *argR*.

foods, most notably watermelon, and it is also produced by humans. When dietary L-citrulline is digested, it is converted to L-arginine in the kidneys and then released into the bloodstream. L-Citrulline is currently produced on an industrial scale by the fermentation of strains of either *Pseudomonas putida* or *Bacillus subtilis*. In an effort to find a more efficacious source of L-citrulline, scientists have genetically engineered a strain of *C. glutamicum*.

To engineer *C. glutamicum* to synthesize an increased amount of L-citrulline, several modifications were made to the wild-type strain (Fig. 8.9): (i) The potential synthesis of L-arginine from L-citrulline was increased by the deletion of the arginine repressor gene *argR*. This increases the metabolic flux in the arginine biosynthesis pathway. (ii) The *argG* gene was deleted so that L-citrulline could not be converted to *N*-arginosuccinate (the immediate precursor to L-arginine). (iii) An *argB* gene encoding a feedback-resistant *N*-acetyl L-glutamate kinase was introduced into the *C. glutamicum* genome in place of the wild-type *argB* gene. (iv) The *argF* gene encoding L-ornithine carbamoylphosphate transferase was overexpressed.

The engineered *C. glutamicum* strain was able to produce high levels of L-citrulline from glucose. In addition, this organism could also efficiently produce L-citrulline using starch, xylose, and glucosamine as a carbon source.

Modified *Corynebacterium*

The majority of bacterial genomes contain one or more prophages (a complete or partial bacteriophage genome inserted into the bacterial chromosomal DNA). Prophages, along with transposable elements, are considered to be major sources of bacterial genomic instability in that they can (i) express unneeded and unwanted genes, (ii) promote horizontal gene transfer, and (iii) cause the loss or inactivation of some bacterial genes. The presence of prophages in an industrial bacterial strain such as a strain

of *C. glutamicum* that is used for the industrial production of amino acids may have a significant negative impact on the industrial performance of that strain. For example, one industrial strain of *C. glutamicum* contains three separate prophages that are 3.9, 13.5, and 187.3 kb in size, respectively (Fig. 8.10). During normal growth of the bacterium, the majority of the prophage-encoded genes are either not expressed or else expressed at only a very low level. However, it is well known that expression of prophage genes is often induced under stressful conditions. When the wild-type bacterium was grown under various stress conditions (pH stress, oxidative stress, phosphate limitation, iron starvation, osmotic stress, and heat shock), the expression of several proteins (notably a restriction-modification system) encoded within the largest prophage was upregulated. To prevent this, one group of researchers deleted the three prophages from the genome of *C. glutamicum* one at a time by homologous recombination. Following deletion of the three prophages, the resultant bacterium had a genome that was ~6% smaller and, when grown under stressful conditions, grew approximately 15 to 20% faster than the wild-type strain. The *C. glutamicum* strain with the reduced genome could be transformed with a much higher efficiency and significantly increased the production of recombinant proteins compared to the wild type. The modified strain provides a more efficient means of producing both foreign proteins and metabolites such as amino acids.

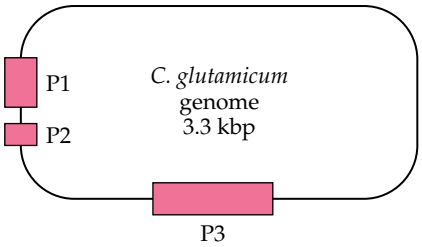


Figure 8.10 The genome of a strain of *Corynebacterium glutamicum* containing three prophages (not drawn to scale). The prophages are labeled P1 to P3 for prophages 1 to 3.

Lycopene

Lycopene is a bright red carotenoid pigment (Fig. 8.11A) that is commonly found in tomatoes and other fruits, including watermelons, pink grapefruit, pink guavas, papayas, and rose hips. It is a powerful antioxidant that has been suggested to decrease low-density lipoprotein oxidation in humans and thereby to lower the risk of atherosclerosis and coronary heart disease. In addition, lycopene, and several carotenoids derived from lycopene, have been proposed as treatments for some types of cancer. It would be useful if lycopene (and carotenoids produced from lycopene) could be produced in microorganisms so that the large-scale-processing problems that exist when lycopene is isolated from tomatoes might be avoided. It is estimated that the global market for lycopene will be ~$135 million in 2023.

For the production of lycopene in *E. coli*, the 2-*C*-methyl-D-erythritol 4-phosphate pathway provides the precursors isopentenyl diphosphate and dimethylallyl diphosphate (Fig. 8.11B). However, by introducing *Saccharomyces cerevisiae* genes encoding the mevalonate pathway under the control of *E. coli* transcriptional promoters, the levels of these precursor molecules are increased dramatically. By subsequently introducing *Pantoea agglomerans* (a Gram-negative bacterium) genes encoding the biosynthesis of lycopene from the above-mentioned precursor metabolites, a relatively high level of lycopene can be produced (Fig. 8.11B). With this engineered *E. coli* strain, containing the two introduced biosynthetic pathways, it was possible to obtain approximately 60 mg of lycopene per liter of bacterial culture. While additional optimization of this system is still necessary before it can be the basis of a commercial system for lycopene production, this work is an important step in the development of such a system. More recently, scientists have engineered the yeast strain *Yarrowia lipolytica* to produce comparable levels of lycopene to what is described above. In this

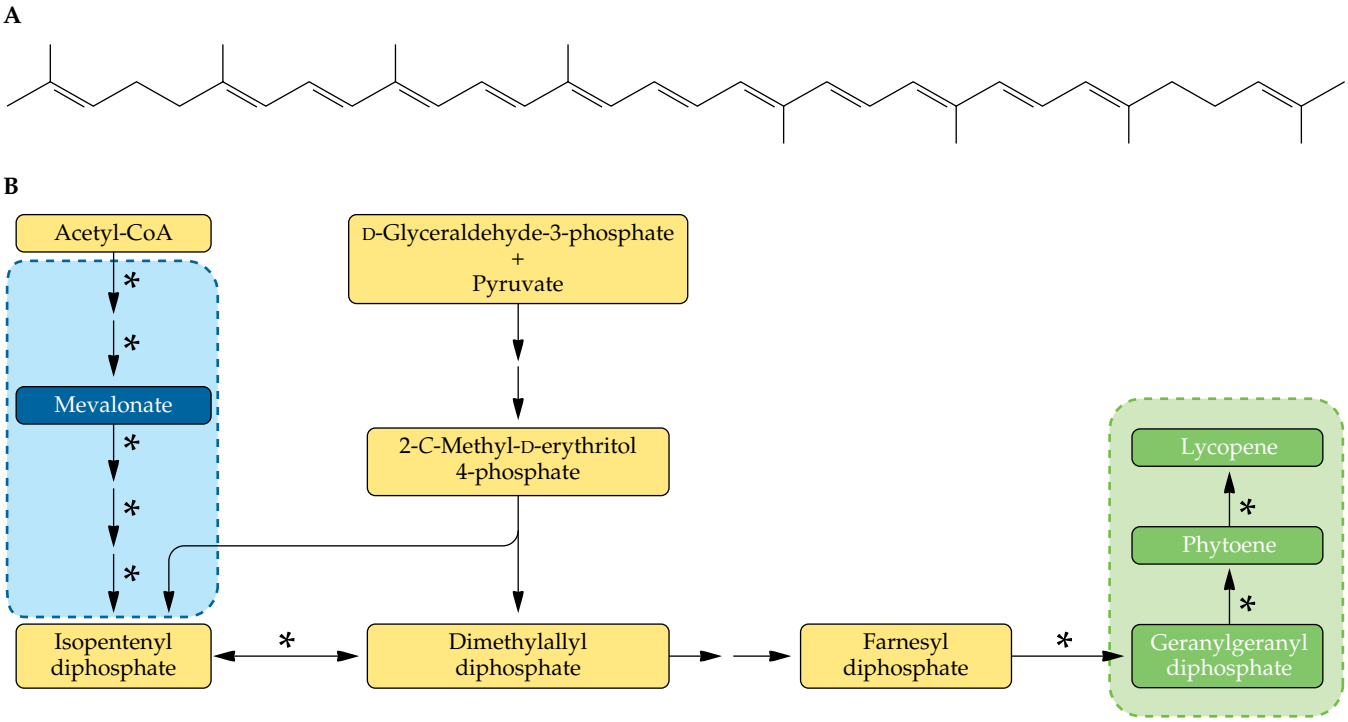


Figure 8.11 **(A)** The chemical structure of lycopene. **(B)** Overview of the production of lycopene in an engineered strain of *E. coli*. The yellow metabolites are normally produced in *E. coli*. The portion of the pathway highlighted in blue represents the enzymes encoded by *S. cerevisiae* genes, as well as some of the metabolites that they produce. The pathway highlighted in green represents enzymes encoded by *P. agglomerans* and the metabolites that they produce. Introduced enzymes are depicted, adjacent to the arrows, by asterisks.

case, genes that were not expressed well in the yeast host were resynthesized using yeast codons in order to increase their expression levels.

Antibiotics

Since the discovery of penicillin in 1928, more than 12,000 antibiotics with different specificities and a variety of modes of action have been isolated from various microorganisms. The universal use of antibiotics to treat bacterial infections has resulted in an enormous improvement in human health and has undoubtedly saved millions of lives. The majority of the most important antibiotics have been isolated from the Gram-positive soil bacterium *Streptomyces*, although fungi and other Gram-positive and Gram-negative bacteria are also sources of antibiotics. Worldwide, over 125,000 tons of antibiotics are produced per year, with annual gross sales of about $85 billion in 2020, including antibiotics used in animal feed as animal growth promoters. The antibiotic market is driven by the sales of four leading drug classes: the cephalosporins (27%), macrolides (20%), quinolones (17%), and penicillins (17%).

An estimated 200 to 300 new antibiotics are discovered each year, primarily through labor-intensive research programs in which thousands of

different microorganisms are screened to find those that produce unique antibiotics. However, with the high costs of development and clinical testing, only the compounds that show significant therapeutic and economic promise are marketed. Therefore, only about 1 to 2% of newly discovered antibiotics are added annually to the disease-fighting arsenal. In fact, the pharmaceutical industry has been reluctant to invest in research and development in this area, and some companies have either abandoned or scaled down their efforts since 1999. In addition, to date, nearly all of the genetic improvements to industrially important antibiotic-producing strains have been achieved by the use of classical mutagenesis and selection. While the yields of antibiotics from many strains have been significantly improved—the original penicillin-producing fungus isolated by Alexander Fleming yielded 2 units per milliliter of culture, while the strains used today synthesize approximately 70,000 units per milliliter of culture—this yield improvement takes many years and requires the use of considerable human and financial resources. Recent studies have attributed much of the increase in penicillin yield that resulted from multiple rounds of conventional mutagenesis and selection to increases in the number of copies of the genes that encode the synthesis of this antibiotic. Recombinant DNA technology can have a positive impact on this endeavor in two ways. First, the technology can be used to develop new, structurally unique antibiotics with increased activities against selected targets and decreased side effects. Second, genetic manipulation can be used to rapidly and inexpensively enhance yields and hence lower the cost of production of existing antibiotics.

For the genetic manipulation of *Streptomyces*, it is essential that this bacterium can be transformed and that the transformed cells can be readily selected. However, unlike *E. coli*, *Streptomyces* strains exist not as individual cells, but as extended aggregates called mycelial filaments. The cell wall must be removed to release individual cells (protoplasts) before DNA transformation (Fig. 8.12). Without this step, it would not be possible to distinguish transformed from nontransformed cells, because visible colonies on a solid medium would each have started from a cell aggregate rather than from an individual cell. Thus, colonies that grew in the presence of a selective antibiotic would contain a mixture of transformed and nontransformed cells. However, as a consequence of protoplast formation prior to transformation, all colonies that grow in the presence of a selective antibiotic contain only transformed cells. The uptake of plasmid DNA into *Streptomyces* protoplasts is enhanced by the addition of polyethylene glycol, which slightly disrupts the cell membrane. Following transformation, the protoplasts are first plated onto a solid medium to enable the bacterial cell walls to regenerate and are then overlaid with a selective medium that often contains either neomycin or thiostrepton, both of which act as selection agents for transformed cells.

Antibiotic Biosynthesis Genes

The biosynthesis of an antibiotic typically includes 10 to 30 separate enzyme-catalyzed steps, so cloning all the genes for the synthesis of a particular antibiotic is not an easy task. One early strategy for isolating the complete set of antibiotic biosynthesis genes from a bacterium whose genome has not been sequenced consists of first transforming a strain of *Streptomyces*, which has been mutagenized so that it is no longer able to synthesize an antibiotic, with a genomic DNA library prepared from the

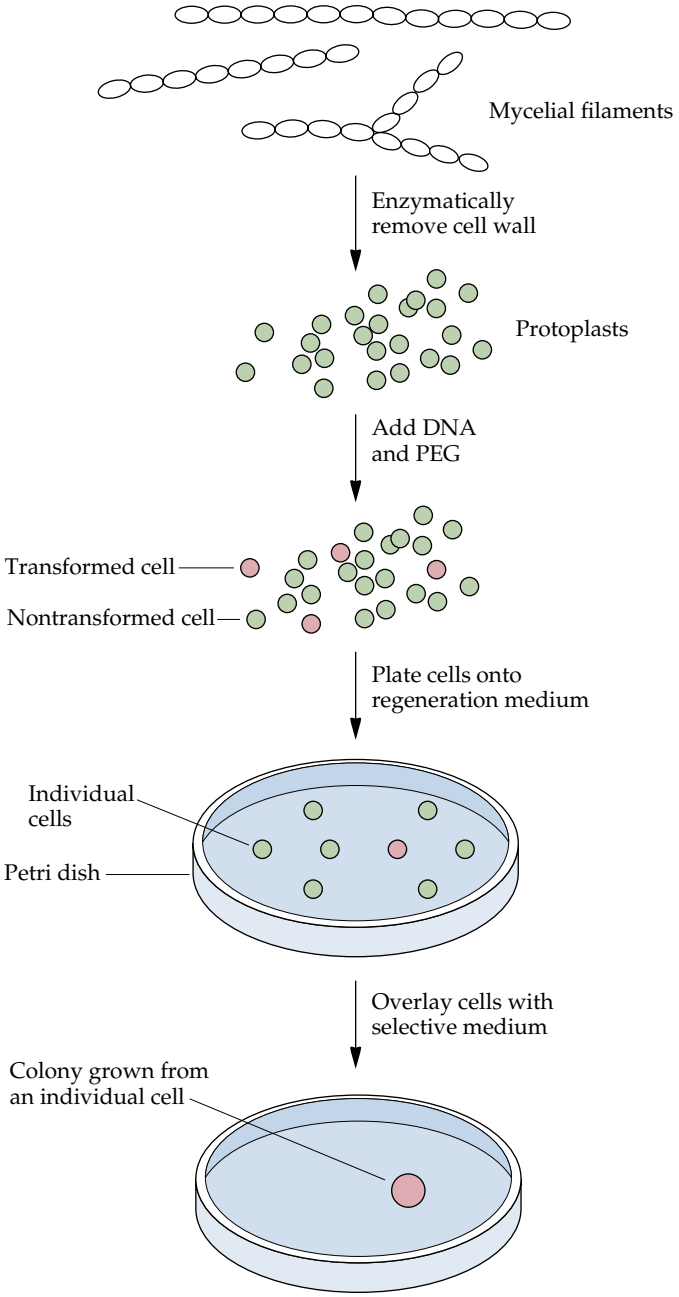


Figure 8.12 Schematic representation of DNA transformation and selection of transformants of *Streptomyces* strains. The pink circles represent transformed cells, and the green circles represent nontransformed cells. PEG, polyethylene glycol.

antibiotic-producing wild-type strain of the same organism. Following the introduction of the DNA library into mutant cells, transformants are screened for the ability to produce the antibiotic. This process is facilitated if the antibiotic is colored and imparts this color to antibiotic-producing cells. Then, the plasmid DNA from the clone that supplies a functional gene and gene product, i.e., complements the mutant strain, is used as a DNA hybridization probe to screen another clone bank of wild-type

chromosomal DNA (i.e., one in which the average-size fragment is around 10 kb) to isolate clones with regions that overlap the probe sequence. In this way, DNA segments that are adjacent to and usually bigger than the initial complementing DNA can be identified and cloned. A complete gene cluster can be reconstructed from the overlapping clones. The process of identifying adjacent genes is sometimes called walking down the chromosome. Generally, the antibiotic biosynthesis genes are clustered at a single site on the chromosomal DNA so that the genes that are adjacent to the complementing gene are also likely to be involved in the biosynthesis of the target antibiotic.

The complementation approach has been used to isolate the genes for the biosynthesis of the red-colored antibiotic undecylprodigiosin from *Streptomyces coelicolor* A3 (Fig. 8.13A). In this case, the complementation assay is simple and entails assessing the color of the colonies. Colonies of wild-type organisms are red (Fig. 8.13B) because of the presence of the antibiotic, and mutant colonies are cream colored. Complementation of the mutant produces a red colony (Fig. 8.13B).

In addition to cloning antibiotic biosynthesis genes by complementation, more direct strategies can be employed, including knowledge of the partial or complete sequence of the genome of the antibiotic-producing bacterium. One or more of the key enzymes in a biosynthetic pathway can be identified through either genetic or biochemical studies and then purified. The N-terminal amino acid sequence of the enzyme can then be determined, and with this information, oligonucleotide probes for the gene can be prepared.

Despite the difficulty, there are a number of examples of the cloning and transfer of large fragments of DNA encoding entire antibiotic biosynthetic pathways. In these cases, it is usually necessary to use a vector that can accept and maintain pieces of DNA as large as 100 kb. For this purpose, researchers have employed bacterial artificial chromosomes that have been engineered to replicate autonomously in *E. coli* and, when they are introduced into *Streptomyces*, to integrate into the chromosome.

Engineering Polyketide Antibiotics

The term "polyketide" defines a class of antibiotics that are synthesized through the successive enzymatic condensation of small carboxylic acids, such as acetate, propionate, and butyrate. Polyketide drugs include the antibiotic erythromycin, the immunosuppressive drug FK506, and the cholesterol-lowering drug lovastatin. While some polyketide antibiotics are produced by plants or fungi, the majority are produced by actinomycetes as secondary metabolites. To create new polyketide antibiotics with different properties, the functioning of the enzymes that synthesize these antibiotics must be understood before the genes encoding the enzymes can be manipulated.

Polyketide synthesis consists of a number of enzymatic steps that are each repeated several times, including ketoreduction, dehydration, and enoylreduction of the β-group of the growing polyketide chain. Each condensation cycle results in the formation of a β-keto group on a growing carbon chain (Fig. 8.14). Polyketide synthesis is analogous to the mechanisms used for the synthesis of long-chain fatty acids. The complete synthesis of a polyketide antibiotic generally requires the participation of several of these multifunctional enzymes that make up the subunits of the polyketide synthase.

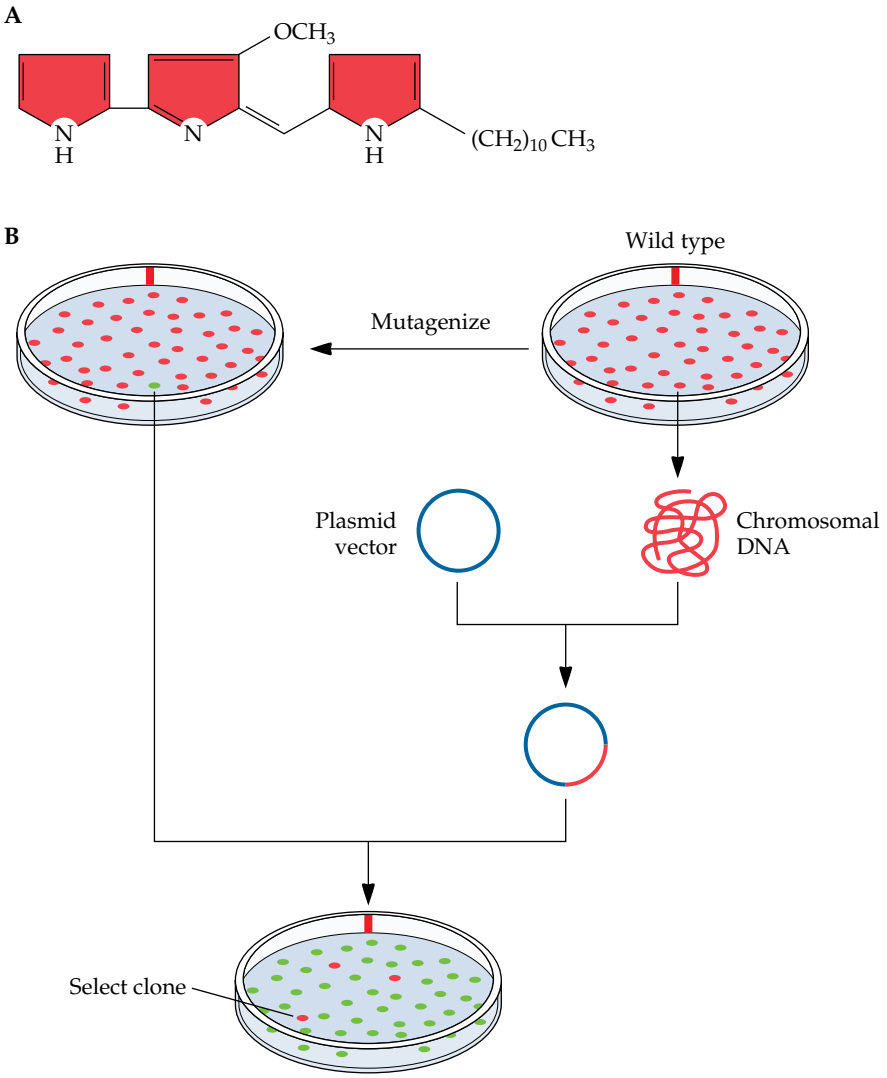


Figure 8.13 (A) Chemical structure of the red-colored antibiotic undecylprodigiosin. **(B)** Method for cloning genes involved in the biosynthesis of the antibiotic undecylprodigiosin. Chromosomal DNA from wild-type antibiotic-producing cells is spliced into a *Streptomyces* cloning vector. The clone bank is used to transform a noncolored (i.e., non-antibiotic-producing) mutant of the wild type. Red transformants (in which the mutant has been complemented) are selected, and the plasmid DNA insert is characterized.

The gene clusters for synthesis of aromatic polyketides all contain a set of three genes encoding a so-called minimal polyketide synthase (Fig. 8.15). Each minimal polyketide synthase contains the activities for one cycle of polyketide chain elongation. The minimal module has a ketosynthase (with an acyltransferase domain), a chain length factor, and an acyl carrier protein. The minimal polyketide synthase is responsible for the synthesis of the aromatic polyketide backbone. Modifications to the basic structure are catalyzed by other enzymes acting in concert with the minimal polyketide synthase. The order of the modules in a polyketide synthase specifies the sequence of the distinct two-carbon units, and the

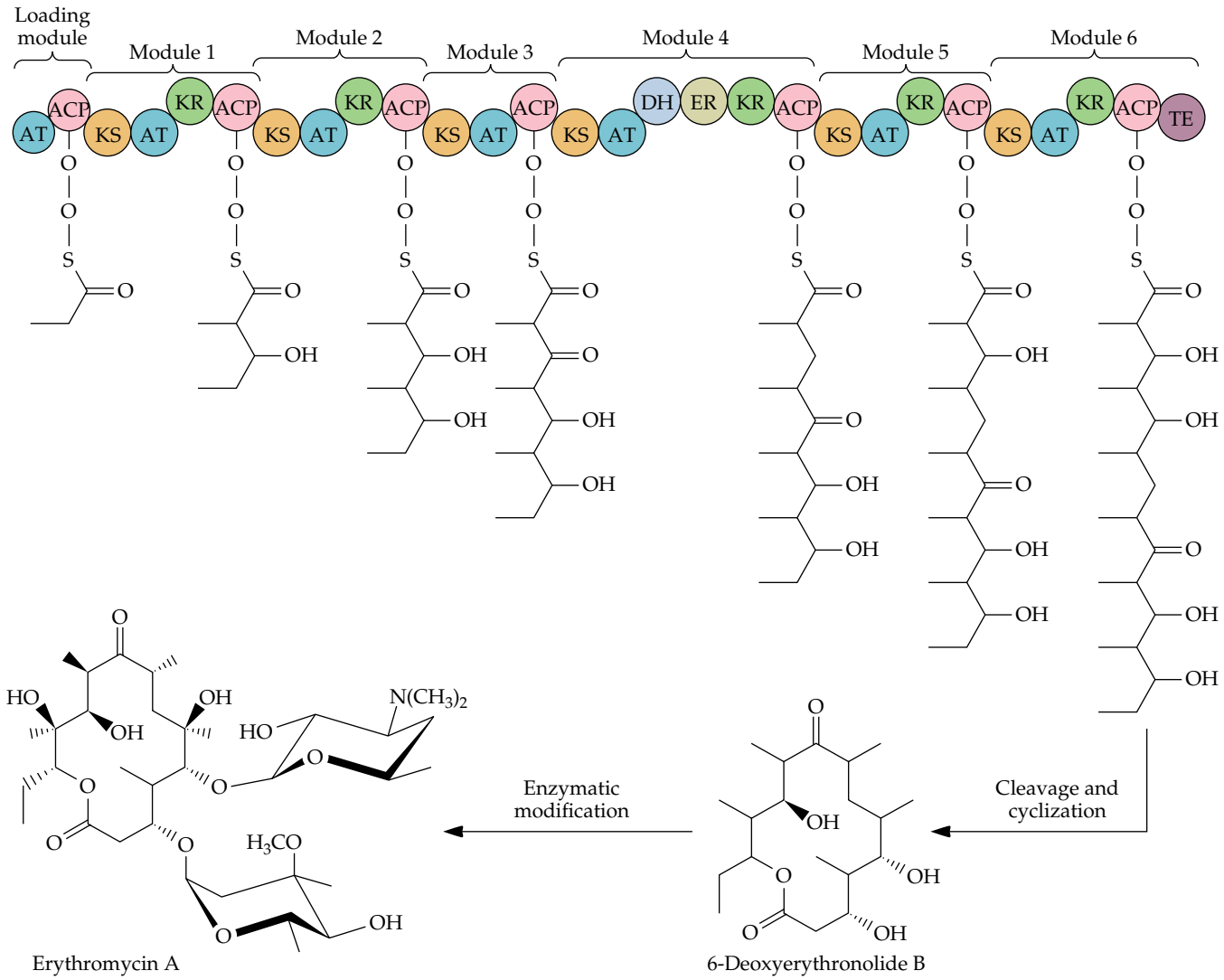


Figure 8.14 Assembly line biosynthesis of the antibiotic erythromycin A. Each biosynthetic module is responsible for adding one acyl unit onto the precursor molecule. Each of the assembly line modules consists of functional enzymatic domains (with each one having a different function) that provide the required activity for polyketide biosynthesis. For the synthesis of erythromycin, the acyltransferase domain (AT) loads a propionyl-CoA unit onto an acyl carrier protein (ACP). The ATs of modules 1 to 6 are specific for methylmalonyl-CoA, which they load onto ACPs. The ketosynthase (KS) domains catalyze the condensation of the different acyl units. In some cases, a module contains additional domains such as ketoreductase (KR), dehydratase (DH), or enoyl reductase (ER), so that the growing polypeptide may be derivatized while it is still tethered to an ACP. Module 6 contains a terminal thioesterase (TE) that cleaves the growing molecule from the last ACP and cyclizes it before it is enzymatically modified to erythromycin. The protein modules are not drawn to scale.

number of modules determines the size of the polyketide chain. The genes encoding a complete set of these proteins for the synthesis of a particular polyketide are generally organized into a single cluster (Fig. 8.15). Each minimal polyketide synthase gene cluster encodes the synthesis of a

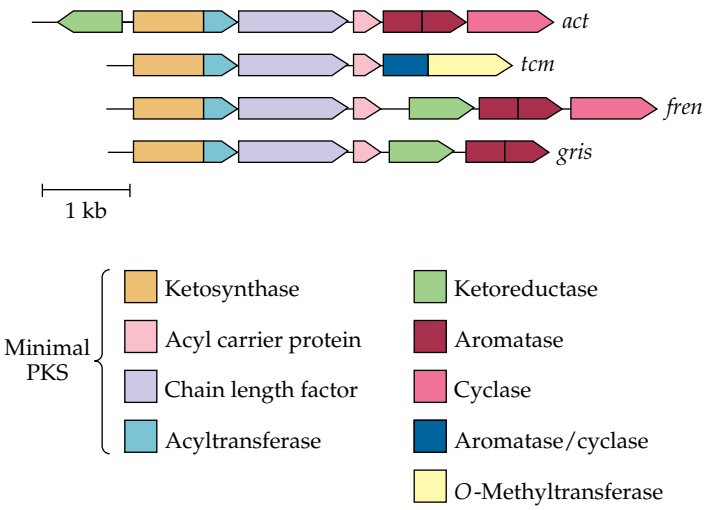


Figure 8.15 Gene clusters for the biosynthesis of the aromatic polyketide antibiotics actinorhodin (*act*), tetracenomycin (*tcm*), frenolicin (*fren*), and griseusin (*gris*). Each cluster contains genes encoding a minimal polyketide synthase (PKS), which is responsible for the synthesis of the polyketide backbone. The enzymes encoded by the other genes act to modify the growing polyketide chain. Each gene is shown pointed in the direction in which it is transcribed.

particular antibiotic. By interchanging genes for different polyketides between clusters, new aromatic polyketide antibiotics may be created, an approach that promises to dramatically accelerate the process of discovery of new antibiotics.

If each of the enzymatic reactions that is catalyzed by a domain on a multifunctional polyketide synthase subunit catalyzes only a single biochemical step in the pathway, the loss of any one activity should affect only a single step in the overall synthesis. Moreover, alteration of a catalytic domain whose function has been established should allow researchers to make predictable changes to the structure of the synthesized antibiotic. For example, a detailed knowledge of the genetics and biochemistry of the components involved in the synthesis of the antibiotic erythromycin allowed researchers to alter the biosynthetic genes in a predetermined manner and to produce predictably altered derivatives of erythromycin. Erythromycin is normally synthesized by the actinomycete *Saccharopolyspora erythraea*, and the entire 56-kb DNA segment that contains the *ery* gene cluster has been sequenced. The erythromycin polyketide synthase was altered in two different ways: either (i) a DNA region that encoded β-ketoreductase activity was deleted or (ii) a DNA region encoding enoylreductase was mutated. With the β-ketoreductase deletion, the erythromycin intermediates that accumulated had a carbonyl moiety rather than a hydroxyl group at the C-5 carbon of the ring (Fig. 8.16). With the enoylreductase mutation, a carbon–carbon double bond was introduced at positions C-6 and C-7 of the ring (Fig. 8.16). These experiments indicate that once the cluster of genes encoding the biosynthesis of a particular polyketide antibiotic has been isolated and characterized, it is possible, by altering specific DNA fragments, to modify an antibiotic biosynthesis pathway and thereby to alter the structure of the antibiotic in a predictable manner to create novel polyketide antibiotics.

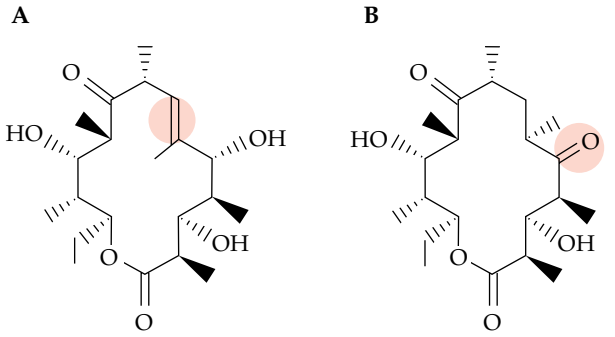


Figure 8.16 Altered erythromycin derivatives produced through genetic manipulation. **(A)** A mutation in an enoylreductase gene caused a carbon–carbon double bond to be introduced at positions C-6 and C-7 of the ring (highlighted). **(B)** A deletion in a β-ketoreductase gene caused the erythromycin to have a carbonyl moiety rather than a hydroxyl group at the C-5 carbon of the ring (highlighted). Data from Katz and Donadio, *Annu. Rev. Microbiol*. **47:**875–912, 1993.

Unique Antibiotics

In recent years, there has been an enormous proliferation in the prevalence of antibiotic-resistant bacterial pathogens. At present in the United States, more deaths are attributable to infections by methicillin-resistant strains of *Staphylococcus aureus* than to human immunodeficiency virus/AIDS. An important component of *S. aureus* virulence is the carotenoid pigment staphyloxanthin, which is synthesized by the bacterium. As a consequence of the large number of conjugated double bonds that this compound possesses, it can detoxify the reactive oxygen species (including peroxides, superoxide, hydroxyl radical, and singlet oxygen) that are often produced by the host immune system in response to the bacterial infection. On the other hand, strains of *S. aureus* that do not contain staphyloxanthin are rapidly inactivated by the reactive oxygen species produced by host neutrophils (a type of white blood cell). This observation suggests that the disruption of *S. aureus* staphyloxanthin synthesis might be a suitable target to prevent the proliferation and toxicity of the bacterium.

In the first committed step in the biosynthesis of *S. aureus* staphyloxanthin, the enzyme dehydrosqualene synthase condenses two molecules of farnesyl diphosphate to produce presqualene diphosphate (Fig. 8.17). In *S. aureus*, this compound is modified to yield dehydrosqualene, which is converted into 4,4′-diaponeurosporene and eventually into staphyloxanthin. Interestingly, the synthesis of cholesterol in humans also proceeds through presqualene diphosphate (Fig. 8.17), and the human enzyme squalene synthase is structurally similar to bacterial dehydrosqualene synthase. In order to determine whether the *S. aureus* enzyme dehydrosqualene synthase that catalyzes the first committed step could be a target for a designer antibiotic, the bacterial gene was cloned and overexpressed, the enzyme was purified to homogeneity, and its X-ray crystallographic structure was determined to 1.58-Å resolution. Based on the three-dimensional structure of the enzyme, as well as the presence of inhibitors of the human squalene synthase (from the cholesterol synthesis pathway) that performed the same function, a series of eight chemical inhibitors was designed, synthesized chemically, and tested. Three of the inhibitors, when tested at levels below 1 μM, dramatically blocked the conversion of farnesyl diphosphate

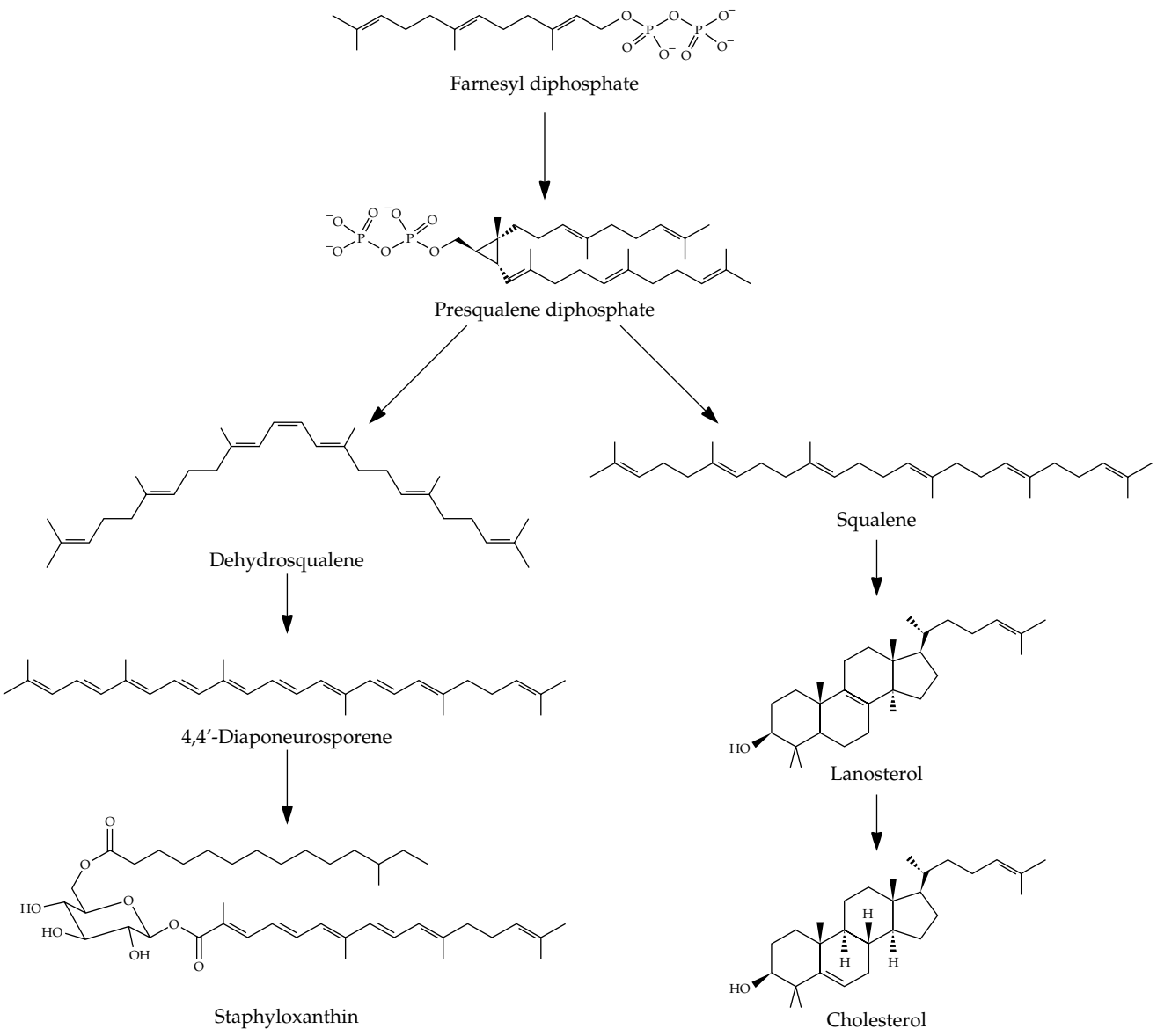


Figure 8.17 Simplified overview of the biosynthesis of staphyloxanthin in *S. aureus* and cholesterol in humans. The first step (the committed step) is the same in both pathways.

to presqualene diphosphate (which is highly pigmented and therefore easy to visualize). In fact, one of the best inhibitors turned out to be a drug candidate that had already undergone preliminary testing in humans for its ability to lower cholesterol levels. Despite the fact that these results are preliminary, the compound that was selected using this strategy caused a 98% decrease in surviving *S. aureus* in infected mice. Squalene synthase is also inhibited by a small molecule called TAK-475 or lapaquistat (Fig. 8.18), which unfortunately caused liver damage when used in high doses in clinical trials. Thus, while much remains to be done, this work demonstrates that by choosing a nonconventional target, it is possible to develop

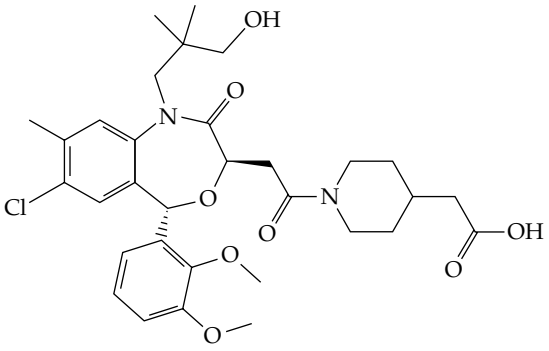
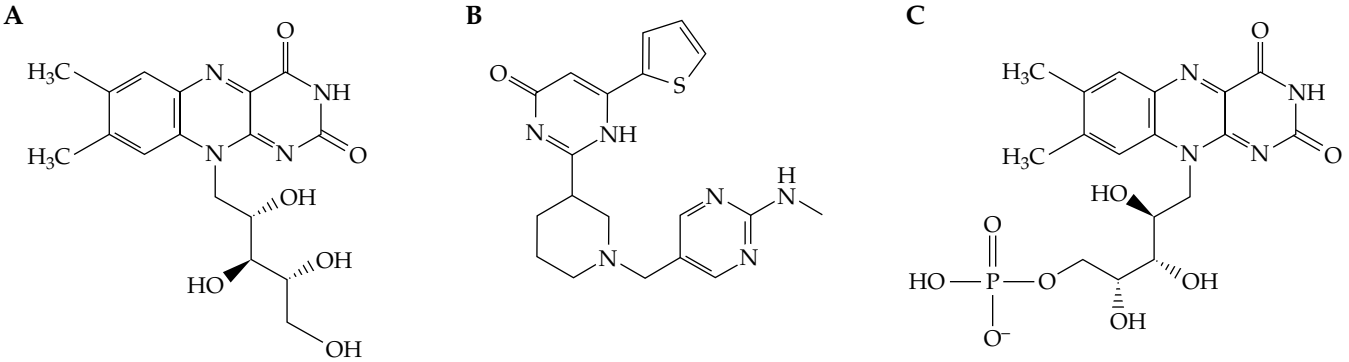


Figure 8.18 Chemical structure of the squalene synthase inhibitor lapaquistat.

designer antibiotics that render pathogenic bacteria susceptible to human and animal immune systems.

Riboflavin (vitamin B_2) is essential for bacterial growth; it is generally synthesized *de novo* but may also be taken up from the environment. Initially, in an effort to identify compounds that could inhibit the growth of pathogenic strains of *E. coli* by inhibiting the riboflavin (Fig. 8.19A) biosynthetic pathway, scientists from a large pharmaceutical company screened a library of ~57,000 synthetic small molecules for antibacterial activity. This screen led to the discovery of a compound named ribocil (Fig. 8.19B). Ribocil is not a close structural analog of flavin mononucleotide (FMN) (Fig. 8.19C), which is the usual regulator of riboflavin biosynthesis. However, ribocil can bind to a riboswitch (instead of FMN) in the mRNA upstream of a synthase enzyme gene that is involved in riboflavin biosynthesis, thereby inhibiting riboflavin biosynthesis (Fig. 8.20). Bacterial riboswitches are regulatory elements that are located in the 5′ untranslated regions of some bacterial mRNAs. Part of the riboswitch includes an aptamer-like molecule that binds to a specific metabolite such as an amino acid, an enzyme cofactor, or certain ions. When the metabolite binds to the riboswitch, a conformational change in a portion of the mRNA is induced, thereby altering the expression of the gene that is downstream from the riboswitch. Normally, when FMN binds to this riboswitch, it prevents the expression of the riboflavin synthase, stopping additional riboflavin synthesis when

Figure 8.19 The chemical structures of **(A)** riboflavin, **(B)** ribocil, and **(C)** flavin mononucleotide.



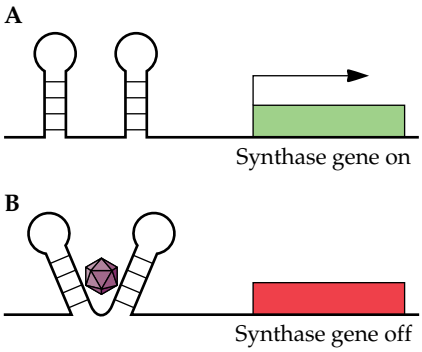


Figure 8.20 Modulation of riboflavin biosynthesis by a riboswitch translational modulator that is part of the mRNA. **(A)** In the absence of flavin mononucleotide (or ribocil), translation of the synthase gene occurs. The arrow indicates the direction of transcription of the synthase gene. **(B)** When bound to flavin mononucleotide (a derivative of riboflavin) or ribocil (shown as a purple shape), the riboswitch conformation changes, and translation of the synthase gene is prevented so that riboflavin is not synthesized. Riboflavin (not shown) can displace ribocil and activate riboflavin synthesis.

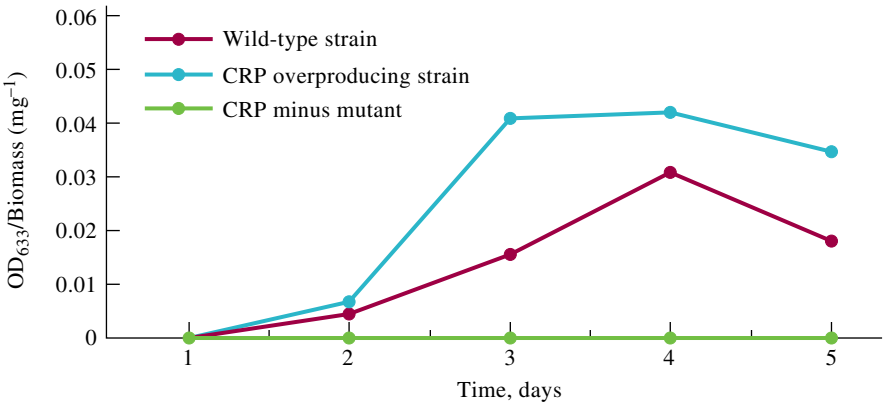
there is a sufficient amount of FMN present (i.e., this is a form of transcription attenuation). Although ribocil is not a structural analogue of the natural ligand of the riboswitch, it is a functional analogue. In this case, a riboswitch becomes a unique target for an antibiotic. With the intent of commercializing this and similar chemical compounds, the discoverers of ribocil have filed patent applications describing the synthesis and utility of ribocil and its chemical derivatives. The novel component of this invention is that it clearly demonstrates that targeting various riboswitches provides a novel means of screening for new unique antibiotics.

Overproducing Antibiotics

As pointed out above, in addition to producing new antibiotics, genetic manipulation can be used to rapidly and inexpensively enhance yields and thereby lower the production costs of existing antibiotics. This may be achieved in a number of different ways, described below.

It has been known for some time that the cyclic AMP receptor protein (Crp) is an important transcriptional regulator in various *Streptomyces* species. For example, in these organisms Crp plays a key role in spore germination and colony development. Scientists recently showed that Crp is also involved in regulating secondary metabolism and antibiotic production in *Streptomyces coelicolor*. In one set of experiments, both Crp-underproducing and -overproducing mutants of *S. coelicolor* were constructed, and then the behavior of these mutant strains was compared to the wild-type strain regarding the ability to synthesize the antibiotics actinorhodin, undecylprodigiosin, and calcium-dependent antibiotic. As expected, the Crp underproducing mutant yielded dramatically lower antibiotic levels, while the overproducing mutant yielded levels of these three antibiotics up to 2.5 times higher than those of the wild type (Fig. 8.21). Importantly, the introduction of the *crp* gene, encoding the Crp protein (which is highly conserved in *Streptomyces* species), into other *Streptomyces* species stimulated secondary metabolite production, including antibiotic production in all of the tested strains. In these experiments, the *crp* gene was cloned under the control of a strong constitutive promoter (*ermE*) on an integrating plasmid vector whose target chromosomal integration

Figure 8.21 Antibiotic actinorhodin production (measured as the optical density at 633 nm [OD_{633}]) per amount of biomass in wild-type and CRP-mutant strains of *Streptomyces coelicolor*.

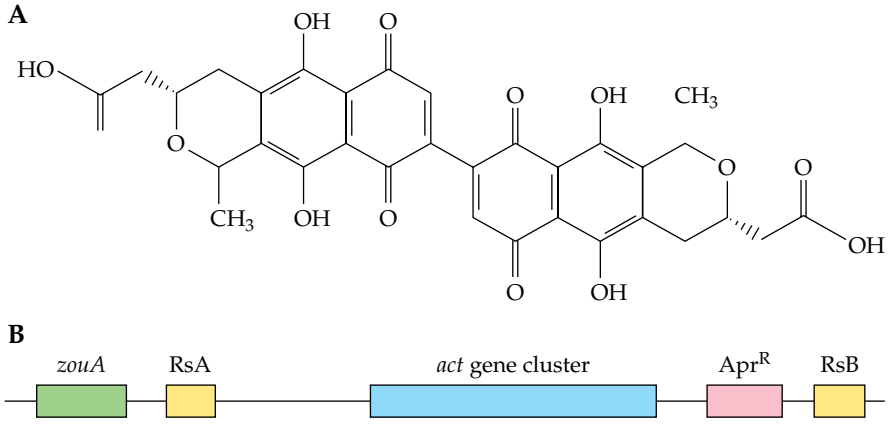


sequence is found in all *Streptomyces* species that have been examined. This latter result means that overproducing Crp may be a simple and effective strategy to increase the synthesis of many different *Streptomyces*-encoded antibiotics.

Detailed analysis of strains of *Penicillium chrysogenum* that produced increasingly higher levels of penicillin, following repeated rounds of mutagenesis and screening, revealed the presence of increasing numbers of tandem copies of the antibiotic gene cluster. Based on this analysis, researchers sought to develop a strategy for the targeted amplification of clusters of various microbial genes that encode the production of selected antibiotics. Such a procedure would ideally decrease the time that it takes to produce a bacterial strain that synthesizes high levels of a target antibiotic from the years that it took to accomplish this with penicillin to as little as several weeks or months. Thus, starting with the knowledge of the complete genome sequence of a strain of *S. coelicolor* that produces actinorhodin (Fig. 8.22A), a homologous recombination system was engineered to catalyze the tandem amplification of the actinorhodin gene cluster, which contains 22 genes for the biosynthesis of this antibiotic (Fig. 8.22B). Following insertion of the gene cluster shown in Fig. 8.22B in the chromosomal DNA and growth of the cells on increasing amounts of the antibiotic apramycin, strains of *S. coelicolor* with from 4 to 12 tandem copies of the actinorhodin biosynthesis gene cluster were produced. Some of the engineered *S. coelicolor* strains were able to produce up to 20 times the amount of actinorhodin synthesized by the parental strain. Moreover, it is possible that this technique may be synergistic with other approaches to increasing antibiotic productivity.

Large-scale production of antibiotics by *Streptomyces* spp. is often limited by the amount of oxygen available to cells. The low solubility of oxygen in aqueous media and the highly dense nature of filamentous *Streptomyces* cultures can result in an oxygen-depleted culture medium, a condition that causes poor cell growth and reduced antibiotic yield.

Figure 8.22 **(A)** Structure of the antibiotic actinorhodin. **(B)** The antibiotic actinorhodin gene (*act*) cluster (23 kb in size) flanked by recombination sites A and B (RsA and RsB), the gene (*zouA*) encoding a site-specific recombinase, and a gene for resistance to the antibiotic apramycin (apr^R). The entire cluster of genes shown is in the *S. coelicolor* genome. To amplify the cluster shown, the bacterium was passaged five times in medium with increasing amounts of the antibiotic apramycin.



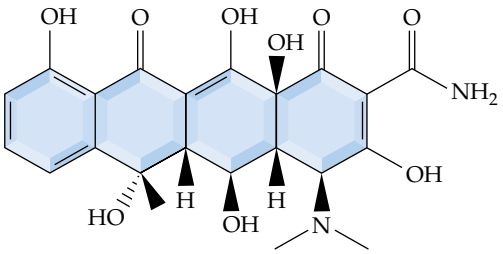


Figure 8.23 Structure of the antibiotic oxytetracycline.

To overcome this problem, *Streptomyces* strains better able to utilize the available oxygen were developed by genetic manipulation. To this end, the hemoglobin gene from the aerobic bacterium *Vitreoscilla* was isolated and expressed in *S. coelicolor*. This hemoglobin is a homodimeric heme protein that is functionally similar to eukaryotic hemoglobin and can sequester oxygen from the medium and then deliver it to the cells. When both wild-type and recombinant *S. coelicolor* cultures were grown in the presence of a low level of dissolved oxygen, the recombinant cells with a functional *Vitreoscilla* hemoglobin produced 10 times more of the antibiotic actinorhodin per gram of cells and had greater cell densities than did the wild-type cells. Thus, the expression of the *Vitreoscilla* hemoglobin gene in oxygen-starved antibiotic-producing microbial cells provides these cells with a mechanism for obtaining sufficient oxygen to allow both cell proliferation and increased antibiotic yields under otherwise limiting conditions.

The broad-spectrum antibiotic oxytetracycline (Fig. 8.23) was discovered in the 1940s, patented in 1949, and commercialized in 1950. Unfortunately, in recent years its effectiveness has been reduced because numerous bacterial strains developed resistance to it. However, this antibiotic is still used to treat infections caused by *Chlamydia* and *Mycoplasma*, as well as being a treatment for severe acne, and commonly is used as a feed additive in veterinary and aquaculture applications. Oxytetracycline production by *Streptomyces rimosus* requires 21 different proteins, with the antibiotic backbone being synthesized by a minimal polyketide synthase (Fig. 8.15) and an aminotransferase. Unfortunately, the industrial fermentation of oxytetracycline by *S. rimosus* is relatively slow, and the bacterium requires approximately 8 to 10 days to reach the maximum antibiotic production level. Thus, shortening the fermentation time would significantly improve the industrial productivity of this antibiotic. To decrease the oxytetracycline production time, scientists isolated the entire oxytetracycline biosynthetic gene cluster of *S. rimosus* and transferred all of these genes into a much faster-growing strain, *Streptomyces venezuelae*, which normally synthesized the antibiotic jadomycin. To improve the synthesis of oxytetracycline in its new host, *S. venezuelae* was engineered to increase the supply of malonyl-CoA, a key precursor of oxytetracycline. Following these modifications, the final engineered strain (Fig. 8.24) produced, in 2 days, slightly more antibiotic than the original strain did in 8 days. Thus, a conceptually simple, although technically somewhat complex, genetic manipulation significantly improved the rate of antibiotic production, a result that can significantly lower the cost of producing this antibiotic.

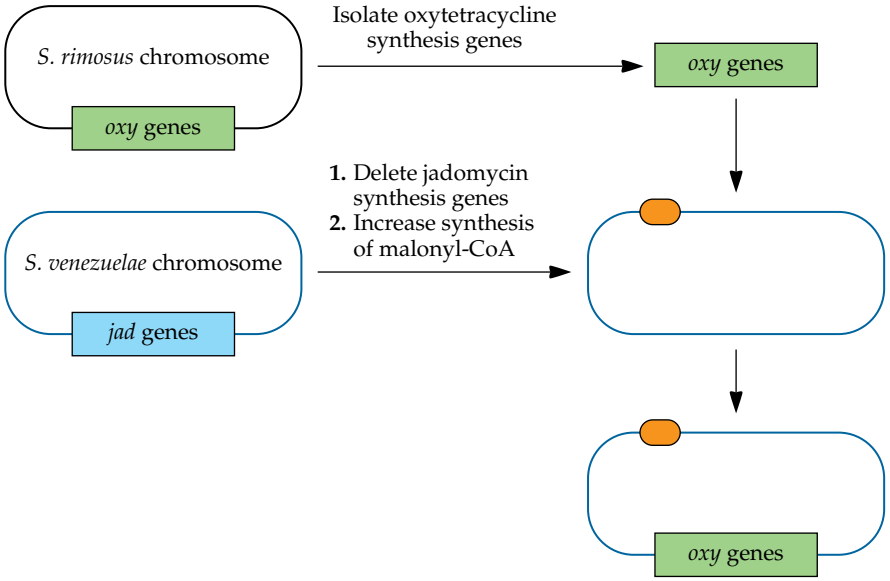


Figure 8.24 Engineering the more rapid synthesis of oxytetracycline. The cluster of genes encoding oxytetracycline biosynthesis (shown in green) in *Streptomyces rimosus* is isolated. The cluster of genes encoding jadomycin biosynthesis (shown in blue) in *S. venezuelae* is deleted. The altered *S. venezuelae* strain is engineered to overproduce malonyl-CoA (this change in the chromosome is shown in orange). The engineered *S. venezuelae* strain is transformed to express the oxytetracycline gene cluster.

Biopolymers

Biopolymers are large, multiunit macromolecules synthesized by microorganisms, plants, and animals. Some of these polymers have physical and chemical properties that are useful to the food-processing, manufacturing, and pharmaceutical industries. The ability to genetically engineer organisms has stimulated researchers to design new biopolymers, replace synthetic polymers with biological equivalents, modify existing biopolymers to enhance their physical and structural characteristics, and find ways to increase the yields and decrease the costs of biopolymers produced by industrial processes.

Xanthan Gum

Xanthomonas campestris is a Gram-negative obligatory aerobic soil bacterium that produces the commercially important biopolymer xanthan gum, a high-molecular-weight exopolysaccharide, as a by-product of its metabolism. This polymer has a cellulosic backbone made up of a straight chain of glucose units. Each of its trisaccharide side chains includes one glucuronic acid and two mannose residues, which are attached to every second glucose residue of the backbone (Fig. 8.25A). Xanthan gum has high viscosity, is stable in extreme physical and chemical environments, and exhibits physical and chemical properties similar to those of a plastic. In particular, its physical properties make it useful as a stabilizing, emulsifying, thickening, or suspending agent. For successful commercial production of xanthan gum, *X. campestris* should be grown on an inexpensive and plentiful carbon source. Wild-type *X. campestris* can efficiently utilize glucose, sucrose, and starch, but not lactose, as a carbon source.

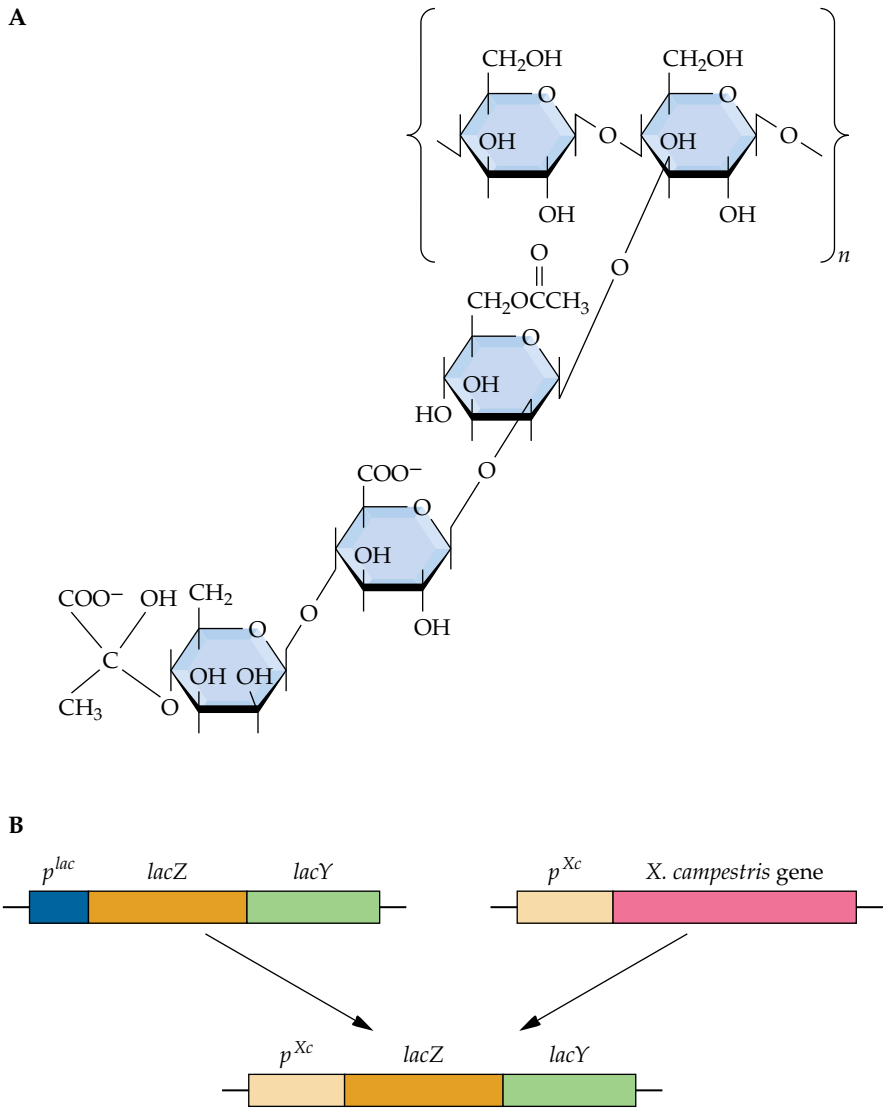


Figure 8.25 **(A)** Structure of xanthan gum. The repeating unit, designated *n*, forms a chain of glucose units. The trisaccharide is attached to alternate glucose residues of the repeating chain. **(B)** Engineering *E. coli lacZ* (encoding β-galactosidase) and *lacY* (encoding lactose permease) genes for constitutive expression in *X. campestris* (*Xc*).

Whey is a waste by-product of the cheese-making process that consists of water (94 to 95%), lactose (3.5 to 4%), and small amounts of protein, minerals, and low-molecular-weight organic compounds. Enormous quantities of whey are generated by the dairy industry, and its disposal is a major problem. In North America, whey has been used extensively as a "filler" in the preparation of prepared foods; however, with the increasing awareness that large numbers of individuals are lactose intolerant, it is imperative that alternative uses be found for this material. Moreover, disposing of whey by releasing it into rivers and lakes can increase microbial activity and deplete the amount of available oxygen, thereby killing many of the aquatic organisms. Transporting whey to landfill sites is exceptionally expensive, and potential groundwater contamination by the discarded

Table 8.3 Production of xanthan gum by wild-type and transformed *X. campestris*[a]

X. campestris	Amount of xanthan gum produced (μg/ml) with:		
	0.4% Glucose	0.4% Lactose	10% Whey
Wild type	3,530	245	224
Transformant	3,711	3,608	4,241

[a]The amount of product is expressed as micrograms per milliliter of culture grown on a minimal medium with either 0.4% glucose or 0.4% lactose added or on diluted whey (10%), which contains approximately 0.44% lactose. The transformant carries the *E. coli lacZY* genes on a plasmid.

whey is a major concern. Finally, the costs of removing the solid component of whey are prohibitive. Consequently, many schemes have been devised to find creative approaches to use whey.

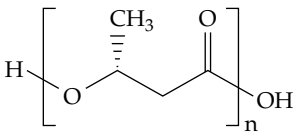
Theoretically, whey could be used as a carbon source for growing industrially important microorganisms. With this in mind, the bacterium *X. campestris* was genetically engineered to grow on whey. The *E. coli lacZY* genes, which encode the enzymes β-galactosidase and lactose permease, were cloned onto a broad-host-range plasmid under the transcriptional control of a strong *X. campestris* promoter (Fig. 8.25B). This construct was introduced into *E. coli* and then transferred from *E. coli* to *X. campestris* by triparental mating (see chapter 2, Fig. 2.9). Transformants maintained the plasmid, expressed β-galactosidase and lactose permease at high levels, utilized lactose as the sole carbon source, and produced large amounts of xanthan gum with glucose, lactose, or whey as the carbon source (Table 8.3). In contrast, wild-type *X. campestris* produces large amounts of xanthan gum only when grown on glucose (Table 8.3). Importantly, optimization of the fermentation conditions yields approximately ten times as much xanthan gum (Table 8.3), thereby increasing the prospects for the commercialization of this technology.

Polyhydroxyalkanoates

Polyhydroxyalkanoates (PHA) are a class of biodegradable polymers that are produced by a large number of different microorganisms and used as an intracellular carbon and energy storage material. These compounds have thermoplastic or elastic properties, depending on the polymer composition, and are being considered for use in the synthesis of a range of biodegradable plastics. It is estimated that in 2019 the global market for biodegradable plastics was around $3.3 billion per year and that this market is expected to reach $6 to 7 billion by 2027. However, at present, the very high cost of raw materials (30 to 40% of production costs) for PHA production by bacterial synthesis makes this process economically marginal. Therefore, scientists have sought to find more cost-effective methods to produce PHA, including using cyanobacteria and transgenic plants as a source of this material.

Poly(3-hydroxybutyric acid) is the most thoroughly studied and characterized polyhydroxyalkanoate (Fig. 8.26). Both the polymer and the (Gram-negative) *Alcaligenes eutrophus* (now named *Cupriavidus metallidurans*) genes that encode its synthesis have been characterized. Poly(3-hydroxybutyric acid), its copolymer [poly(3-hydroxybutyrate-*co*-3-hydroxyvalerate)], and another polyhydroxyalkanoate [poly(3-hydroxyvaleric acid)] (Fig. 8.27) are produced commercially in the United Kingdom by the fermentation of *C. metallidurans*. Poly(3-hydroxybutyric acid) is biocompatible with human

Figure 8.26 The chemical backbone of poly(3-hydroxybutyric acid).



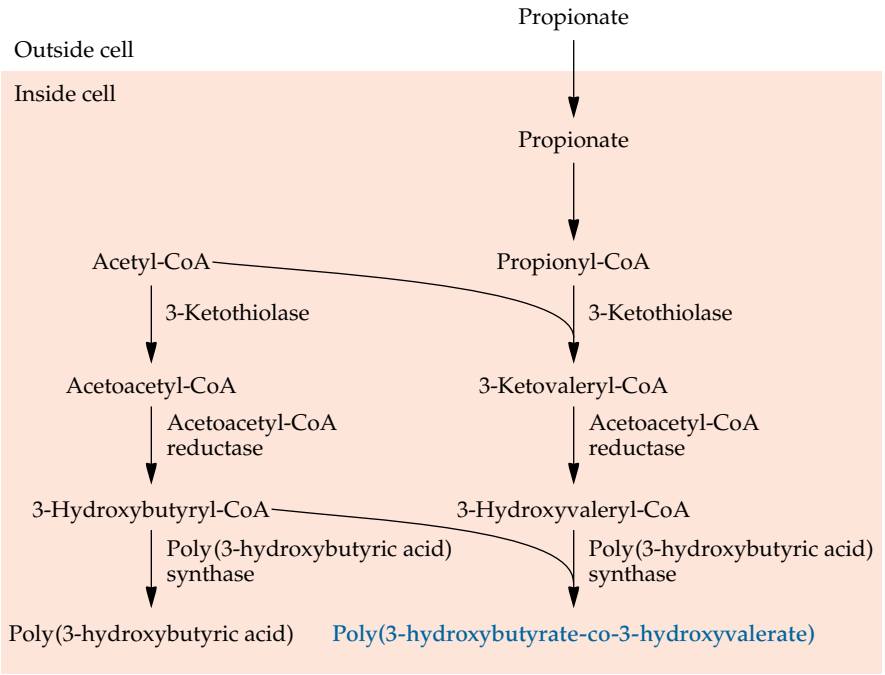


Figure 8.27 Bacterial synthesis of the polymer poly(3-hydroxybutyric acid) and of the (genetically engineered) copolymer poly(3-hydroxybutyrate-*co*-3-hydroxyvalerate).

tissue and can be used without inducing any inflammatory response. Moreover, the degradation of poly(3-hydroxybutyrate) is slow inside the body, so it can be used to facilitate the slow release of drugs and in tissue engineering. Unfortunately, poly(3-hydroxybutyrate) is rigid and brittle and has little chemical resistance. Therefore, it has been blended with other polymers to reduce its brittleness and improve some of its properties (Table 8.4).

Although it is possible to produce poly(3-hydroxybutyric acid) as a by-product of the growth of *C. metallidurans*, the organism grows relatively slowly, requires a relatively low growth temperature (so that the fermentation vessel must be cooled), is difficult to lyse [making the purification of poly(3-hydroxybutyric acid) granules difficult], and utilizes only a limited number of carbon sources for growth (making production costs relatively high). On the other hand, when the genes for the biosynthesis of this

Table 8.4 Poly(3-hydroxybutyrate)-based (PHB-based) polymer blends and their applications

PHB-based blends[a]	Application
PHB/starch	Textiles
PHB/chitosan	Drug delivery
PHB/polycaprolactone	Absorption properties
PHB/polyethylene glycol	Tissue engineering
PHBV/starch	Disposable goods (e.g., food trays, cups)
PHBV/polylactic acid	Biomedical devices and packaging

[a]PHBV is the copolymer poly(3-hydroxybutyrate-*co*-3-hydroxyvalerate).

polymer were transferred to *E. coli*, the resultant transformant grew rapidly to a high cell density and accumulated very large amounts (up to 95% of the dry cell weight) of poly(3-hydroxybutyric acid). Poly(3-hydroxybutyric acid) is synthesized from acetyl-CoA in three steps catalyzed by three enzymes (Fig. 8.27), and the operon encoding these genes was cloned into a plasmid as part of a 5.2-kb insert. Unfortunately, plasmids expressing the poly(3-hydroxybutyric acid) operon in *E. coli* were unstable. In the absence of selective pressure, such as the addition of antibiotics to the growth medium, about half of the *E. coli* cells lost the plasmid after approximately 50 generations. Plasmid loss of this magnitude, while not a major concern in small-scale batch cultures, becomes more of a problem with large-scale or continuous cultures (see chapter 9). This problem was overcome by inserting the *parB* genetic locus from another plasmid onto plasmids carrying the poly(3-hydroxybutyric acid) operon. This gene mediates plasmid stabilization by postsegregational killing of plasmid-free cells. The modified plasmids were quite stable even though poly(3-hydroxybutyric acid) was produced constitutively, which places a metabolic load on the cells. An added benefit of producing poly(3-hydroxybutyric acid) in *E. coli* instead of *C. metallidurans* is that when the poly(3-hydroxybutyric acid) is recovered by extraction with an alkaline hypochlorite solution, the polymer is degraded to a much lesser extent than when it is produced in *C. metallidurans*. This is probably because most of the poly(3-hydroxybutyric acid) in *E. coli* is produced in a crystalline state, while in *C. metallidurans* it is amorphous. Nevertheless, the polymers extracted from the two organisms had identical polymer properties. In addition, *E. coli* transformants synthesizing poly(3-hydroxybutyric acid) produced very little acetate, which can be deleterious for cell growth, presumably because all of the excess acetyl-CoA of the cell was converted to poly(3-hydroxybutyric acid) rather than acetate.

The copolymer poly(3-hydroxybutyrate-*co*-3-hydroxyvalerate) has physical properties that are similar to those of polypropylene. Consequently, there is considerable commercial interest in the biological production of poly-(3-hydroxybutyrate-*co*-3-hydroxyvalerate). However, *E. coli* strains that expressed the three polymer biosynthetic genes synthesized only poly-(3-hydroxybutyric acid) and not the copolymer. This limitation was overcome with *E. coli* cells that were mutated at both the *fadR* and *atoC* loci. The FadR protein is a negative regulator of fatty acid biosynthesis, and the *fadR* mutant activates the glyoxylate shunt, enhancing the capacity for energy metabolism and biosynthesis, which leads to a reduction of acetate excretion and improvement of the biomass yield. The *atoC* gene product is a positive regulator of fatty acid uptake, and the gene product from the *atoC* mutation turns on the synthesis of the proteins encoded by *atoA* and *atoD*, whose gene products facilitate the uptake of propionate from the growth medium into the cell. The propionate is converted to propionyl-CoA and then condensed with acetyl-CoA to form 3-ketovaleryl-CoA, which can be converted into 3-hydroxyvaleryl-CoA before its incorporation into the copolymer (Fig. 8.27). The amount of 3-hydroxyvalerate in the copolymer is dependent on the percentage of propionate that is added during the fermentation, but it never exceeds 40%.

In addition to engineering the composition of polyhydroxyalkanoates to produce polymers with specific desired properties, it is also essential that the chain lengths of these polymers be regulated. Some polymerizing

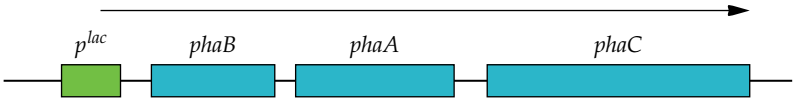
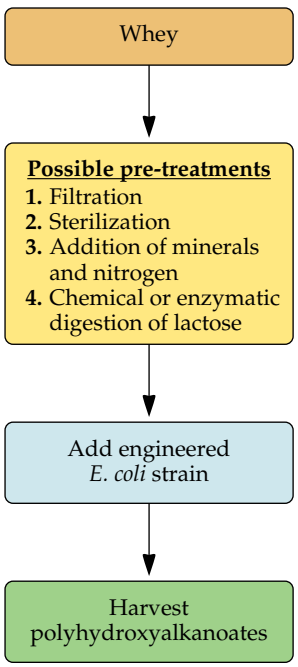


Figure 8.28 *Azotobacter* sp. genes encoding enzymes responsible for the biosynthesis of polyhydroxybutyrate under the transcriptional control of the *E. coli lac* promoter. The arrow indicates the direction of transcription. When a plasmid containing this construct is introduced into an *E. coli* strain that does not encode the *lac* repressor, the *pha* genes are expressed constitutively. *phaA*, 3-ketothiolase; *phaB*, acetoacetyl-CoA reductase; *phaC*, polyhydroxyalkanoate synthase.

enzymes yield primarily short-chain polymers with approximately 4 or 5 of the 3-hydroxyalkanoate monomeric units, while other enzymes from different bacteria, including *Pseudomonas oleovorans*, produce medium-chain polymers that include 6 to 14 monomeric units. In addition, mutations in enzymes that catalyze the various steps of fatty acid oxidation can be generated to produce monomeric units with a modified composition compared with the wild-type strain. These modified monomers may then be incorporated into the polyhydroxyalkanoate. As a consequence of these manipulations, a wide range of polyhydroxyalkanoates with different physical and chemical properties have been synthesized.

It would be economically advantageous if bacteria could be engineered to efficiently produce polyhydroxyalkanoates using industrial waste products as a carbon source. To achieve this goal, polyhydroxybutyrate biosynthesis genes from a strain of an *Azotobacter* sp. were inserted into a plasmid under the transcriptional control of the *E. coli lac* promoter (Fig. 8.28). The constructed plasmid was introduced into an *E. coli* strain that contained genes for the uptake and assimilation of lactose but that did not encode the *lac* repressor. Thus, both lactose uptake and assimilation genes, as well as polyhydroxybutyrate biosynthesis genes, were expressed constitutively. The *E. coli* strain that carried the constructed plasmid was able to grow on either 25% lactose (from whey) or corn steep liquor (a by-product of corn [maize] processing) and to produce a significant level of polyhydroxybutyrate. By growing the transformed *E. coli* strain aerobically in a fed-batch culture (see chapter 9), after 24 hours, the cells accumulated polyhydroxybutyrate to 73% of their cell dry weight. Moreover, the physical properties of the polyhydroxybutyrate that was produced were similar to the properties of the polymer isolated from the *Azotobacter* sp. Engineered *E. coli* strains that can break down lactose may be used to produce polyhydroxyalkanoates from industrial waste products in an industrial process to produce those polymers (Fig. 8.29). To increase the productivity of the *E. coli* (or other bacterium) that is used in this process, a number of different pretreatments have been employed (although not all pretreatments are used in the same experimental approach). One important pretreatment step involves the removal of proteins and other solids from the whey by ultrafiltration (see chapter 9). In addition, minerals and a nitrogen source may be added to increase the growth of the bacteria, and the lactose may be partially predigested to ensure that all of it is subsequently utilized. Besides using lactose as a carbon source for the synthesis of polyhydroxyalkanoates, some researchers have employed crude sludge palm oil (the oil that is not readily extracted during the milling of palm oil), xylose, and wheat straw (the stalk that remains after wheat grains are harvested) hydrolysate as starting materials for the synthesis of these polymers.

Figure 8.29 Flow chart of the industrial production of polyalkanoates by a genetically engineered strain of *E. coli*.



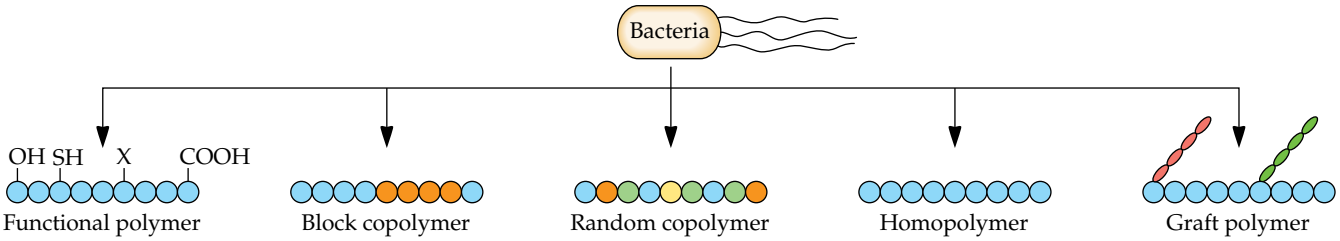


Figure 8.30 Schematic representation of some of the different types of polyalkanoates that can be synthesized by bacteria. A homopolymer is exclusively one type of polyalkanoate; a graft polymer is the addition of one or more molecules grafted onto a homopolymer; a functional polymer is a homopolymer containing one or more different types of side chains; a block copolymer consists of blocks (each with several identical monomers) of two or more different types of polyalkanoates in the same polymer chain; and a random copolymer consists of two or more different types of polyalkanoates randomly inserted into the same polymer chain.

Recently, scientists have reported engineering the synthesis of a poly(3-hydroxybutyrate-*co*-3-hydroxyhexanoate) in both *C. metallidurans*, using palm oil as a substrate, and in *E. coli*, using glucose as a carbon source. Given that poly(3-hydroxybutyrate-*co*-3-hydroxyhexanoate) is a tougher and more flexible plastic than poly(3-hydroxybutyric acid), this polymer could serve as a useful biodegradable plastic. In fact, since ~2010, there has been a significant effort by scientists around the world to develop a range of polyalkanoates (Fig. 8.30), synthesized by bacteria, with many different properties, some of which, it is hoped, will be useful in various industrial applications. As all of the details to the approaches shown schematically in Fig. 8.30 are worked out, there is every reason to expect that within the next 10 to 15 years scientists will be able to synthesize a myriad of new synthetic biodegradable commercial plastics.

Solvent Tolerance

For microorganisms to be able to efficiently synthesize a large number of different organic compounds, it is essential that these strains are able to tolerate relatively high concentrations of these organic solvent-like compounds, which might otherwise be toxic to the producing cells. An understanding of the key traits that make some bacteria naturally solvent-tolerant should facilitate the engineering of many commercial bacterial strains to make them more solvent-tolerant. If successful, this approach should facilitate the microbial production of both solvents (Table 8.5) and other organic molecules.

To improve the solvent tolerance of bacteria (either Gram-negative or Gram-positive strains), a number of different bacterial traits may be modified (Fig. 8.31). These include changing membrane permeability, altering some of the membrane proteins, decreasing the stress induced by solvent molecules, altering the bacterial cell bioenergetics, and altering the cell morphology. Membrane permeability may be decreased by changing lipid composition toward more saturated and transunsaturated fatty acids, which leads to membrane compaction. Membrane proteins may be modified to include more membrane-stabilizing proteins and an increased membrane efflux (so that it is easier for small synthesized molecules to be removed from inside the cell). A decreased level of stress inside the bacterial cell may come from increasing the cell's molecular chaperones and

Table 8.5 Some microbially synthesized solvents and their industrial uses

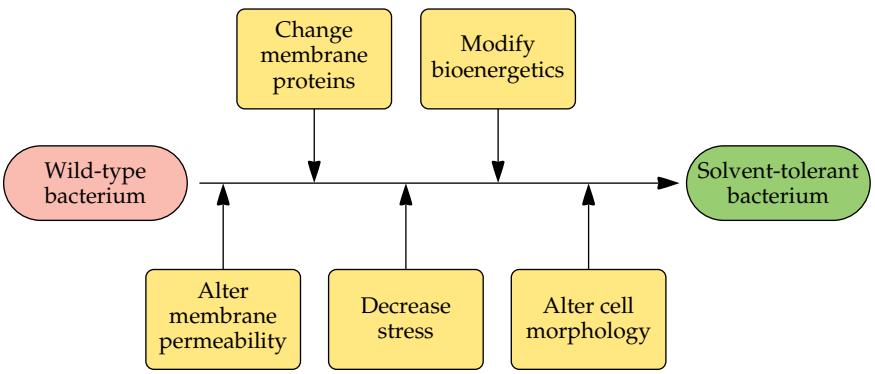
Solvent	Industrial use
Acetone	Cosmetic, pharmaceutical, medical, and domestic uses
Ethyl acetate	Formulation for epoxies, urethanes, acrylics, and vinyls
n-Butanol	Biofuel
Phenol	Plastic precursor
Butyl acetate	Cosolvent
Benzene	Substrate for the production of 3-methylcatechol
Toluene	Substrate for the production of 3-methylcatechol, o-cresol, and p-hydroxybenzoate
Styrene	Substrate for the production of (S)-styrene oxide; precursor of polystyrene
1-Octanol	Cosolvent
Ethylbenzene	Production of paints, varnishes, and lacquers
Cyclohexane	Precursor to nylon, adipic acid
m-Xylene	Substrate for the production of 3-methylcatechol
n-Hexane	Extraction solvent for vegetable oil, cleaning agent
1-Decanol	Cosolvent

general stress response proteins. Cellular biogenetics may be altered by upregulating the tricarboxylic acid (TCA) cycle and thereby increasing the amount of available NAD(P)H. When the cell size is decreased, the cell surface-to-volume ratio increases, facilitating more efficient nutrient uptake and solvent extrusion. Clearly, increased solvent tolerance requires a large number of cellular and biochemical changes. This work is still at a very early stage; however, given the large number of possible mechanisms for increasing solvent tolerance, it is expected that as experience is gained, scientists will be able to engineer not only increased solvent tolerance but also significantly increased product yield.

Systems Metabolic Engineering To Optimize Product Yield

By combining a variety of modern approaches and techniques, it is possible to engineer nearly any microorganism to become a host strain for the production of a very wide range of industrial products, from amino acids to antibiotics and from solvents to dyes. The overall approach to this "systems metabolic engineering" procedure is remarkably similar regardless

Figure 8.31 Schematic representation of various traits that might be modified in engineering a bacterium to make it more solvent-tolerant.



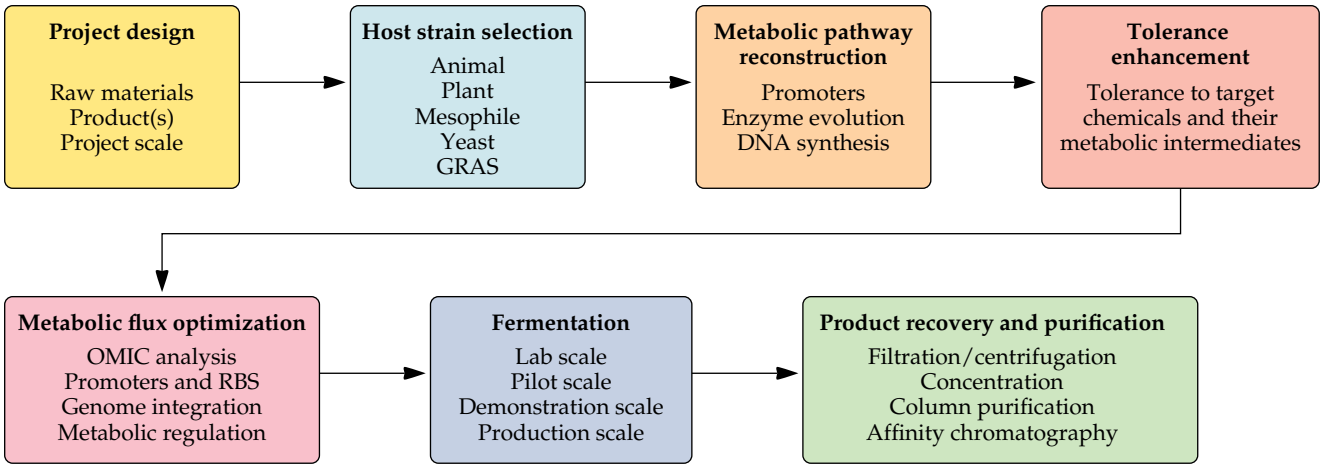


Figure 8.32 Flow chart showing some of the key steps in a systems metabolic engineering strategy for the microbial production of a wide range of industrial chemicals. Abbreviations: GRAS, microbial strains that are "generally regarded as safe"; OMICs, any or all of proteomics, transcriptomics, or metabolomics; RBS, ribosome binding sites.

of the target product (Fig. 8.32). Initially, in the project design phase, researchers select a compound that they wish to produce (e.g., ethanol), the raw materials that will provide the starting substrate (e.g., wheat straw), and the scale of the project (e.g., how large each facility for the conversion of wheat straw to ethanol will be and how many of these facilities will be required). Next, it is necessary to choose a host strain for the production of the desired product; in the example of converting wheat straw to ethanol, one might select an ethanologenic strain of bacteria or a strain of yeast. The next step in the process involves modifying the existing metabolic pathways in the selected microorganism to optimize production. This may mean modifying existing promoters, synthesizing variants of existing enzymes, or selecting better-performing variants of those enzymes through directed evolution (chapter 3). Subsequently, the microorganism may need to be engineered to increase its tolerance to the target chemicals (both the final product and the metabolic intermediates) using the process described in the previous section on solvent tolerance. To assess the effectiveness of the altered traits, the host strain's genome may be sequenced, and the organism's metabolic fluxes may be monitored by assessing the behavior of the microorganism under different conditions using transcriptomics, proteomics, or metabolomics (chapter 2). To optimize the metabolic flux and product yield, it may be necessary to introduce foreign genes into the host microorganism and/or to alter the promoters and/or ribosome binding sites of some of the genes involved in the bioconversion pathway, at the same time eliminating any unwanted feedback inhibition that might limit the yield of the final product. Having produced a microorganism that is "super" efficient on a laboratory scale, it is then imperative to develop a suitable fermentation strategy that can be as effective on a large scale as it is on a small scale, and finally, an efficient strategy for the purification of the final product on a large scale is required (see chapter 9). It is important to note here that this system is not limited to the conversion of wheat straw into ethanol; the same principles apply to hundreds of different situations, for example, if researchers are planning to convert whey into L-citrulline.

Microbial Degradation of Xenobiotics

The problem of toxic waste disposal is enormous. According to the U.S. Environmental Protection Agency, more than 40 million tons of hazardous waste are produced each year in the United States. Of these many hazardous chemicals, the Agency for Toxic Substance and Disease Registration lists 275 compounds as priorities for environmental cleanup. All of these compounds are considered to pose a significant threat to human and animal health. This list ranks environmental contaminants not solely on their toxicity but, rather, on a combination of their toxicity and their concentration in the environment. The organic compounds include trichloroethylene, trinitrotoluene, benzene, toluene, xylene, polycyclic aromatic hydrocarbons, polychlorinated biphenyls, pesticides, and insecticides. For example, worldwide production in 1985 of just one chemical that is released into the environment—pentachlorophenol—was more than 50,000 tons. Incineration and chemical treatment have been used to break down many toxic chemicals, but these methods are costly and often create new environmental difficulties. With the discovery in the mid-1960s of a number of soil microorganisms that are capable of degrading xenobiotic ("unnatural," or synthetic; from the Greek *xenos*, meaning "foreign") chemicals, such as herbicides, pesticides, refrigerants, solvents, and other organic compounds, the notion that microbial degradation might provide an economical and effective means of disposing of toxic chemical wastes gained credence.

Members of the genus *Pseudomonas* are the predominant group of soil microorganisms that degrade xenobiotic compounds. Various *Pseudomonas* strains can break down and, as a consequence, detoxify hundreds of different organic compounds. In many cases, one strain can use any of several different related compounds as its sole carbon source. Recently, it has been shown that many strains of *Burkholderia* spp. are also able to degrade a wide range of aromatic compounds (Fig. 8.33). The degradative pathways used by these bacteria include some of the same pathways utilized by various *Pseudomonas* spp., as well as some pathways that are unique to *Burkholderia* spp. As these degradative pathways are elaborated, it will expand the possibilities, using genetic manipulation of (i) combining several degradative pathways into one bacterium and (ii) increasing the rate of biodegradation of some compounds.

The biodegradation of complex organic molecules generally requires the concerted efforts of several different enzymes. The genes that code for the enzymes of these biodegradative pathways are sometimes located in the chromosomal DNA, although they are more often found on large (approximately 50- to 200-kilobase) plasmids (Table 8.6). In some organisms, the genes that contribute to the degradative pathway are found in both chromosomal and plasmid DNA.

Degradative pseudomonads, in most cases, enzymatically convert xenobiotic, nonhalogenated aromatic compounds, such as those listed in Table 8.6, to either catechol (Fig. 8.34A) or protocatechuic acid (Fig. 8.34B). Then, through a series of oxidative cleavage reactions, these compounds are processed to yield either acetyl coenzyme A (acetyl-CoA) (Fig. 8.34C) and succinate (Fig. 8.34D) or pyruvate (Fig. 8.34E) and acetaldehyde (Fig. 8.34F), compounds that are readily metabolized by almost all organisms. Halogenated aromatic compounds, the main constituents of many herbicides and pesticides, are broken down by the same enzymes that degrade the nonhalogenated compounds. However, for the halogenated

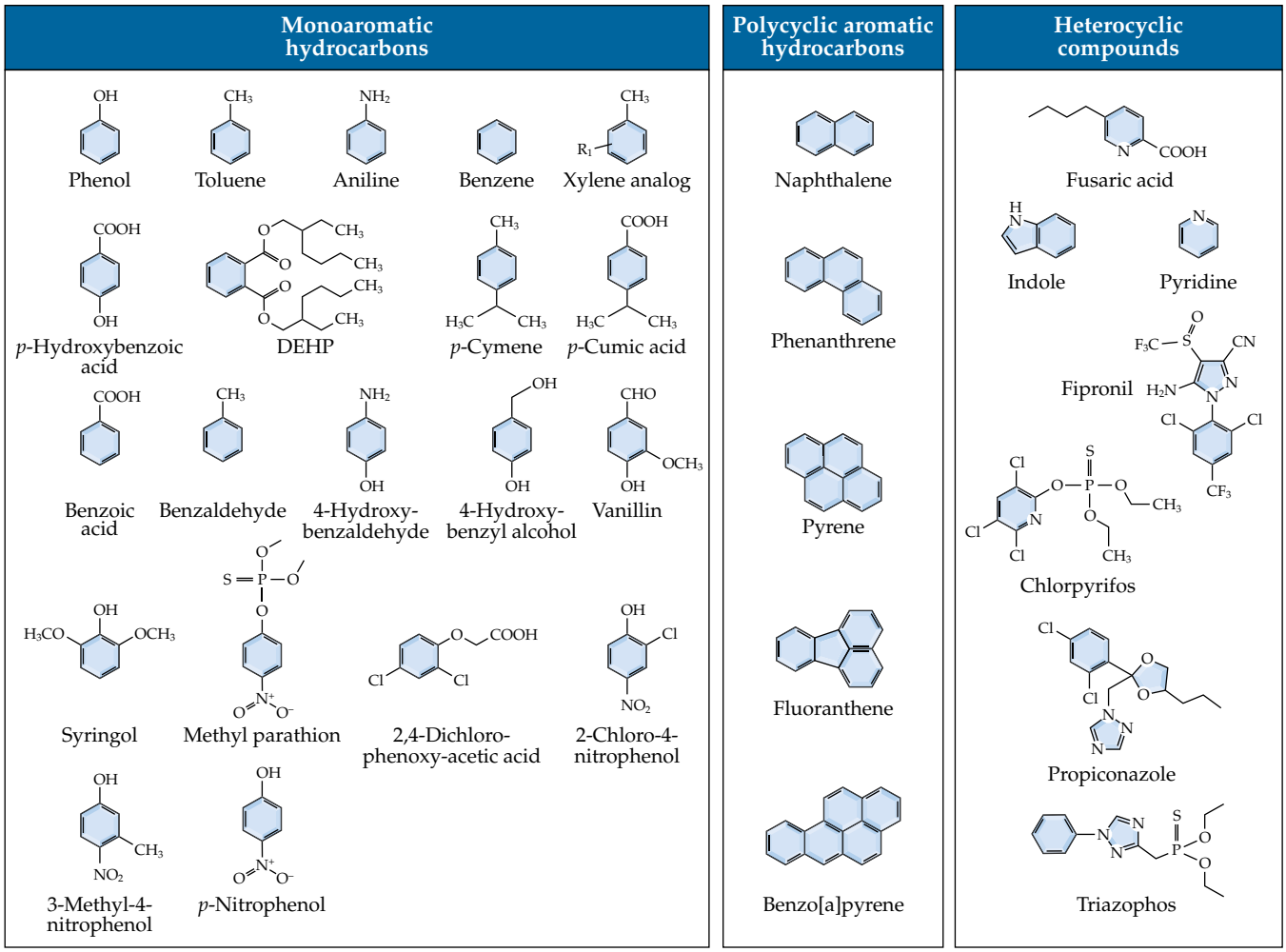


Figure 8.33 A schematic overview of some of the aromatic compounds that are degraded by various strains of *Burkholderia* spp.

compounds, the rate of degradation is much slower, being inversely related to the number of halogen atoms that are initially present on the target compound. Dehalogenation, the removal of a halogen substituent from an organic compound, is the critical requirement for detoxification and often occurs by a nonselective dioxygenase reaction that replaces the halogen on a benzene ring with a hydroxyl group. This step may occur either during or after the biodegradation of the original halogenated compound.

Genetic Engineering of Biodegradative Pathways

Despite the ability of many naturally occurring microorganisms to degrade a number of different xenobiotic chemicals, there are limitations to the biological treatment of these waste materials. For example, (i) no single microorganism can degrade all organic wastes; (ii) high concentrations of some organic compounds can inhibit the activity or growth of degradative microorganisms; (iii) most contaminated sites contain mixtures of chemicals, and an organism that can degrade one or more of the components

Table 8.6 Some *Pseudomonas* plasmids, their degradative pathways, and sizes[a]

Name of plasmid	Compound(s) degraded	Plasmid size (kilobases)
SAL	Salicylate	60
SAL	Salicylate	68
SAL	Salicylate	72
SAL	Salicylate	83
TOL	Xylene and toluene	113
CAM	Camphor	225
XYL	Xylene	15
NAH	Naphthalene	69
OCT	Octane, D-camphor	~500
NAH7	Napthalene, salicylate	83
PJP1	2,4-Dichlorophenoxyacetic acid	87
PJP2	2,4-Dichlorophenoxyacetic acid	54
PJP3	2,4-Dichlorophenoxyacetic acid	78
pP51	1,2-Di, 1,4-di, and 1,2,4-trichlorobenzene	110
pAC31	3,5-Dichlorobenzoate	108
pAC25	3-Chlorobenzoate	102
pWW0	Xylene and toluene	117
pWW100	Biphenyl	200
pWWO	Xylene and toluene	176
pXYL-K	Xylene and toluene	135
pVI150	Phenol	>200
pNL1	Xylene, naphthalene, biphenyl	184
pAC27	3-Chlorobenzoate	110
pHMT112	Benzene	112
pTDN1	Aniline, *m*- and *p*-toluidine	79

[a]Plasmids with the same name encode similar degradative pathways, even though they have different sizes and were described in different laboratories.

of the mixture may be inhibited by other components; (iv) many nonpolar compounds adsorb onto particulate matter in soils or sediments and become less available to degradative microorganisms; and (iv) microbial biodegradation of organic compounds is often quite slow. One way to address some of these problems is to transfer, by conjugation into a recipient strain, plasmids that carry genes for different degradative pathways (Fig. 8.35). If two resident plasmids contain homologous regions of DNA, recombination can occur, and a single, larger fusion or cointegrant plasmid with combined functions can be created. Alternatively, if two plasmids do not contain homologous regions and, in addition, belong to different incompatibility groups, they can coexist within a single bacterium.

Manipulation by Transfer of Plasmids

Bacterial strains with expanded degradative capabilities were first created in the 1970s by Dr. Ananda Chakrabarty and coworkers. They used different plasmids to construct a bacterial strain that degraded a number of the hydrocarbon components found in petroleum (Fig. 8.35). This strain has been called a "superbug" because of its increased metabolic capabilities. The CAM (camphor-degrading) plasmid was transferred by conjugation into a strain carrying the OCT (octane-degrading) plasmid. These two plasmids were incompatible and could not be maintained in the same cell

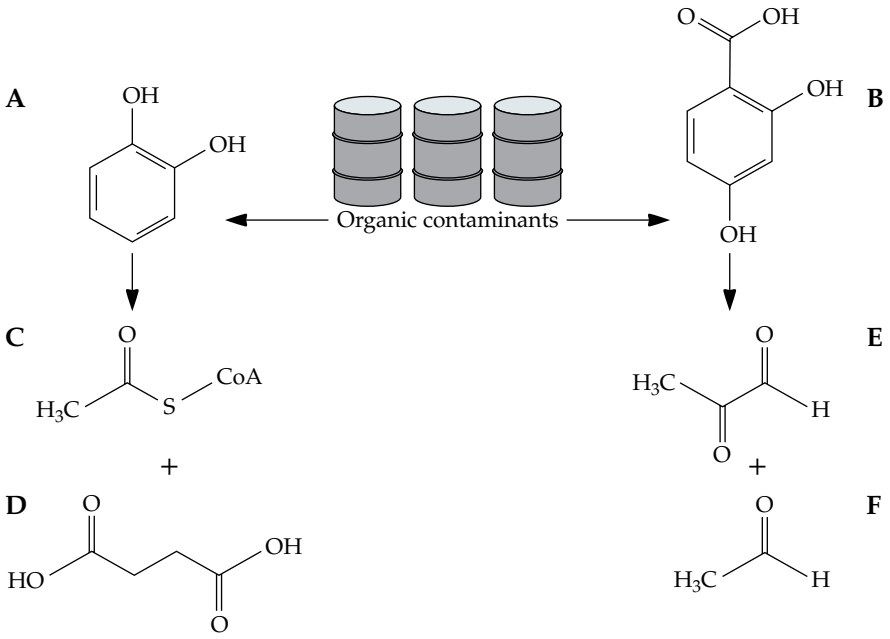


Figure 8.34 Biodegradation of organic contaminants to **(A)** catechol, **(B)** protocatechuic acid, **(C)** acetyl-CoA, **(D)** succinic acid, **(E)** pyruvic acid, and **(F)** acetaldehyde. Each of the arrows shown represents multiple biodegradative steps.

as separate plasmids. However, when recombination occurred between the two plasmids, the resulting single plasmid was perpetuated and carried both camphor- and octane-degradative activities. The NAH (naphthalene-degrading) plasmid was transferred by conjugation into a strain carrying the XYL (xylene-degrading) plasmid. The NAH and XYL plasmids were compatible and could therefore coexist within the same host cell. Finally, the CAM/OCT fusion plasmid was transferred by conjugation into the strain carrying the NAH and XYL plasmids. The final result of these manipulations was the generation of a strain that grew better on crude oil than did any of the single-plasmid strains either alone or in combination.

Although this particular multiple-degradative strain has not been used to clean up environmental oil spills, it has played a critical role in the development of the biotechnology industry. The inventor of this "superbug" was granted a U.S. patent describing its construction and use. This was the first patent ever granted for a genetically engineered microorganism and represented a watershed court decision, because it implied that biotechnology companies could protect their inventions in the same way as the chemical and pharmaceutical industries had in the past (see chapter 13).

Manipulation by Gene Alteration

Bringing together different plasmid-encoded degradative pathways by conjugation is only one way to create bacteria with novel properties. It is also possible to extend the degradative capability of a strain by altering the genes of an existing degradative pathway. The feasibility of this approach was examined for the toluene- and xylene-degrading pathway of plasmid pWWO. This plasmid encodes a "*meta*-cleavage" pathway involving 12 different genes and enables pseudomonads carrying the plasmid to

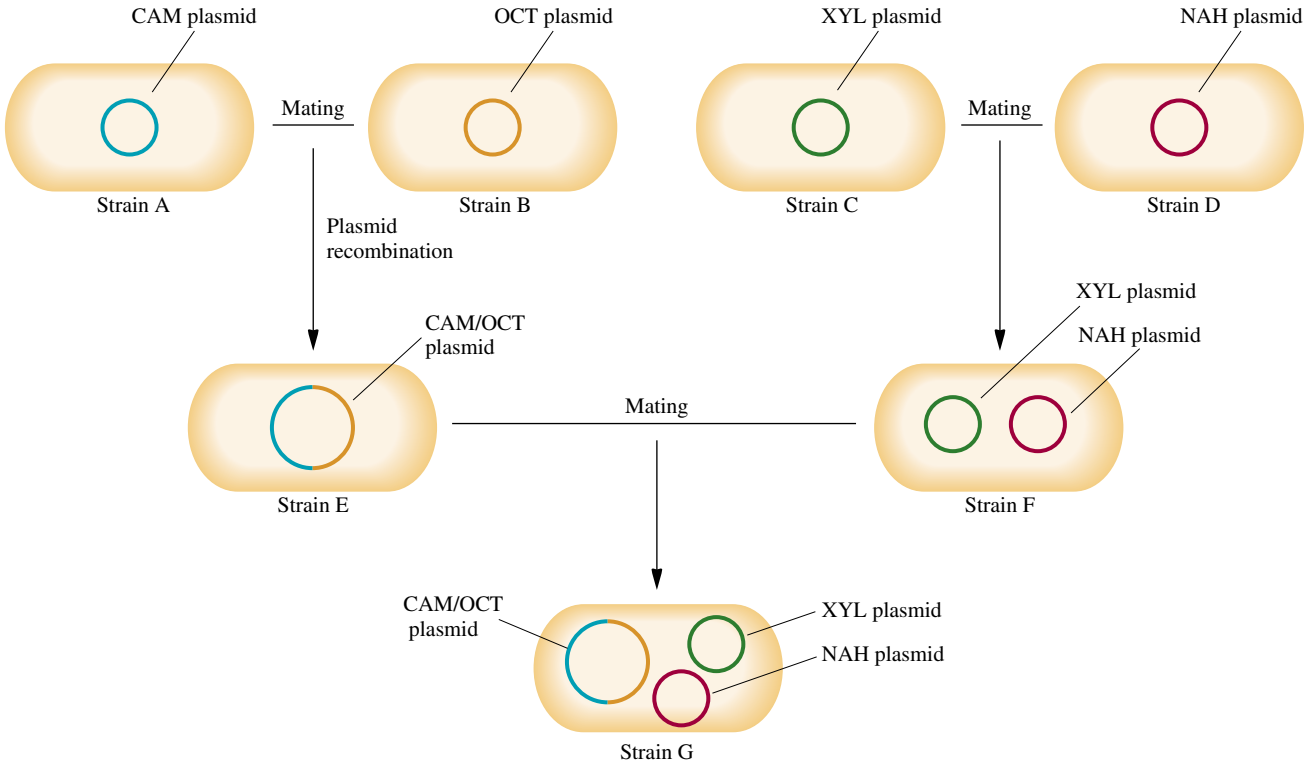


Figure 8.35 Schematic representation of the development of a bacterial strain that can degrade camphor, octane, xylene, and naphthalene. Strain A, which contains a CAM (camphor-degrading) plasmid, is mated with strain B, which carries an OCT (octane-degrading) plasmid. Following plasmid transfer and homologous recombination between the two plasmids, strain E carries a CAM and OCT biodegradative cointegrant plasmid. Strain C, which contains an XYL (xylene-degrading) plasmid, is mated with strain D, which contains an NAH (naphthalene-degrading) plasmid, to form strain F, which carries both of these plasmids. Finally, strains E and F are mated to yield strain G, which carries the CAM/OCT plasmid, the XYL plasmid, and the NAH plasmid.

utilize various alkylbenzoates as carbon sources (Fig. 8.36). The genes in the toluene–xylene pathway of pWWO are part of a single operon, called the *xyl* operon, under the control of the p^m promoter. Transcription from the p^m promoter, by RNA polymerase, is positively regulated by the *xylS* gene product, which is activated by most of the initial substrates, such as benzoate and 3-methylbenzoate, of the pathway (Fig. 8.36). Detailed biochemical and genetic analyses showed that bacteria carrying pWWO could degrade 4-ethylbenzoate, albeit slowly, to 4-ethylcatechol, which accumulated in the medium, but no further. 4-Ethylcatechol prevented its own degradation by inactivating one of the most important enzymes in the biodegradative pathway, catechol 2,3-dioxygenase, the product of the *xylE* gene. In addition, 4-ethylbenzoate, unlike most other alkylbenzoates, does not activate the XylS protein; consequently, transcription of the operon from the p^m promoter did not occur to any significant extent when 4-ethylbenzoate was the only substrate. Thus, there are two major problems with the naturally occurring *meta*-cleavage pathway system: (i) how to overcome the inactivation of an important enzyme in the degradative process by 4-ethylbenzoate

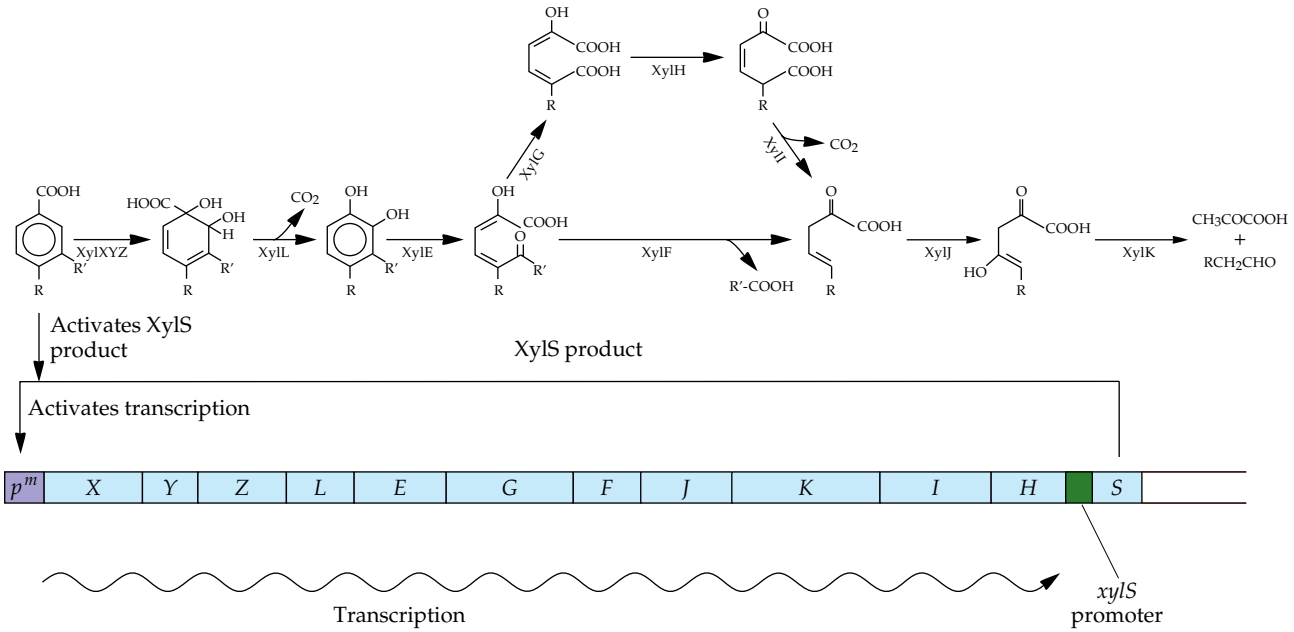


Figure 8.36 The *meta*-cleavage pathway and the *xyl* operon of the toluene- and xylene-degrading plasmid pWWO. Transcription of the *xyl* operon is controlled by the p^m promoter, which is regulated by the XylS gene product, which in turn must be activated by one of the initial pathway substrates. The genes from *xylX* to *xylH* (*X* to *H*) are under the control of the *pm* promoter. The *xylS* gene, which is not part of this operon, is constitutively expressed. Some of the primary substrates are benzoate, where R and R′ = H; 3-methylbenzoate, where R = H and R′ = CH_3; 3-ethylbenzoate, where R = H and R′ = CH_2CH_3; and 4-methylbenzoate, where R = CH_3 and R′ = H. The *xylXYZ* genes encode toluene dioxygenase, *xylL* encodes dihydroxycyclohexadiene carboxylate dehydrogenase, *xylE* encodes catechol 2,3-dioxygenase, *xylF* encodes hydroxymuconic semialdehyde hydrolase, *xylG* encodes hydroxymuconic semialdehyde dehydrogenase, *xylH* encodes 4-oxalocrotonate tautomerase, *xylI* encodes 4-oxalocrotonate decarboxylase, *xylJ* encodes 2-oxopent-4-enoatehydratase, and *xylK* encodes 2-oxo-4-hydroxypentonate aldolase.

and (ii) how to induce transcription of the genes of this pathway with 4-ethylbenzoate as the inducer.

To find a mutant that could solve the second problem, a tetracycline resistance gene was placed under the control of the p^m promoter on one plasmid, which also carried an ampicillin resistance gene. The *xylS* gene was cloned onto another plasmid carrying a kanamycin resistance gene. Transformants carrying both of these plasmids were selected on the basis of resistance to both ampicillin and kanamycin (Fig. 8.37A). The *E. coli* cells carrying both of these plasmids were treated with the chemical mutagenic agent ethyl methanesulfonate and plated onto a medium containing both tetracycline and 4-ethylbenzoate. The only cells that could grow on this medium carried an altered XylS protein (S* in Fig. 8.37A) that could interact with 4-ethylbenzoate and cause the tetracycline resistance gene to be transcribed. Thus, the degradative pathway that includes this mutated *xylS* gene can be induced by 4-ethylbenzoate. To address the problem of catechol 2,3-dioxygenase inactivation by 4-ethylcatechol, the mutated *xylS* gene was subcloned onto a broad-host-range plasmid carrying a kanamycin resistance gene and introduced into *P. putida* cells carrying pWWO (Fig. 8.37B).

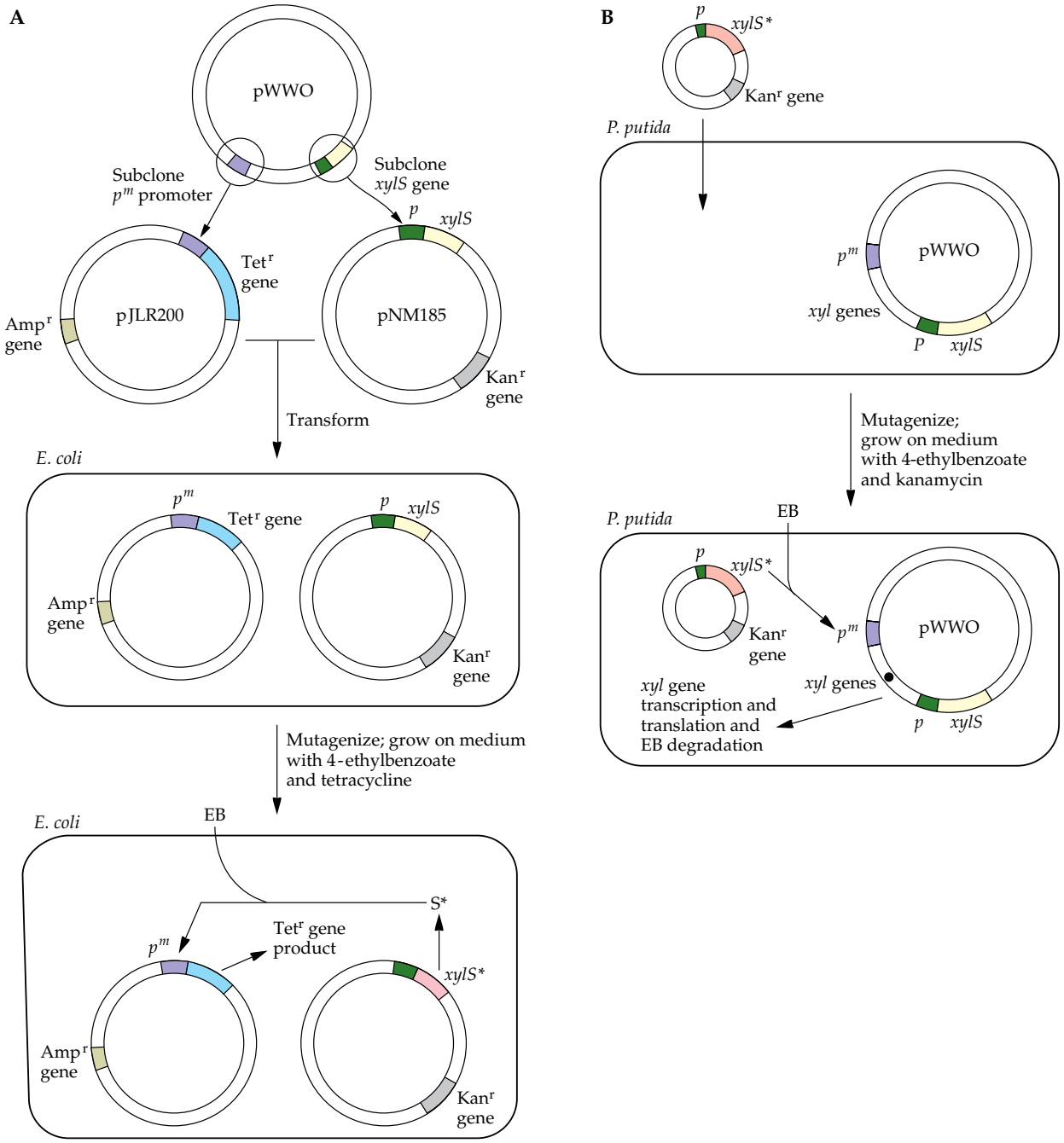


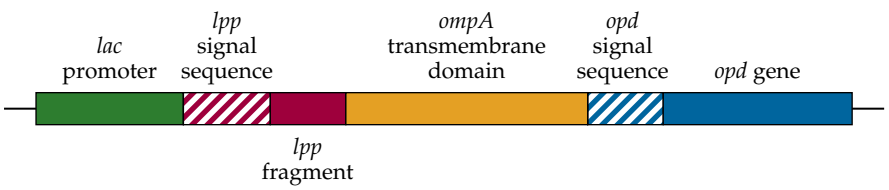
Figure 8.37 **(A)** Protocol used to create an XylS protein that is activated by 4-ethylbenzoate. The p^m promoter is cloned onto plasmid pBR322 and replaces the tetracycline resistance (Tet^r) gene promoter to form plasmid pJLR200; the *xylS* gene and its promoter are spliced onto a broad-host-range plasmid containing a kanamycin resistance (Kan^r) gene. *E. coli* is transformed with both of these plasmids. Transformants are selected by their resistance to both ampicillin (Amp^r) and kanamycin and then chemically mutagenized with ethyl methanesulfonate. Only cells with a mutation (S*) in the *xylS* gene that enables the XylS protein to be activated by 4-ethylbenzoate (EB) can grow on medium that contains both 4-ethylbenzoate and tetracycline, because only these cells are resistant to tetracycline. **(B)** Protocol used to create a modified catechol 2,3-dioxygenase that is not inhibited by 4-ethylcatechol. A *P. putida* strain carrying pWWO is transformed with a broad-host-range plasmid carrying the mutated *xylS** gene, whose product can activate the p^m promoter. Transformants are chemically mutagenized and then grown on minimal medium that contains 4-ethylbenzoate as the sole carbon source and kanamycin. Any cell that can grow on this medium carries a mutated catechol 2,3-dioxygenase gene. This mutation is indicated by a dot in the middle of the *xyl* gene cluster.

The transformed cells were plated, at a high cell density, onto a minimal medium containing 4-ethylbenzoate as the sole carbon source, kanamycin to select for the presence of the plasmid, and ethyl methanesulfonate. Cells that were able to grow on this medium produced an altered form of the enzyme catechol 2,3-dioxygenase that was not inhibited by 4-ethylcatechol. Additional analysis confirmed that the catechol 2,3-dioxygenase gene on pWWO had been mutated and that mutant versions of both the *xylS* gene and the catechol 2,3-dioxygenase gene were required for the degradation of 4-ethylbenzoate.

An important aspect of this work is the fact that the two genes that were altered, i.e., those encoding XylS and catechol 2,3-dioxygenase, are the major determinants of the range of compounds that can be degraded by this pathway. The work with 4-ethylbenzoate demonstrates that by combining recombinant DNA technology, conventional mutagenesis, and the appropriate selection protocols, novel properties can be added to a degradative pathway. Using more current technology to manipulate this pathway, researchers might eliminate the use of a chemical mutagen and instead use either directed mutagenesis or CRISPR technology (see chapters 2 and 3) to alter the target genes.

It is estimated by the World Health Organization that that ~3 million cases of severe pesticide poisoning occur each year, resulting in >200,000 deaths. Currently, detoxification of organophosphate pesticides in the environment is performed by chemical treatment, incineration, or burial in landfill sites. Each of these approaches has serious environmental drawbacks. It would therefore be advantageous if bacteria that are able to degrade these compounds could be utilized in place of these methods. Several soil bacteria, including *Pseudomonas diminuta* MG and *Flavobacterium* spp., possess an enzyme, organophosphorus hydrolase, that catalyzes the hydrolysis of many of these pesticides, including methyl and ethyl parathion, paraoxon, chlorpyrifos (Dursban), coumaphos, cyanophos, and diazinon, to environmentally innocuous compounds. Unfortunately, these bacteria, as well as *E. coli* engineered to express organophosphorus hydrolase, degrade these pesticides relatively slowly because of the low rate of uptake of organophosphate pesticides into the bacterial cells. Thus, a novel approach was developed to solve this problem. *E. coli* cells were engineered to express organophosphorus hydrolase (and its signal peptide) as part of a fusion protein that contained the *E. coli* lipoprotein signal peptide and the N-terminal portion of the lipoprotein and outer membrane protein A (Fig. 8.38). When this fusion protein was synthesized, it was localized on the outer surface

Figure 8.38 The DNA construct used to produce the fusion protein Lpp–OmpA–OPH. This construct includes the *E. coli lac* promoter, DNA encoding the *E. coli lpp* signal sequence and the first 9 amino acids of the mature *E. coli* lipoprotein, the portion of the *E. coli ompA* gene encoding the transmembrane domain, and the gene (*opd*) for the *Flavobacterium* sp. organophosphorus hydrolase (OPH) and its signal peptide.



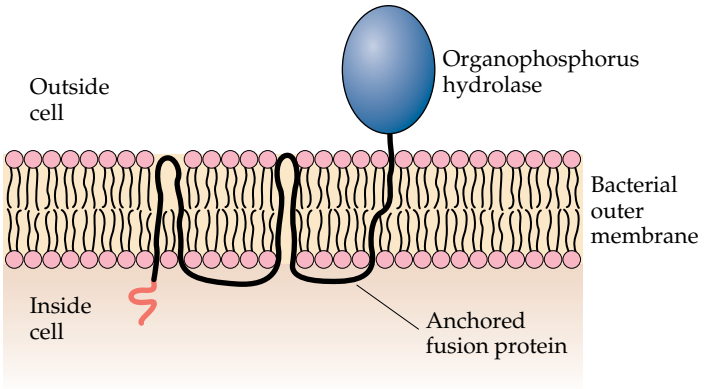


Figure 8.39 Schematic representation of the fusion protein Lpp–OmpA–OPH anchored in the *E. coli* outer membrane with organophosphorus hydrolase on the outside of the cell and therefore exposed to the external medium. Lpp (shown in red) includes the first 9 amino acids from the *E. coli* lipoprotein. OmpA (shown in black) includes the transmembrane domain (amino acids 46 to 159) of *E. coli* outer membrane protein A. OPH (shown in blue) includes *Flavobacterium* sp. organophosphorus hydrolase and its signal peptide.

of the bacterium (Fig. 8.39). This eliminated the problem of a low rate of pesticide uptake into the engineered bacterium. Cells with organophosphorus hydrolase on their outer surfaces had approximately seven times higher activity than cells that expressed a similar amount of the enzyme intracellularly. Moreover, the enzyme activity of cells with organophosphorus hydrolase on their surfaces was extremely stable. Nearly 100% of the activity was retained after 1 month at 37°C. This concept is quite promising; however, to date, it has been tested only on a laboratory scale.

For many years, a large number of different nitroaromatic compounds have been used industrially as dyes, plasticizers, explosives, solvents, and pesticides. Many of these compounds are recalcitrant to breakdown, persist in the environment, and are considered to be toxic and sometimes carcinogenic pollutants. For example, the compound 4-nitrophenol (Fig. 8.40) is formed by the hydrolysis of the insecticide parathion and is considered to be a priority environmental pollutant, in part because it leads to numerous human health problems. Similarly, the compound 3-methyl-4-nitrophenol (Fig. 8.40) is a toxic breakdown product of the agricultural insecticide fenitrothion.

The bacterium *Burkholderia* sp. strain DNT facilitates the breakdown and detoxification of 2,4-dinitrotoluene. First, the enzyme 2,4-dinitrotoluene dioxygenase removes one nitro group to form 4-methyl-5-nitrocatechol (Fig. 8.40), which is then converted to 2-hydroxy-5-methylquinone by the enzyme 4-methyl-5-nitrocatechol monooxygenase. Unfortunately, the 4-methyl-5-nitrocatechol monooxygenase has a very narrow substrate range, which includes 4-nitrocatechol (Fig. 8.40), as well as 4-methyl-5-nitrocatechol. However, this enzyme cannot efficiently use either of the similar substrates 4-nitrophenol or 3-methyl-4-nitrophenol. In an effort to expand the substrate range of the enzyme, the *Burkholderia* sp. strain DNT 4-methyl-5-nitrocatechol monooxygenase gene was isolated and subjected to error-prone PCR (chapter 3), and the randomly mutated genes were cloned into a plasmid vector. The library of mutated 4-methyl-5-nitrocatechol monooxygenase genes was transferred into *E. coli* cells by electroporation, and the transformants were screened for activity

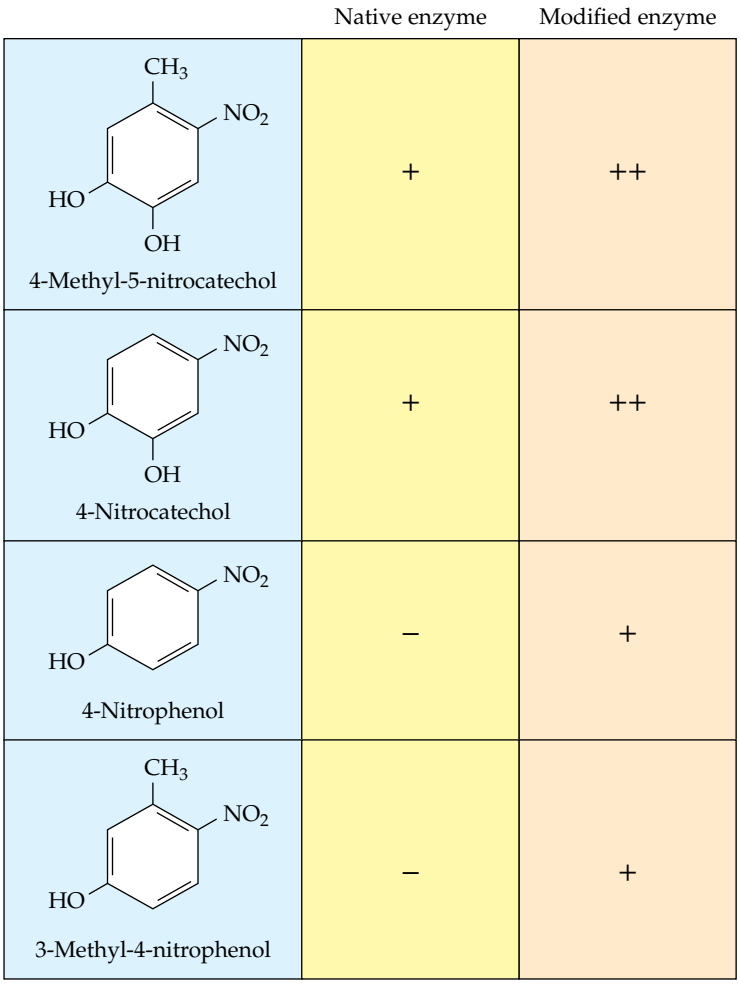


Figure 8.40 Breakdown of nitroaromatic compounds by a strain of *Burkholderia* sp. that produces either the native form or a modified form of the enzyme 4-methyl-5-nitrocatechol monooxygenase.

on agar plates containing 4-nitrophenol (which is colorless to light yellow). Cells that contained the wild-type gene turned the colonies light brown, while one transformant (of 3,000 tested) turned the colony dark brown, indicating that the 4-nitrophenol was being broken down. When the DNA sequence of the 4-methyl-5-nitrocatechol monooxygenase gene from the transformant that turned the medium dark brown was determined, two amino acids within the encoded protein had been altered. Amino acid 22 was changed from methionine to leucine, and amino acid 380 was changed from leucine to isoleucine. These two amino acid alterations resulted in the altered enzyme having 10 times greater activity toward 4-nitrophenol and 4 times greater activity toward 3-methyl-4-nitrophenol than the native form of the enzyme. In addition, the modified enzyme had about 50% more activity than the native enzyme toward 4-nitrocatechol and 4-methyl-5-nitrocatechol. The changes in activity that were found following the mutagenesis of the 4-methyl-5-nitrocatechol monooxygenase gene are an important first step in developing a bacterium that can effectively degrade 4-nitrophenol and 3-methyl-4-nitrophenol.

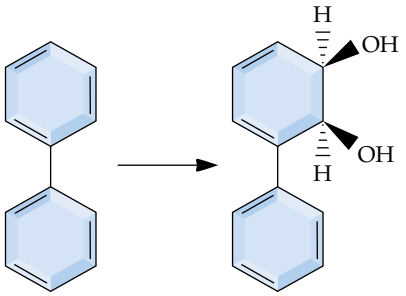
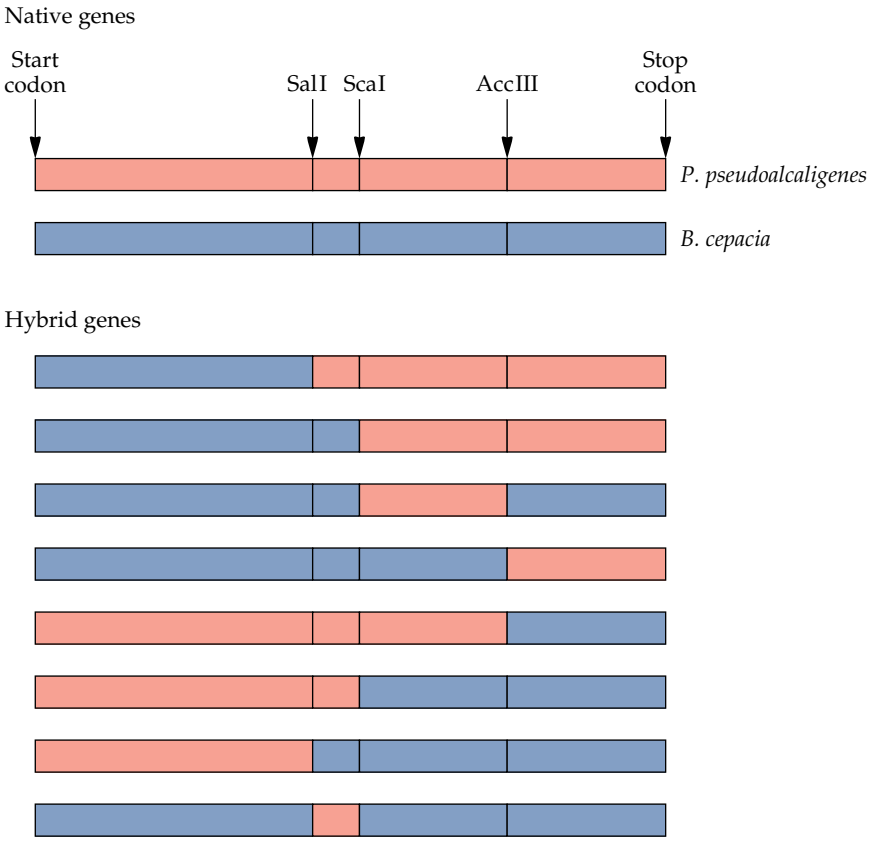


Figure 8.41 The first step in the enzymatic breakdown of the compound biphenyl using the enzyme biphenyl dioxygenase.

Bacterial strains that can degrade the compound biphenyl use the enzyme biphenyl dioxygenase (Fig. 8.41). Biphenyl dioxygenase is a multicomponent enzyme encoded by four genes, *bphA1A2A3A4*, where *bphA1* encodes a large subunit of terminal dioxygenase (an iron-sulfur protein), *bphA2* encodes a small subunit of terminal dioxygenase, *bphA3* encodes ferredoxin, and *bphA4* encodes ferredoxin reductase. BphA1 and BphA2 are associated as a heterotetramer and catalyze the introduction of two oxygen atoms into the biphenyl ring. Ferredoxin and ferredoxin reductase act as an electron transfer system from reduced nicotinamide adenine dinucleotide (NADH) to reduce the terminal dioxygenase. The generation of degradative bacteria with novel biological activities may be achieved by the creation of chimeric versions of the *bphA1* gene (Fig. 8.42). In this case, one *bphA1* gene was from a strain of *Pseudomonas pseudoalcaligenes* with the ability to degrade only a narrow range of polychlorinated biphenyls (PCBs), while the other was from *Burkholderia cepacia*, which can degrade a different range of PCBs. While the BphA1 enzymes produced by these two different strains were quite similar to one another, they nevertheless differed by 20 amino acids. Following gene shuffling (chapter 3), some of the hybrid genes encoded an enzyme with a wider degradative ability than either of the original enzymes and had enhanced abilities to degrade

Figure 8.42 Chimeric (hybrid) *bphA1* genes constructed from *P. pseudoalcaligenes* and *B. cepacia bphA1* genes. The chimeras were constructed by digesting both of the isolated genes with the restriction enzymes shown and then recombining the resultant fragments as indicated. The various transformants carrying these constructs were subsequently assayed for enzymatic (degradative) activity.

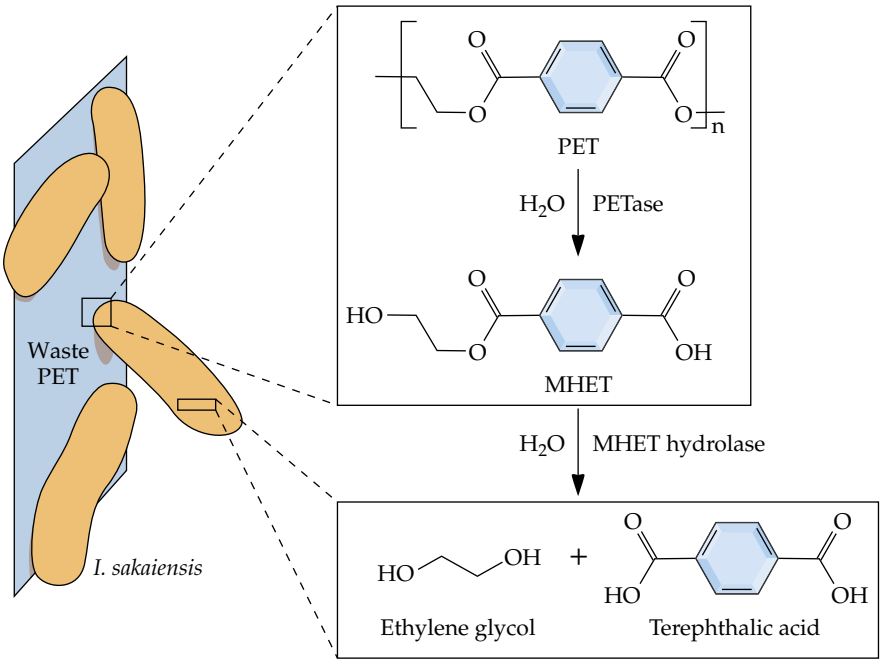


benzene, toluene, and alkyl benzenes, all of which are poorly degraded by the two starting enzymes. The native *bphA1* gene in *P. pseudoalcaligenes* was subsequently replaced with hybrid genes by homologous recombination. Some of these modified *P. pseudoalcaligenes* strains acquired enhanced biodegradative capabilities. It now remains to be seen whether these engineered bacterial strains can degrade a range of PCBs on a large scale.

Plastics

Between the 1950s and 2018, it is estimated that 6.3 billion tonnes of plastic were produced worldwide, with around 9% of the total being recycled and another 12% being incinerated. Unfortunately, the other ~80% of these plastics eventually entered the environment, much of it finding its way into rivers, lakes, and oceans. Some scientists have suggested that by 2050, there could be more plastic than fish in the oceans, by weight. One long-term solution to this problem is to switch to the use of degradable plastics. However, this solution does not address the problem of the massive amount of plastics that are already present in the environment. One way to address this environmental contamination problem is to select and utilize microorganisms that can completely degrade different types of plastics. In a study in 2016, scientists isolated and characterized a bacterium from contaminated sediment near a plastic bottle recycling plant in Japan that completely degrades and assimilates the plastic poly(ethylene terephthalate) (PET). The selected bacterium, *Ideonella sakaiensis*, is a Gram-negative rod-shaped aerobic betaproteobacterium (Fig. 8.43). First, the bacterium binds

Figure 8.43 Schematic representation of the bacterium *I. sakaiensis* binding to the surface of a portion of waste PET. The bacterium secretes the enzyme PETase, which breaks down the PET polymer into the compound mono(2-hydroxyethyl) terephthalic acid (MHET), which may then be taken up by the bacterium, where it is hydrolyzed by MHET hydrolase to ethylene glycol and terephthalic acid.



to the plastic surface. Then, the bacterium secretes the enzyme PETase, which depolymerizes the PET polymer to mono(2-hydroxyethyl) terephthalic acid (MHET), which is taken up by the bacterium and hydrolyzed to ethylene glycol and terephthalic acid by the enzyme MHETase. These two compounds are further metabolized by this bacterium to facilitate bacterial growth. Notwithstanding the exciting potential of this discovery, the complete degradation of a small piece of PET film (~0.2 mm thick) took 6 weeks, which is too slow for this bacterium to be used in a practical way. However, it is hoped that the genetic manipulation of this bacterium (focusing on the genes encoding the PET degradative enzymes) might significantly increase the rate of PET degradation. In addition, it might be possible to develop a process to isolate the terephthalic acid produced following PET degradation, thereby providing a major cost savings in the synthesis of a new PET polymer. Importantly, since the isolation and characterization of *I. sakaiensis*, several other plastic-degrading bacteria have been isolated and are currently being characterized and tested. Moreover, another group of researchers has recently reported genetically manipulating the gene encoding PETase so that the enzyme's degradation ability has been improved.

Utilization of Starch and Sugars

Starch, the major food reserve polysaccharide in plants, consists of a mixture of linear homopolymers (amylose) and branched homopolymers (amylopectin) of D-glucose. Amylose is made up of linear chains of 1×10^2 to 4×10^5 D-glucose residues linked by α-1,4 glycosidic bonds (Fig. 8.44A). Amylopectin consists of short linear chains of approximately 17 to 23 glucose units that are linked by α-1,4 bonds and joined by 1,6 linkages and some 1,3 linkages to form a highly branched structure that contains 1×10^4 to 4×10^7 glucose residues (Fig. 8.44B). The degree of branching and the ratio of amylose to amylopectin vary with the source and age of the starch, although in the majority of plants around 20 to 25% of starch is amylose.

Commercial Production of Fructose and Alcohol

The major use of starch is in the food and brewing industries, where it is usually hydrolyzed to low-molecular-weight components before being converted into other compounds, especially fructose and alcohol. The most important enzymes for breaking down and transforming starch are α-amylase, glucoamylase, and glucose isomerase (Fig. 8.44). Together, these three enzymes account for approximately 30% of the cost of all enzymes currently used for industrial processes.

Fructose and alcohol are produced commercially from starch by multistep processes that include both enzymatic and nonenzymatic reactions, as follows (Fig. 8.45).

1. The procedure begins with the gelatinization of milled grain (often corn [maize], which is approximately 40% starch). This treatment—steam cooking under pressure—exposes the surface area of the starch, thereby making it more readily available for subsequent enzymatic hydrolysis. The product of this process has a gel-like consistency.
2. The gelatinized starch is cooled to 50 to 60°C, and the enzyme α-amylase is added. In this liquefaction step, the gel-like starch is enzymatically digested by the hydrolysis of the available α-1,4-linkages

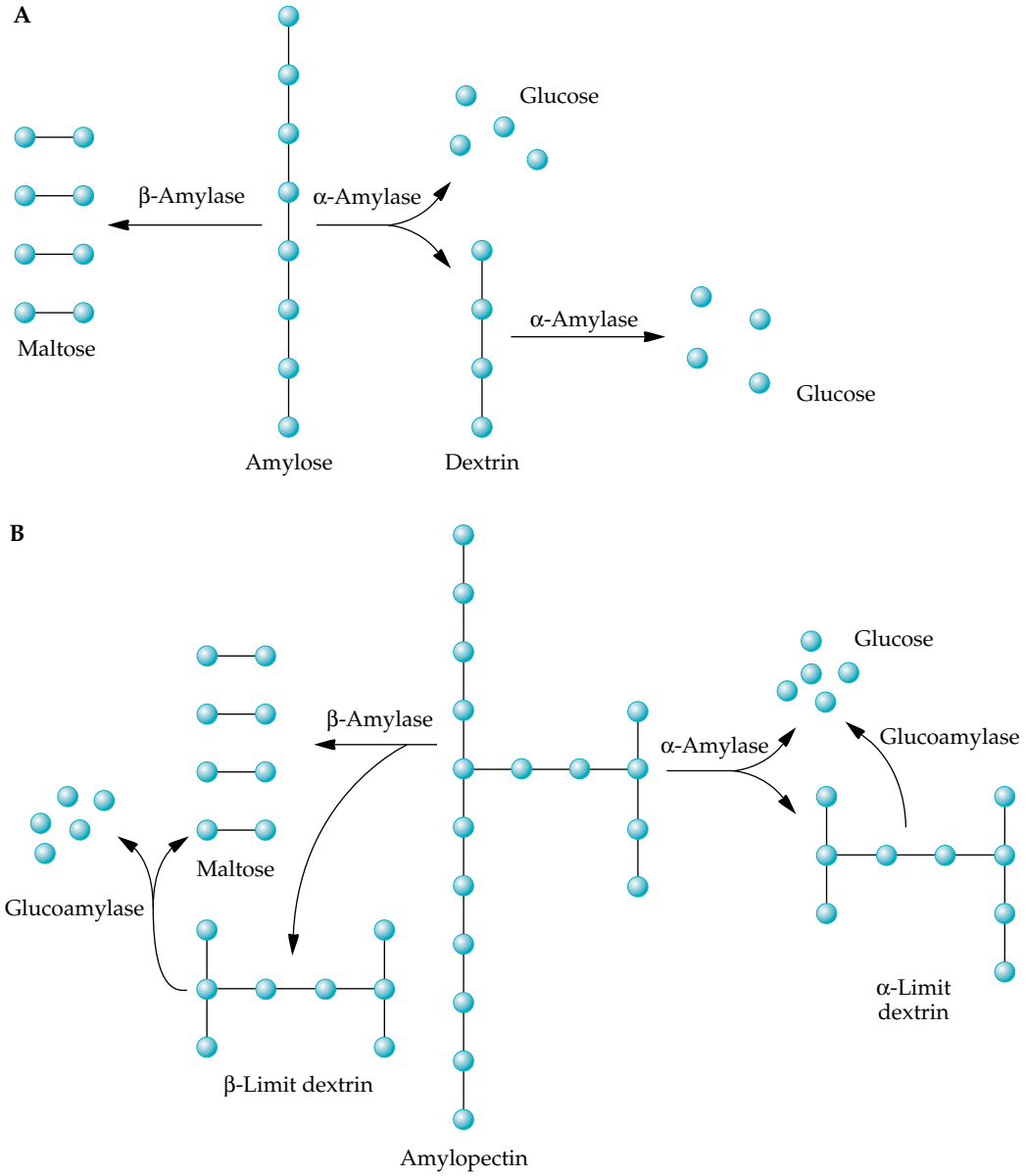


Figure 8.44 **(A)** Pathway for the enzymatic hydrolysis of amylose; **(B)** pathway for the enzymatic hydrolysis of amylopectin. The blue circles represent D-glucose residues.

to form soluble low-molecular-weight polysaccharides. A high temperature is used because enzyme hydrolysis is more rapid at high temperatures and the enzyme does not efficiently penetrate the gelatinized starch at lower temperatures.

3. The final step for the release of glucose includes the addition of glucoamylase and results in the saccharification—complete hydrolysis—of the remaining polysaccharides, including both linear and cross-linked molecules.

The end product of these treatments is glucose, which may be either converted into alcohol as a result of fermentation by yeast cells or transformed into fructose by the enzyme glucose isomerase. The success of the latter

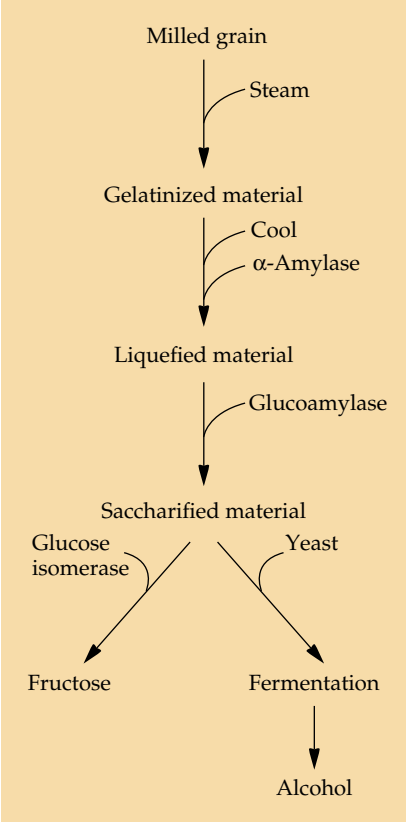


Figure 8.45 Industrial production of fructose and alcohol from starch.

enzymatic conversion has led to the replacement of sucrose by fructose as a sweetener for prepared foods and beverages in North America. This is because of the lower cost of fructose and its perceived greater sweetness, therefore requiring smaller amounts of sweetener. The source of starch for commercial fructose production is generally corn, and the final product is called high-fructose corn syrup or high-fructose syrup, although it contains a mixture of fructose and glucose. High-fructose corn syrup typically consists of either 90% fructose (for baked goods), 55% fructose (in soft drinks), or 42% fructose (in sports drinks), with the remainder being glucose.

The enzyme α-amylase randomly hydrolyzes α-1,4-linkages in both amylose and amylopectin chains, yielding a mixture of glucose, maltose (two glucose molecules joined by an α-1,4 linkage), maltotriose (three glucose molecules joined by α-1,4 linkages), and a series of α-limit dextrins, which are the portions of the amylopectin chains that contain cross-links (Fig. 8.44). Although α-amylase can be isolated from a variety of microorganisms, it is commonly obtained from *Bacillus amyloliquefaciens* for industrial purposes.

For some applications, the enzyme β-amylase, produced by a variety of microorganisms, is used in addition to or in place of α-amylase to digest starch. By hydrolyzing alternate α-1,4 linkages from the ends of amylose and amylopectin chains, β-amylase cleavage yields primarily maltose residues and various β-limit dextrins (Fig. 8.44).

The enzyme glucoamylase hydrolyzes α-1,3, α-1,4, and α-1,6 linkages; however, because it is less efficient than α-amylase in cleaving α-1,4 linkages, it is usually used in conjunction with α-amylase. The major role of glucoamylase is digestion of the cross-links of amylopectin, which results in its complete breakdown to glucose. Glucoamylase and other enzymes are used to reduce the carbohydrate (limit dextrin) content of normal beers to produce the so-called light and dry varieties. Although glucoamylase digestion is usually performed prior to the onset of the fermentation, the two steps may be combined. A number of organisms produce glucoamylase, but for industrial purposes, it is usually obtained from the fungus *Aspergillus niger*.

Increasing Alcohol Production

The enzymes that are used in the production of alcohol or fructose from milled grain are major components of the overall cost of the process. These enzymes are often used only once and then discarded. Thus, innovative approaches to the inexpensive large-scale production of the enzymes could lower the cost of alcohol or fructose production. There are several ways to achieve this end.

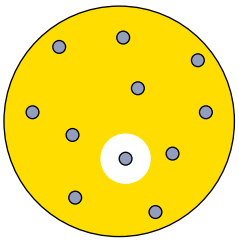
- Each of the enzymes could be overproduced in a fast-growing recombinant microorganism that utilizes an inexpensive substrate, thereby lowering the cost compared with production from native organisms.
- Variants of α-amylase, either naturally occurring or genetically manipulated, that function efficiently at 80 to 90°C could be used to allow the liquefaction step to be performed at this temperature. Heat-resistant α-amylase would speed the hydrolysis of gelatinized starch while decreasing the amount of energy that is required to cool the gelatinized starch to a temperature suitable for starch hydrolysis.

- The α-amylase and glucoamylase genes could be altered so that each enzyme would have the same temperature and pH optimum, thereby enabling the liquefaction and saccharification steps to be performed under the same conditions.
- An enzyme that could efficiently degrade raw starch could be found or engineered, obviating the need for the gelatinization step and thereby saving a large amount of energy.
- A fermentation organism that can synthesize and secrete glucoamylase could be developed, eliminating the need to purify and add this enzyme during fermentation.

A considerable amount of research has been initiated to determine whether these possibilities are feasible.

Genes that code for α-amylase have been isolated from a number of organisms, including *B. amyloliquefaciens* and the high-temperature-tolerant bacterium *Bacillus stearothermophilus*. Briefly, chromosomal DNA was isolated from these bacteria and used to prepare a genomic library (chapter 2) that was transformed into *Bacillus subtilis*, which does not have α-amylase activity. All transformants were tested for the production and secretion of α-amylase as follows. After the transformants had formed colonies at 65°C on solid medium containing starch, the plates were exposed to iodine vapor. The colonies producing α-amylase were surrounded by a distinctive halo, or clear zone, indicating that the starch in the immediate vicinity of these cells had been hydrolyzed (Fig. 8.46). A positive starch–iodine test signifies that the transformed cell contains an α-amylase gene that is transcribed from its own promoter, because the vector does not carry a promoter. Also, a secretion signal is present, because the substrate is too large to enter the cell, and therefore the halo must be due to the activity of a secreted α-amylase. The availability of α-amylase genes from varied sources enables researchers to carry out specific genetic modifications that suit the needs of specific industrial processes.

Figure 8.46 Detection of a *B. subtilis* clone expressing an α-amylase gene. The halo results from the degradation of the starch in the medium in the vicinity of the clone by secreted α-amylase.



With the aim of bypassing the saccharification step that requires the addition of purified glucoamylase during the production of alcohol from starch, researchers isolated a full-length glucoamylase cDNA from the fungus *Aspergillus awamori* and cloned it into a *Saccharomyces cerevisiae* plasmid under the control of the promoter and transcription terminator regulatory signals from the yeast enolase (*ENO1*) gene (Fig. 8.47). A laboratory strain of *S. cerevisiae* that was transformed with the plasmid carrying the glucoamylase cDNA was able to express this activity and to ferment soluble starch to alcohol, thus demonstrating that the approach is feasible. Unfortunately, this laboratory strain of *S. cerevisiae* has a number of properties that make it ill suited for use in a commercial process, including an inability to tolerate high levels of alcohol, inefficient expression of the glucoamylase cDNA, and loss of plasmids unless selective pressure is used for their maintenance. These problems were overcome as follows. First, the level of glucoamylase expression was increased approximately 5-fold by deleting a 175-base-pair (bp) negative regulatory region from the *ENO1* promoter on the plasmid. Second, the plasmid was modified by deleting its yeast origin of replication and adding a segment of yeast DNA that is homologous to a yeast chromosomal site, thereby converting it to an integrating vector (chapter 3). With this form of the plasmid, the complete glucoamylase construct was incorporated into a chromosomal site and stably maintained. Third, another *S. cerevisiae* strain (brewer's yeast) that tolerates

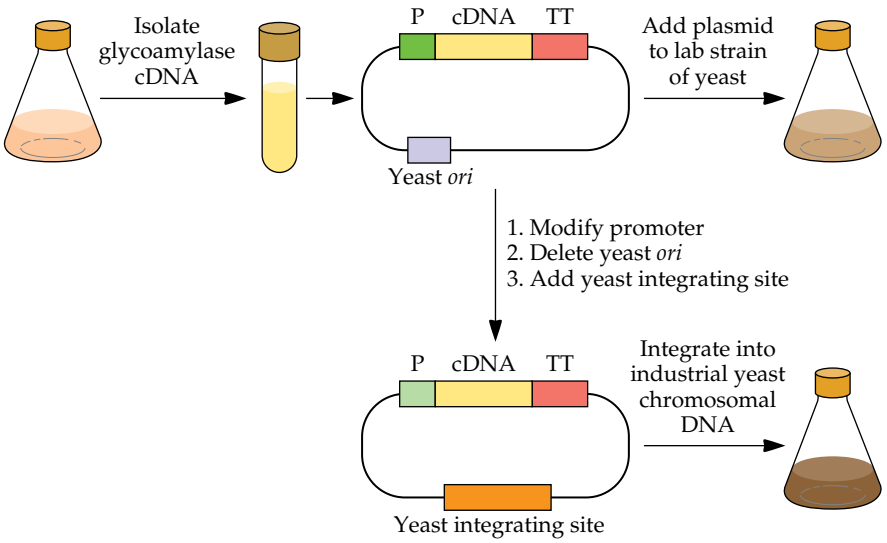


Figure 8.47 Expressing an *Aspergillus awamori* glucoamylase cDNA in a laboratory and an industrial strain of *Saccharomyces cerevisiae*. The cDNA is first cloned into a yeast expression plasmid (using an *ENO1* promoter [P] and transcription termination [TT]) that is used is used to transform a laboratory strain of yeast. The promoter is modified, the yeast origin of replication (*ori*) is deleted, and a yeast chromosomal integrating site is added. The modified vector is integrated into the chromosomal DNA of an industrial strain of yeast.

high levels of alcohol was used as the host cell. The integrating vector was used to transform this yeast strain.

As a result of these modifications, investigators created two novel yeast strains that performed better than a naturally occurring amylolytic (starch-hydrolyzing) yeast, *Saccharomyces diastaticus*, which is closely related to *S. cerevisiae* and can hydrolyze and ferment soluble starch (Table 8.7). The performance of the brewer's strain, which contained the integrated glucoamylase gene, was superior to that of the laboratory strain with the same gene on a multicopy plasmid. This difference may reflect plasmid instability with the concomitant loss of the introduced glucoamylase gene. Neither the laboratory strain nor the brewer's strain of *S. cerevisiae* was able to utilize soluble starch unless the strain was transformed with the

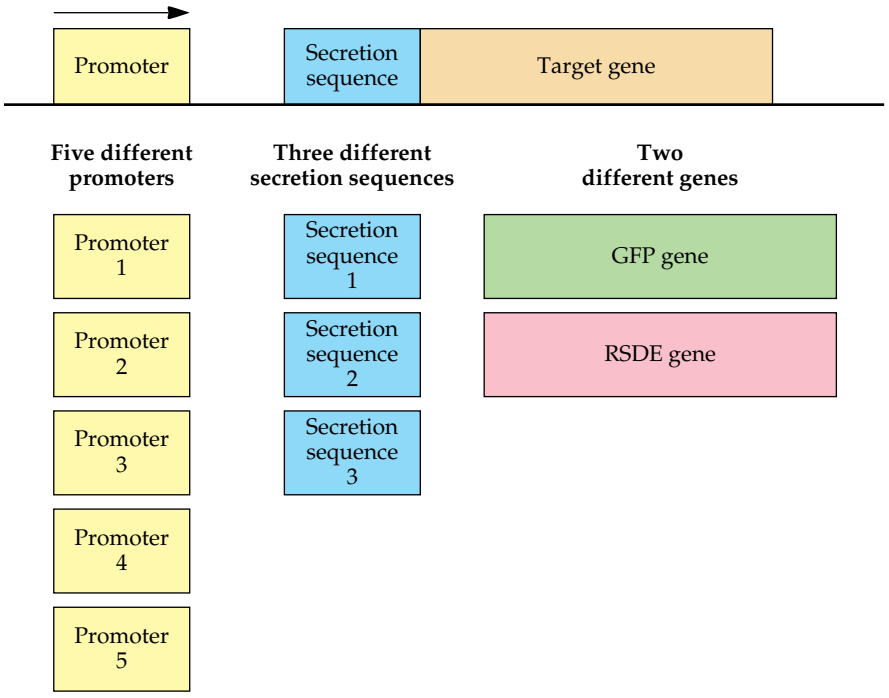
Table 8.7 Fermentation of soluble starch (25% [wt/vol]) by various yeast strains

Strain	Carbohydrate utilized (%)	Ethanol produced (g/liter)	Ethanol yield (g/g of substrate)
Laboratory yeast	5	<0.1	0
Laboratory yeast + glucoamylase cDNA on a plasmid	68	75.6	0.41
Brewer's yeast	<1	3.1	0
Brewer's yeast + integrated glucoamylase cDNA	93	118.2	0.48
S. diastaticus (naturally amylolytic yeast)	43	44.2	0.38

cloned glucoamylase gene. In both the plasmid and integrated forms, the *A. awamori* glucoamylase cDNA was under the control of regulatory signals of *ENO1* from which the 175-bp negative regulatory region had been removed. When it was tested, in the strain with the plasmid, the plasmid had to be maintained under selective pressure.

It would be advantageous if, instead of using two different enzymes to digest starch before converting it to ethanol or fructose, the starch could be completely and rapidly converted to glucose in a single step. Moreover, it would be a significant energy savings if raw starch could be converted to glucose without having to employ the energy-intensive gelatinization step. In fact (as shown in the last line of Table 8.7), a number of organisms (including some strains of *Bacillus*, yeast, and fungi) contain enzymes that have the ability to degrade raw starch without any pretreatment. Unfortunately, most of these raw starch-degrading enzymes (RSDEs) are produced in relatively low levels. To overcome this problem, scientists increased the expression of an RSDE enzyme (a glucoamylase) that is synthesized by the fungus *Penicillium oxalicum*. These researchers first tested five strong promoters from this fungus and used these promoters to direct the expression of the green fluorescent protein gene (Fig. 8.48). This screening led to the selection of an inducible promoter of an endoglucanase gene (which was significantly more active than the second most active promoter). Then, three different secretion signal sequences were tested, in conjunction with the selected promoter, for their ability to produce secreted green fluorescent protein. In this case the signal peptide of

Figure 8.48 Schematic representation of the protocol used to optimize the expression of a raw starch-degrading enzyme (RSDE) with glucoamylase activity. Five different *Penicillium* promoters and three different secretion signals were tested and used to direct the synthesis of a green fluorescent protein (GFP) gene. The best promoter and secretion signal was used to direct the synthesis of the RSDE gene.



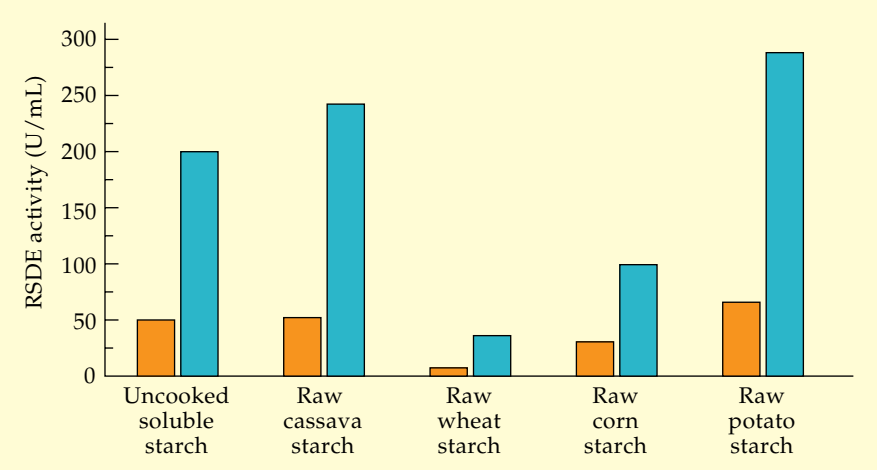


Figure 8.49 Use of the unpurified enzyme produced as described in Fig. 8.48 to digest various raw starches (blue bars) compared to the unpurified enzyme from the wild-type fungus (orange bars).

a raw starch-degrading glucoamylase was found to be the most effective of the three that were tested. Finally, the gene encoding the target RSDE was placed under the control of the selected promoter and secretion signal. With this construct, a crude enzyme extract from the engineered fungal strain exhibited a high level of activity toward raw starch from several different sources, including cassava, wheat, corn, and potato (Fig. 8.49). This experiment suggests that, while this system still requires additional improvements, these genetic manipulations may eventually provide the basis for an alternative to traditional energy-intensive methods for producing glucose from raw starch.

In an effort to produce large amounts of glucoamylase, several copies of the glucoamylase gene were integrated into the chromosomal DNA of the fungus *A. niger*. Surprisingly, there was no correlation between the number of copies of the glucoamylase gene in the chromosomal DNA and the amount of measured enzyme activity. On the other hand, the level of enzyme activity was strongly dependent upon the sites where the genes were inserted. Thus, merely increasing the gene copy number was not sufficient to produce a larger amount of active enzyme.

While most strains of the yeast *S. cerevisiae* encode only an intracellular glucoamylase, which is generally expressed only during sporulation, a few strains encode a secreted glucoamylase. Unfortunately, none of these enzymes contains a starch-binding domain. Thus, while *S. cerevisiae* glucoamylases can hydrolyze soluble starch (dextrins), they are unable to break down the larger molecules of insoluble starch. In an attempt to construct a strain of *S. cerevisiae* that could hydrolyze insoluble starch and therefore be of use in a variety of industrial processes, the coding region of a secreted *S. cerevisiae* glucoamylase gene was fused to a DNA fragment encoding a starch-binding domain from the fungus *A. niger* (Fig. 8.50). When this construct was expressed in *S. cerevisiae*, the presence of the starch-binding domain increased the ability of the enzyme to degrade insoluble starch by about 6-fold. While this system is far from optimized, this simple genetic manipulation is an important initial step in the development of more efficient industrial systems for the production of alcohol from starch.

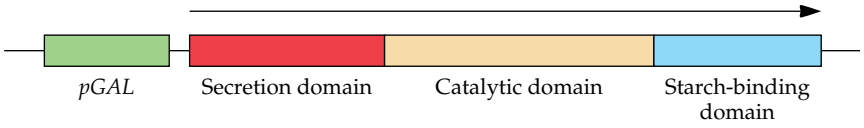


Figure 8.50 The genetic construct used to produce in *S. cerevisiae* a secreted glucoamylase containing a starch-binding domain (from *A. niger*). The construct is transcribed under the control of the yeast *pGAL* promoter in the direction indicated by the arrow.

Alcohol Tolerance

Conventional mutagenesis and selection have historically been used to improve the useful behavior of a range of microorganisms. This approach does not predetermine which gene(s) will be altered; however, it is slow and tedious, requiring an inordinate amount of testing of mutated strains. Moreover, mutagenesis and selection introduce only a limited number of changes at each round of mutagenesis. Sometimes changing a fundamental property of a microorganism may require altering the expression of many, seemingly unconnected, genes. The reprogramming of a microorganism also may be realized by mutagenizing one of the proteins that is responsible for regulating global transcription of the microorganism.

The toxicity of ethanol and other alcohols to the yeast *S. cerevisiae* is a primary factor limiting the titer and productivity of industrial production. However, in order to engineer yeast to more efficiently produce alcohol, it is necessary that the yeast strain be able to tolerate high concentrations of both glucose and ethanol. To achieve this end, it is likely that the expression of a large number of yeast proteins must be altered. Moreover, it is by no means clear which proteins need to have their levels of expression either increased or decreased. Instead, one group of researchers undertook to reprogram a significant portion of yeast metabolism by generating a large number of randomly mutated yeast transcription factor genes. In a small number of cases, the modified transcription factor will alter yeast gene expression in a manner that increases the tolerance of yeast for high levels of both glucose and ethanol. In yeast, 15 different proteins bind to DNA and regulate the promoter specificity of RNA polymerase II. To dramatically modify yeast metabolism, the gene encoding one of these transcription factors, *SPT15*, was first isolated and then altered by error-prone PCR (Fig. 8.51). This reaction randomly introduces a range of different mutations in the *SPT15* gene. The mutagenized *SPT15* genes were cloned into a plasmid vector and introduced into wild-type yeast (*S. cerevisiae*) cells. The transformed cells were grown on agar medium in the presence of 6% ethanol and 120 g/liter of glucose. Any transformant that grew better on this medium than the native yeast strain was then more fully characterized. One particular mutant displayed a higher level of cell viability and grew much faster than the native yeast strain at ethanol concentrations of between 10 and 20%. When this strain was examined in detail, it was found that the *SPT15* gene had three separate mutations, all of which were required for this activity. In addition, when the mutant was transcriptionally profiled (using microarray technology; chapter 2), several hundred genes were found to be differentially regulated compared to the native strain, with the majority of genes being upregulated. Interestingly, a detailed analysis of the genes whose expression was significantly altered did not reveal that a particular pathway or genetic network was primarily

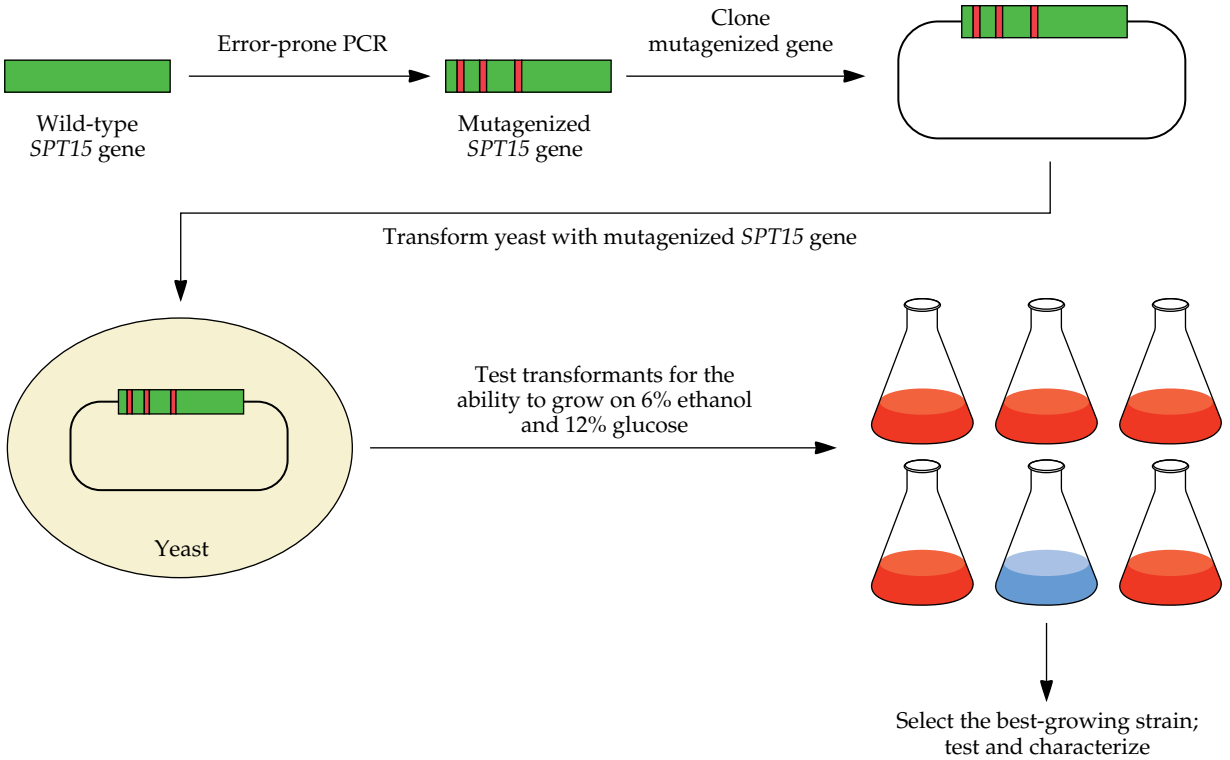


Figure 8.51 Engineering the yeast *S. cerevisiae* to tolerate high levels of glucose and ethanol. The yeast transcription factor *SPT15* gene was mutagenized by error-prone PCR, cloned into a plasmid vector, and used to transform a laboratory yeast strain. The transformants were selected for the ability to grow on medium containing both 6% ethanol and 12% glucose. The portions of the *SPT15* gene shown in red represent the introduced mutations.

responsible for the observed reprogramming. Finally, the yeast strain that was used in these experiments was a standard laboratory strain, so that to incorporate this approach into a commercial process, it will be necessary to repeat the work with an industrial strain of yeast.

Another way to protect *S. cerevisiae* from inhibition by high concentrations of alcohol might be to alter the chemical composition of the yeast plasma membrane, which is often sensitive to high concentrations of alcohol. In one series of experiments, researchers added various salts to high-cell-density and high-glucose cultures, thereby modifying the yeast plasma membrane. They observed large improvements in ethanol productivity as a result of increases in the K^+ concentration and in the pH of the fermentation medium. In fact, these conditions resulted in the complete utilization of the added glucose and a concomitant decrease in the formation of acetic acid and glycerol, two unwanted fermentation by-products. This increase in ethanol productivity resulted from an increase in yeast cell ethanol tolerance and occurred under both aerobic and anaerobic conditions and in a wide range of both laboratory and industrial yeast strains. The conclusion from these experiments was that, independent of any genetic changes, appropriate medium adjustment can facilitate a significant increase in ethanol tolerance in ordinary yeast strains.

Researchers then reasoned that genetic modifications of cellular ATP-dependent pumps might increase yeast alcohol tolerance similar to that

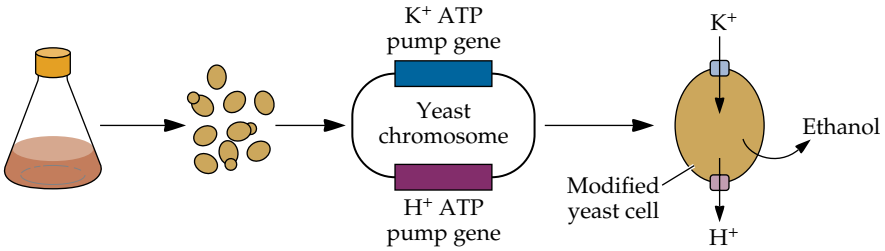


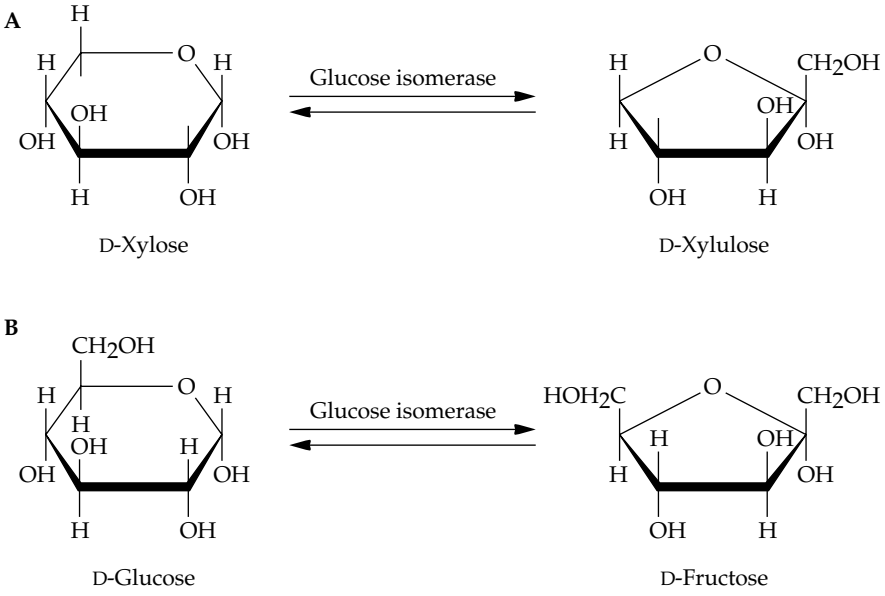
Figure 8.52 Schematic representation of the genetic modification of yeast cells to increase alcohol tolerance and alcohol productivity. Genes encoding a yeast K^+ ATP pump and a yeast H^+ ATP pump are isolated, genetically modified to increase their efficiency, and then reinserted into chromosomal DNA in place of the wild-type genes. The net result is an increase in K^+ and H^+ flux and a subsequent increase in alcohol tolerance, thereby allowing increased alcohol productivity.

observed following increases in growth medium K^+ and pH. In fact, it was found that by genetically modifying the genes encoding the K^+ and H^+ transporters to upregulate their expression, the resultant modified yeast strains had increased the ethanol titers, to the point where a laboratory strain of yeast that normally produced a low level of ethanol now produced a level of ethanol that exceeded that of several higher-ethanol-producing industrial strains (Fig. 8.52).

Improving Fructose Production

The enzyme glucose isomerase should really be called xylose/glucose isomerase, because it primarily catalyzes the conversion of the five-carbon sugar D-xylose to D-xylulose, with the conversion of D-glucose to D-fructose being a secondary or side reaction of this enzyme (Fig. 8.53). Kinetically, xylose/glucose isomerase has a lower k_{cat} (catalytic rate

Figure 8.53 Conversion of **(A)** xylose to xylulose and **(B)** glucose to fructose, catalyzed by glucose isomerase.



constant) and a higher K_m (binding constant) for glucose than for xylose, which means that xylose is bound more tightly to the enzyme than is glucose and that xylose is converted more rapidly to xylulose than glucose is converted to fructose.

Xylose/glucose isomerases are intracellular enzymes and, as such, do not yield the same quantities or purity of product as do the extracellular, or secreted, enzymes that are used in many industrial processes. Most industrial enzymes are used without any extensive purification. An extracellular enzyme preparation generally contains many fewer proteins than an intracellular extract. In addition, the preparation of an intracellular protein extract requires separation of the cells from the growth medium, mechanical disruption of the cells, and removal of cell debris following disruption. These factors lead to higher production costs for xylose/glucose isomerase than for many other industrial enzymes. One way to overcome this problem is to use a batch of xylose/glucose isomerase more than once. This recycling can be achieved by immobilizing the enzyme on a solid support, which both stabilizes the enzyme and facilitates its reuse.

The isomerization of glucose to fructose is a reversible reaction, and the final fructose content is dependent on the reaction temperature. The higher the temperature, the greater the fructose content in the final product. Most commercial processes use conversion temperatures of around 60°C. The enzyme is typically used in an immobilized state that is obtained by cross-linking it to itself with glutaraldehyde and then using it in a continuous process in a packed bed reactor. Under these conditions, a batch of enzyme can be used for approximately 150 to 200 days before it is discarded. Consequently, increasing the temperature optimum for the enzymatic activity and the thermostability of xylose/glucose isomerase is one way to make it more efficient.

The thermophilic bacterium *Thermus thermophilus* produces a xylose/glucose isomerase that not only is active at 95°C, but also is very stable at high temperatures. Therefore, this enzyme is a good candidate for use in industrial processes. Unfortunately, wild-type *T. thermophilus* does not produce large amounts of the enzyme. To circumvent this problem, the *T. thermophilus* xylose/glucose isomerase gene was isolated and expressed in *E. coli* and *Bacillus brevis* on plasmids with different copy numbers, and under the control of various promoters and ribosome-binding sites. One of the recombinant strains overproduced xylose/glucose isomerase more than 1,200-fold relative to the amount found in the original wild-type *T. thermophilus*. The host organism was *B. brevis*, the plasmid copy number per cell was approximately 20, the constitutive promoter was from *B. brevis*, and the ribosome-binding site was from *T. thermophilus*. Therefore, with this strain, high yields of a thermostable xylose/glucose isomerase can be produced for the industrial synthesis of fructose from glucose. In fact, one company has commercialized and is currently selling an immobilized thermostable xylose/glucose isomerase with a temperature optimum of 90°C, under the name Sweetzyme, that is used in the production of high-fructose corn syrup.

Utilization of Cellulose and Hemicellulose

With an increasing world population and dramatic increases in the standard of living in parts of the developing world, meeting the growing worldwide demand for energy for heating, transportation, and industry has become a

major challenge for all countries for the 21st century and beyond. In addition to being sustainable, the future energy supply needs to be nonpolluting so that we may realize a reduction in greenhouse gas emissions. To this end, many countries around the world have begun to produce large amounts of alternative fuels in an effort to at least partially replace nonrenewable fossil fuels, such as oil and gas. In this regard, a major effort has been directed toward producing bioethanol. The United States produced 15.8 billion gallons and Brazil produced 7.1 billion gallons of ethanol in 2017. At the present time, Brazil produces large amounts of ethanol from the fermentation of sucrose derived from sugarcane, and the United States produces ethanol from corn starch. However, the reduction of greenhouse gases that results from the use of sugar- or starch-based ethanol is not as high as desired, and many socially conscious individuals have criticized the strategy of converting land from the production of food to the production of ethanol/energy. Moreover, sugar- or starch-based ethanol is unlikely to provide more than a small fraction of what we require. In this regard, in 2008, the Chinese government announced that it would not allow any further increase in starch-based ethanol production because of competing uses as food. If the world is going to produce ethanol on a large enough scale to significantly lower our use of fossil fuels, that ethanol will have to be produced from lignocellulosic waste products, such as corn stover, barley straw, oat straw, wheat straw, sorghum straw, bagasse (the fibrous matter that remains after sugarcane stalks are crushed to extract their juice), and wood chips. Thus, there is now, more than ever before, a tremendous amount of both political and scientific activity directed toward trying to produce ethanol from lignocellulosic materials.

Lignocellulosics

The polymers lignin, hemicellulose, and cellulose combine in various proportions to form a "lignocellulosic" structural support system for nearly all terrestrial plants (Fig. 8.54; Table 8.8). This material constitutes a vast biomass that is often a waste product of agriculture and timber processing and needs to be disposed of in a safe and efficient manner or used as a resource. It has been estimated that in 2016 there was more than 2.0×10^{13} kg of lignocellulosic wastes produced by different industries in the biosphere, with an energy content that is equivalent to around 600 billion liters of ethanol.

Lignin is a three-dimensional, globular, irregular, insoluble, high-molecular-weight (>10,000 Da) polymer made up of phenylpropane subunits with no chains of regular repeating units or any bonds that are easily hydrolyzed either enzymatically or chemically (Fig. 8.54). The lignin polymer molecule has many different types of chemical linkages between aromatic phenylpropane units. The physical and chemical characteristics of lignin are generally attributed to the last step in lignin biosynthesis, the nonenzymatic free-radical-based joining of the phenylpropane units in a more or less random fashion. In plants, lignin is chemically (covalently) bonded to hemicellulose and wraps around fibers composed of cellulose. Lignin is responsible for the rigidity of plants and for their resistance to mechanical stress and microbial attack.

Hemicelluloses are short-chain, heterogeneous polymers that contain both hexoses (six-carbon sugars, such as glucose, mannose, and galactose) and pentoses (five-carbon sugars, such as xylose and arabinose) (Fig. 8.54). The three major types of hemicelluloses are xylans, which

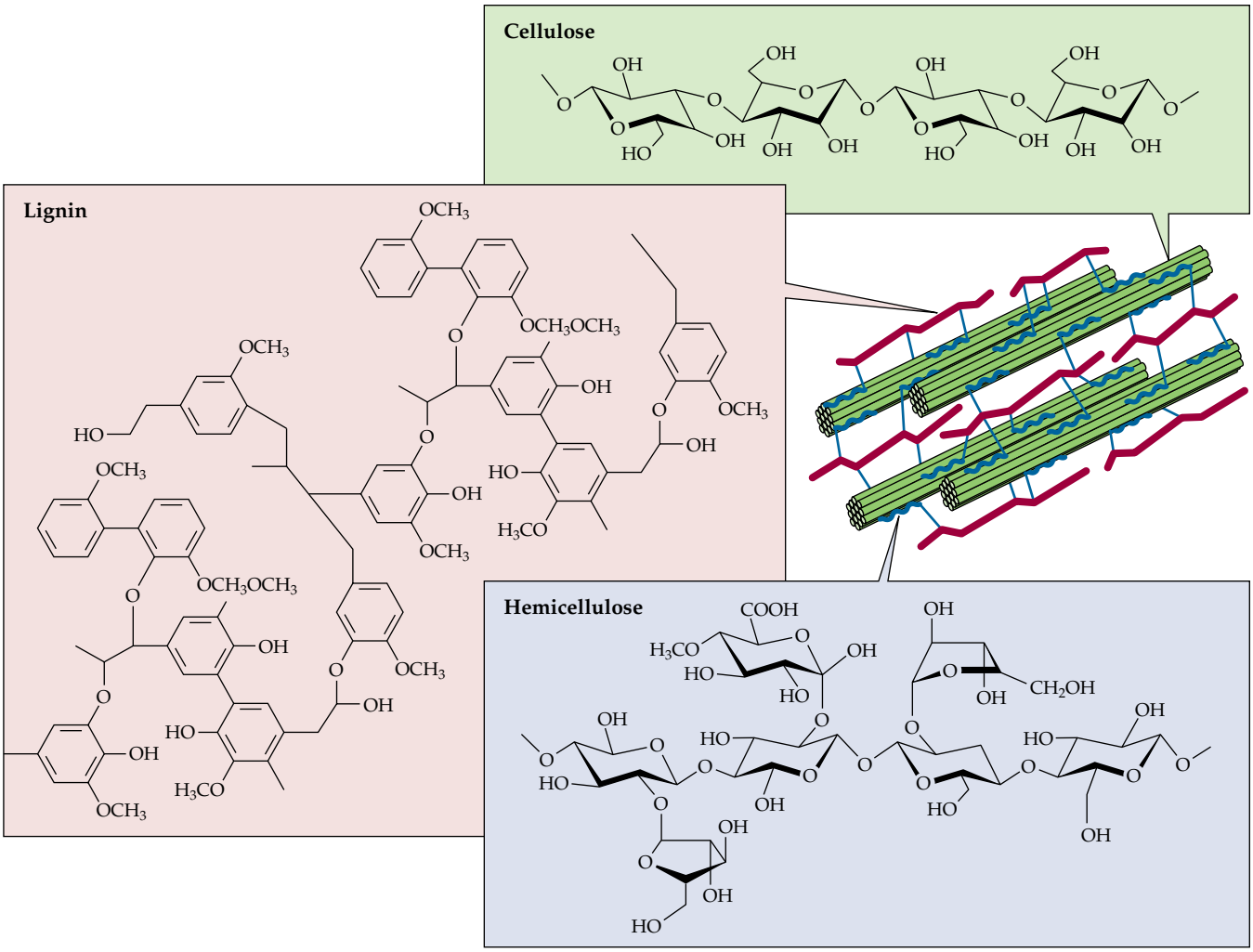


Figure 8.54 Schematic representation of the organization of the lignocellulosic material (lignin, cellulose, and hemicellulose) that makes up a plant cell wall.

have a backbone of poly-β-1,4-xylan, with side links to arabinose, glucuronic acid, and arabino-glucuronic acid; mannans, which are composed of glucomannans and galactomannans; and arabinogalactans. The origin of the lignocellulosic material usually defines the nature of the hemicelluloses. For example, xylan hemicellulose is particularly common in hardwoods, and glucomannans are characteristic of softwoods (Fig. 8.55). Xylanases that predominantly break down the β-1,4-linked bonds in the

Table 8.8 Lignocellulosic biomass composition from different sources

Biomass source	% Cellulose	% Hemicellulose	% Lignin
Hardwood trees	40.5–54.0	18.5–36.0	15.5–24.0
Softwood trees	42.0–50.0	11.0–27.0	20.0–28.0
Agricultural waste	25.0–45.0	12.0–36.0	7.0–25.0
Grasses	25.0–40.0	25.0–50.0	10.0–30.0

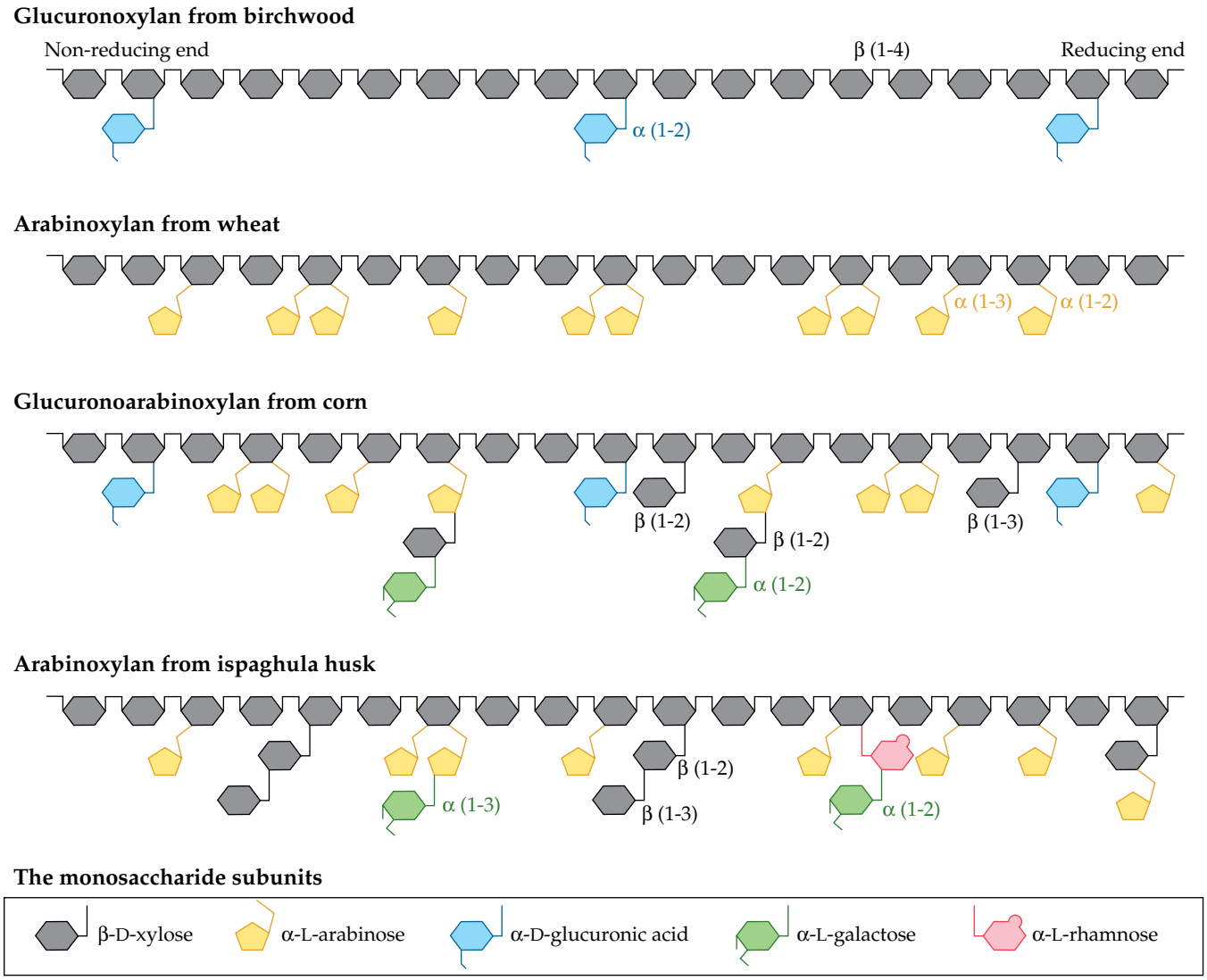


Figure 8.55 Schematic representation of the structures of the main xylans. For clarity, the acetyl, methyl, and feruloyl side chains are not shown. Ispaghula (also psyllium) is a common name for members of the plant genus *Plantago*.

backbone of the xylan polymer into xylose and other sugars (xylanases) are produced by a wide range of organisms, including some fungi, bacteria, yeast, algae, protozoa, and plants. The digestion of xylans by filamentous fungi is typically a component of the chlorine-free bleaching of wood pulp prior to papermaking and the partial digestion of silage prior to composting.

Cellulose, which is the simplest of the components found in lignocellulosic material, is the most abundant polymer in the biosphere. It is composed of long chains of D-glucose molecules linked by β-1,4 glycosidic bonds (Fig. 8.54). Both cellulose and starch can be hydrolyzed to glucose, but their structures are very different, with starch containing only α-1,4 glycosidic bonds. Starch is an energy storage molecule in which the glucose residues are linked in a manner that prevents a tightly ordered arrangement

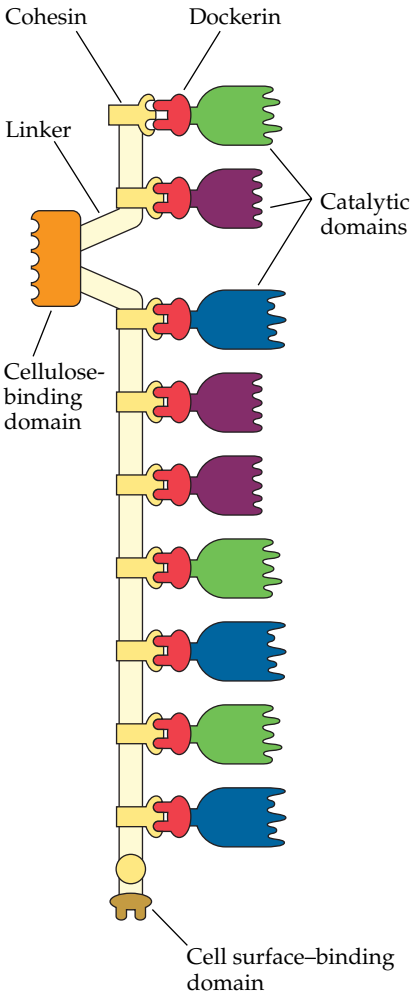


Figure 8.56 Schematic representation of a cellulosome protein complex. This complex contains both structural and enzymatically active protein components. The structural components include scaffoldin, a scaffolding protein (shown in yellow) that contains a strong cellulose-binding module (orange), and a number of cohesin molecules (nine are shown). The cohesin molecules act as binding sites (docking sites) for the dockerin proteins (red), which are attached to the catalytic domains (the active enzymatic components, which are shown in green, purple, and blue). Each cellulosome complex is anchored to the microbial cell surface through the cell surface-binding domain (brown).

of the polymer chains. This open mesh-like structure is easily penetrated by water; as a result, many starches are both water soluble and readily hydrolyzable by enzymes. In contrast, cellulose is a plant-supporting structural molecule. The glucose chains in cellulose are arranged in a manner that permits them to pack together in a crystal-like structure cross-linked by hydrogen bonds that is impervious to water. Consequently, the cellulose polymer is both insoluble and resistant to hydrolysis.

Nevertheless, cellulose is still a form of stored glucose, so it is the component of lignocellulosics that has the most potential for conversion into a variety of useful compounds, such as alcohol. However, before cellulose can be utilized, it must be released from its complex with lignin and hemicellulose. For most lignocellulosic materials, this separation requires treatment with either a strong acid or a strong base or the use of high temperature and pressure. Regardless of how the cellulose is separated from the lignocellulose complex, the energy that is necessary to achieve this adds significantly to the cost of the final product. However, since the annual global production of lignocellulosic materials is huge, effective ways of enzymatically degrading cellulose and hemicellulose are being sought. Chemical and enzymatic methods for the selective degradation of lignin also are being investigated, with less success.

Cellulase Genes

A wide range of bacteria and fungi are naturally capable of degrading cellulose through the concerted action of several different enzymes that collectively are referred to as cellulase. Aerobic microorganisms typically secrete large amounts of these cellulase enzymes into the medium outside the cell. On the other hand, in anaerobic microorganisms, cellulase activity is often found as part of a multiprotein complex that is called a cellulosome (Fig. 8.56) that lies on the external surface of the cell. Cellulases consist of multiple copies of several enzymes with different enzymatic activities, including the following.

- Endoglucanase, which hydrolyzes β-1,4 linkages between adjacent glucose molecules within the amorphous (loosely packed) regions of the cellulose polymer, thereby breaking the chain in the middle
- Exoglucanase, which degrades the nicked cellulose chains from their nonreducing ends (the end capable of acting as a reducing agent) and produces glucose, cellobiose (two glucose units), and cellotriose (three glucose units)
- Cellobiohydrolase, which is often found in cellulolytic fungi and is a type of exoglucanase that removes units of 10 or more glucose residues from the nonreducing ends of the cellulose molecule
- β-Glucosidase, or cellobiase, which converts cellobiose and cellotriose to glucose

Isolating cellulase genes

The breakdown of cellulose by microorganisms (either bacteria or fungi) that produce the various components of cellulase (Fig. 8.57) is slow and often incomplete. Therefore, genetic engineering strategies have been used in an attempt to create organisms with more effective cellulase activity. For this purpose, genes coding for the individual enzymatic functions of cellulase activity have been isolated from both prokaryotic and eukaryotic organisms.

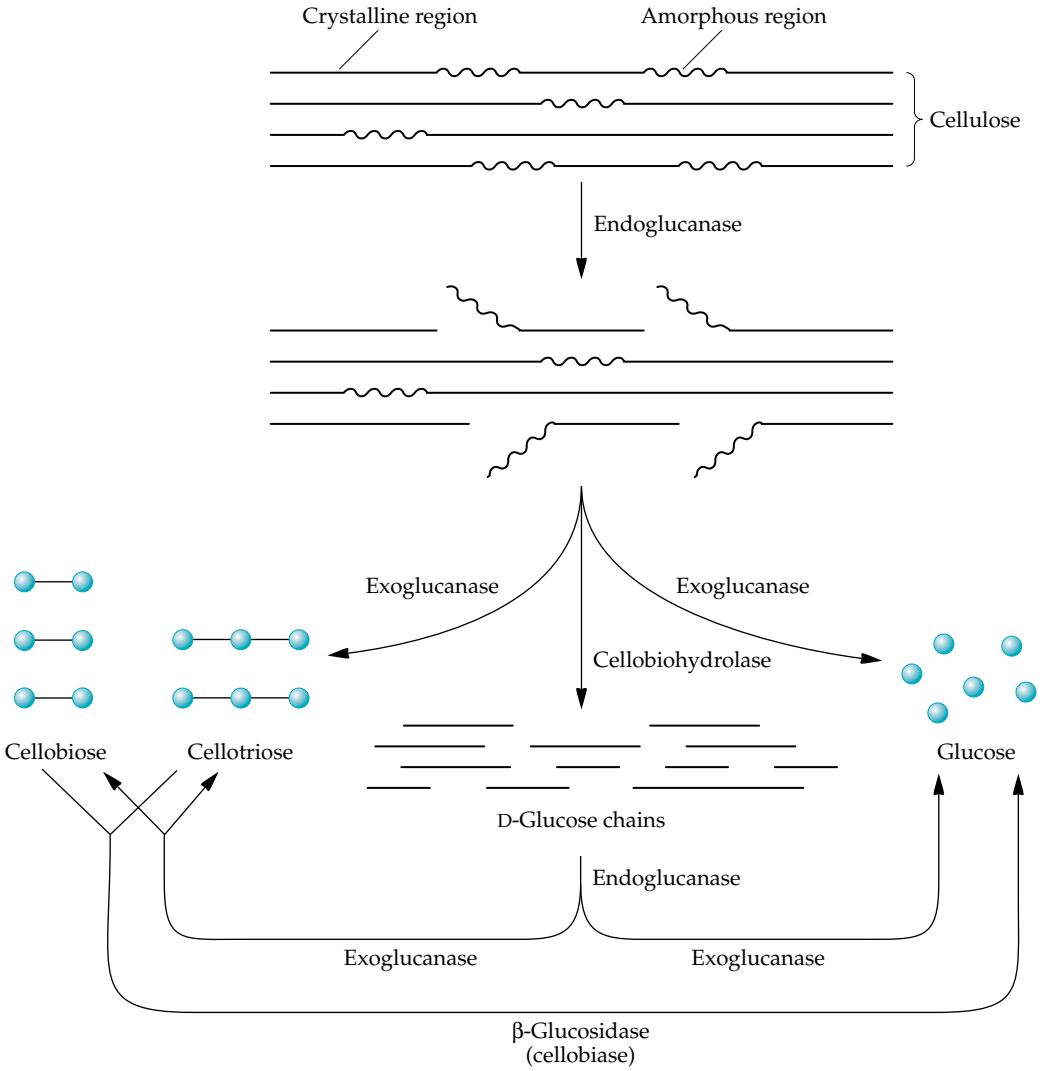


Figure 8.57 Enzymatic biodegradation of cellulose. Cellulose hydrolysis begins with the cleavage of β-1,4 linkages within the accessible amorphous regions of the cellulose chains by endoglucanase(s). This reaction is followed by the removal of oligosaccharides from the partially cleaved cellulose chains by exoglucanase(s) and cellobiohydrolase(s). The degradation of cellulose is completed when the cellobiose and cellotriose are converted to glucose by β-glucosidase.

For a start, constructing a library of the chromosomal DNA of a cellulolytic prokaryote is quite straightforward, employing standard techniques and approaches that might be used in the isolation of any prokaryotic gene (chapter 2). However, screening the genomic library of a cellulolytic prokaryote for different types of cellulose requires a variety of strategies. For example, prokaryotic endoglucanase genes have been cloned by the following simple yet effective identification technique.

1. A genomic DNA library from a cellulolytic prokaryote is constructed in *E. coli* (which cannot digest cellulose), and the host cells are grown overnight on solid medium containing an antibiotic that can be used to select for the presence of the plasmid.

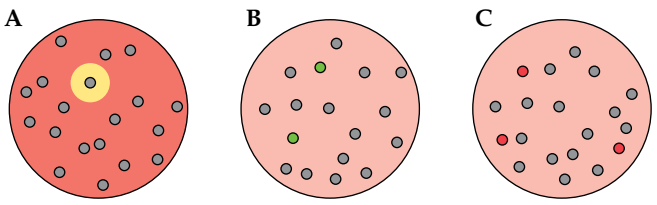
2. The colonies are then overlaid with agar containing carboxymethyl cellulose (CMC), a soluble derivative of cellulose, and the petri plates are incubated at 37°C for several hours. During this time, the CMC molecules that are present in the immediate vicinity of a colony that both synthesizes and secretes an endoglucanase are partially digested. Transformants that synthesize but do not secrete the cloned endoglucanase are not able to degrade the substrate, because it is too large to enter the cell.
3. The digested regions of the CMC are visualized by first flooding the petri plate with a solution of the dye Congo red, which is not toxic to the bacteria, followed by a wash with a solution of sodium chloride. Congo red selectively binds to high-molecular-weight cellulose chains and gives a red color; conversely, it binds weakly to low-molecular-weight polysaccharides and produces a yellow hue. The sodium chloride treatment stabilizes the binding of the dye. If a bacterial colony produces a secreted endoglucanase, it will be surrounded by a yellow halo; the background, where the CMC has not been degraded, will be red (Fig. 8.58A).

There is no convenient plate-based enzyme activity assay for detecting cells that express and secrete an exo-glucanase from a cloned gene, so immunological screening has been used to pick colonies of recombinant bacteria that produce exoglucanase. In this example, the exoglucanase-positive colonies appear as green because of the use of a chromogenic substrate that is produced by the immunological assay (Fig. 8.58B).

Prokaryotic β-glucosidase genes have been isolated by first generating a genomic DNA library from a β-glucosidase-producing bacterium in *E. coli*. The library is then tested for clones that express β-glucosidase activity that can be detected with a chromogenic substrate, such as 5-bromo-4-chloro-3 -indolyl-β-D-glucopyranoside. In this example, when the library is plated on MacConkey–cellobiose agar, the β-glucosidase-positive colonies turn red (Fig. 8.58C).

The strategy of DNA hybridization screening of either cDNA or genomic DNA libraries with a heterologous probe has not been particularly effective for isolating cellulase genes, because the sequences of cellulolytic enzymes from diverse microbial sources are often not very similar. To isolate the messenger RNAs (mRNAs) that encode cellulolytic enzymes from fungi or plants, a novel protocol had to be implemented. Unfortunately, these mRNAs usually constitute only a small fraction of the total mRNA

Figure 8.58 Detection of a recombinant *E. coli* clone expressing a bacterial **(A)** endoglucanase gene, where the yellow halo on a red background indicates the presence of a positive clone that degrades soluble cellulose (CMC) in the growth medium in the vicinity of the clone via the secreted endoglucanase; **(B)** exoglucanase gene, where the green color indicates a positive immunological response to the presence of the exoglucanase; and **(C)** β-glucosidase gene, where the red color indicates the presence of an enzyme that can break down cellobiose.



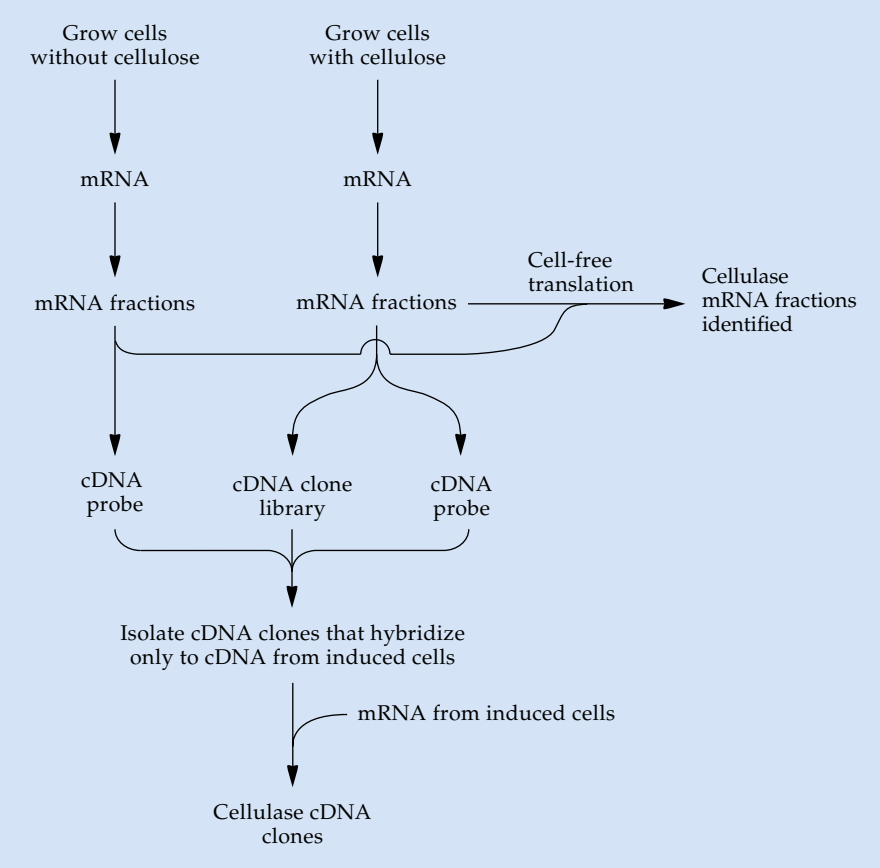


Figure 8.59 Differential hybridization for isolating eukaryotic cellulase cDNA clones.

population. Therefore, it is necessary to enrich for the target mRNA or cDNA and to eliminate cDNA clones that do not carry the target sequences. To meet these ends, the technique of "differential hybridization" has been used for the isolation of a number of different induced eukaryotic cellulase genes, as follows (Fig. 8.59).

1. mRNA is separately isolated both from cells grown without cellulose (i.e., noninduced cells) and from cells grown in the presence of cellulose to enhance the expression of cellulase enzymes (i.e., induced cells).
2. Each mRNA population is fractionated on a sucrose gradient, and each fraction is translated in a eukaryotic cell-free extract that is able to translate added eukaryotic mRNA into protein and is derived from either rabbit reticulocyte or wheat germ extracts. A cellulose-induced protein(s) is identified, after separation of the cell-free translation products on a polyacrylamide gel, by the presence of unique bands that appear from the induced cells but not from the noninduced cells (Fig. 8.60). This step indicates which fractions contain mRNAs that are induced by the addition of cellulose.
3. The mRNA sucrose gradient fractions from cells that direct the synthesis of cellulose-induced proteins and the comparable fractions from the noninduced cells are used separately as templates for the synthesis of cDNA.

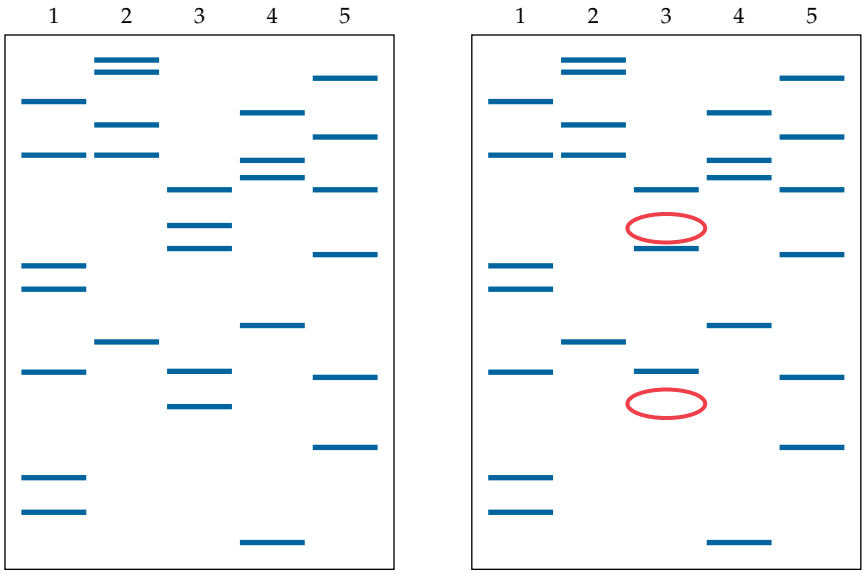


Figure 8.60 Schematic representation of stained polyacrylamide gels of translation products of mRNA fractions, i.e., proteins, following growth of cells either with or without cellulose. The numbers represent the fraction number following sucrose gradient centrifugation of mRNA. The red ovals on the gel without cellulose indicate where bands appear when there is growth in the presence of cellulose.

4. The cDNAs from the cellulose-induced mRNAs are cloned into a plasmid vector, introduced into *E. coli*, replica plated, and then separately screened with labeled cDNA from both the induced and noninduced fractions as hybridization probes. Clones that hybridize only with the cDNA from the induced cells and not with cDNA from the noninduced cells potentially carry cellulose-induced genes and are characterized further.
5. To establish conclusively which of the positive cDNA clones encode cellulase enzymes, the DNA from these clones is introduced into an *E. coli* expression vector, and after *E. coli* is transformed with these constructs, the protein products are detected with antibodies to the enzymes of the cellulase complex.
6. The sequence of each of the positive cDNA clones is determined. In principle, this scheme can be used for the isolation of any induced eukaryotic gene(s).

In one series of experiments, endoglucanase and exoglucanase genes from the bacterium *Cellulomonas fimi* were each put under the control of an *S. cerevisiae* promoter and signal peptide sequence, subcloned onto the same plasmid vector, and introduced into *S. cerevisiae*, which does not normally degrade cellulose. Some transformants secreted about 70% of the activity of each of these enzymes into the growth medium and were able to degrade the cellulose in filter paper and pretreated wood chips; the remainder of the activity was found in the cell cytoplasm. The rate and extent of hydrolysis of both of these substrates were increased by adding β-glucosidase to the mixture, thereby decreasing the amount of cellobiose that accumulated and preventing end-product (feedback) inhibition

of the endoglucanase and exoglucanase activities by cellobiose (Fig. 8.61). Cellobiose acts as a feedback inhibitor of cellulose hydrolysis, and glucose inhibits cellobiose cleavage. These two regulatory mechanisms may prevent complete enzymatic breakdown of cellulose. Instead of adding β-glucosidase to the medium, a β-glucosidase gene could be cloned into the host cell. To this end, a β-glucosidase gene was isolated from the cellulolytic fungus *Trichoderma reesei*, cloned onto a multiple-copy plasmid, and reintroduced into *T. reesei*. The transformed strain overproduced β-glucosidase 5.5-fold, and it degraded microcrystalline cellulose (Avicel), a cellulose derivative, 33% faster than the nontransformed strain. In addition, when a β-glucosidase gene from the yeast *Saccharomycopsis fibuligera* was expressed in *S. cerevisiae*, the transformed strain directed the enzyme to the periplasm. The β-glucosidase-producing *S. cerevisiae* strain was nearly as efficient at utilizing cellobiose to produce ethanol as the nontransformed *S. cerevisiae* strain was at producing alcohol from glucose. Thus, the expression of a recombinant β-glucosidase gene enhances the enzymatic utilization of cellulose and suggests a simple strategy for genetically engineering more effective cellulolytic alcohol-producing microorganisms.

The participation of cellulase enzymes in an industrial process for the bioconversion of wastepaper to alcohol has also been examined. Wastepaper was partially digested by the addition of cellulase enzymes at 45°C; then, the released glucose was fermented by *S. cerevisiae* at 37°C. By extrapolation of small-scale results, yields of 400 liters of ethanol per ton of wastepaper were estimated. If all 100 million tons of wastepaper generated annually in North America were converted into ethanol and used as fuel, approximately 16% of the gasoline that is currently being used in North America could be saved.

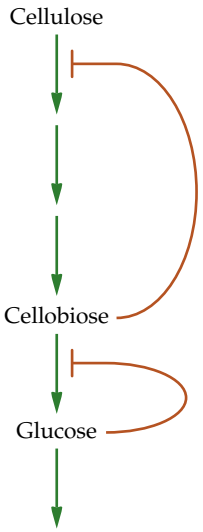


Figure 8.61 Metabolic control of cellulose hydrolysis. The green arrows indicate cellulose degradation. The red lines indicate feedback inhibition by a particular metabolite of cellulose degradation. Relief of feedback inhibition may come from overproducing β-glucosidase, thereby preventing the accumulation of cellobiose, and removal of glucose by converting it into other metabolites.

Designer cellulosomes

Scientists hoping to exploit the activity of microbial cellulosomes to degrade cellulosic waste materials have attempted to manipulate some of the genes involved in the formation of this complex and create designer cellulosomes whose degradative activities are directed toward specific substrates. One of the key ingredients in the assembly and functioning of a cellulosome is the calcium-dependent high affinity (~10^9 M^{-1}) of the cohesin domain of the scaffoldin molecule for the dockerin domain (Fig. 8.56). Unfortunately, within a given species of microorganism, the dockerin domain binds to all of the cohesin domains with the same affinity. Thus, in nature, different catalytic domains attached to the dockerins are randomly incorporated into the cellulosome. However, it is possible, in the laboratory, to engineer cellulosomes that contain certain enzymatic activities designed to facilitate the degradation of specific substrates. In one series of experiments, designer cellulosomes were constructed that were more effective than the free enzymes (i.e., the enzymatic components not assembled into a cellulosome complex) and slightly less effective than native cellulosomes in degrading crystalline cellulose. However, these designer cellulosomes were not especially active at degrading straw (an agricultural waste product). When a xylanase was incorporated by genetic engineering into the cellulosome complex, the ability of the designer cellulosome to degrade straw, which contains hemicellulose as well as cellulose, increased significantly (Fig. 8.62). In addition to engineering microbes with designer cellulosomes that have both cellulase and xylanase activity, scientists have recently reported isolating several

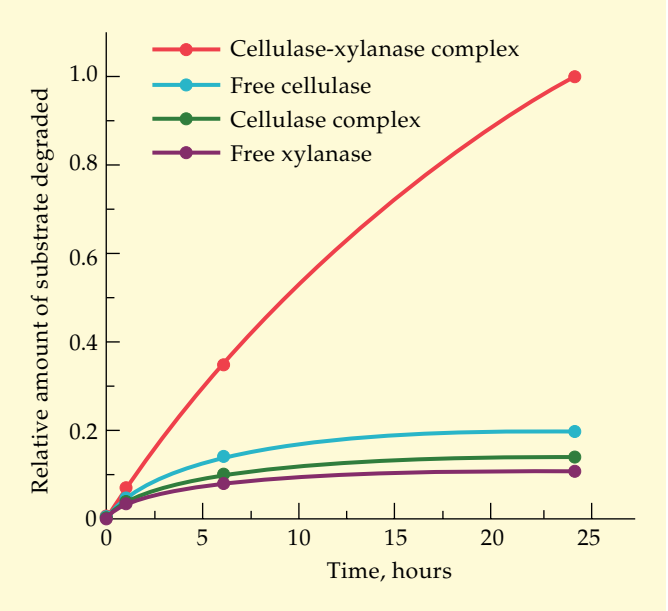
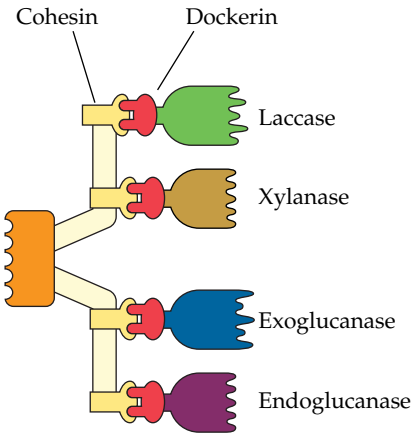


Figure 8.62 Digestion of straw by a cellulase-plus-xylanase designer cellulosome complex free cellulase enzyme, a cellulase-based designer cellulosome, and free xylanase enzyme. Although the data are presented as the relative amount of substrate degraded, in this experiment, the substrate was only partially degraded in all cases.

bacteria and fungi that contain both of these enzymes. These naturally occurring microbes are currently being tested for their effectiveness in a variety of industrial processes, including agro-waste treatment, deinking of wastepaper, biofuel production, and biomass saccharification.

The lignin that is present in most lignocellulosic material physically limits the accessibility of enzymes and chemicals and therefore hinders the breakdown of cellulose and hemicellulose into usable sugars. To increase the amount of fermentable sugars that can be released from lignocellulosics, it is necessary to at least partially remove the recalcitrant lignin molecules. With this in mind, one group of scientists designed and constructed a cellulosome, which already had cellulase and hemicellulase activity, to also contain the enzyme laccase, a copper-containing oxidase that can catalyze the oxidation of a wide range of phenolic compounds, including lignin. This designer cellulosome contained endoglucanase Cel5A from *Thermobifida fusca*, exoglucanase Cel48A from *T. fusca*, xylanase Xyn11V from *Clostridium thermocellum*, and laccase from *T. fusca* (Fig. 8.63). The aerobic bacterium *T. fusca* is typically found in decaying plant matter, where it is involved in plant cell wall degradation. This bacterium survives at both high temperature and a wide range of pH, making it ideal for use in an industrial process. The anaerobic bacterium *C. thermocellum* can also function at high temperatures. To construct a designer cellulosome with these four enzymes, it was necessary to fuse their N termini with dockerins (from either *C. thermocellum* or *Ruminococcus flavefaciens*). This designer cellulosome increased the amount of reducing sugars released from wheat straw by a factor of two compared to a designer cellulosome that did not contain the laccase enzyme and is unique in that, for the first time, a single degradative complex is able to simultaneously digest cellulose, hemicellulose, and lignin.

Figure 8.63 Schematic representation of a designer cellulosome that has been engineered to degrade all three components of plant cell walls, i.e., cellulose, hemicellulose, and lignin.



Although the designer cellulosomes that have so far been constructed are not yet ready to be used in a commercial process, researchers continue to improve their performance. For example, a designer cellulosome that is both stable and efficient at degrading cellulose at high temperatures (75°C) for extended periods of time (making it potentially useful in some industrial processes) has recently been reported. Finally, it is envisioned that various designer cellulosomes could one day be the central component of large-scale industrial processes aimed at efficiently and economically converting different types of cellulosic wastes into usable chemicals.

Modifying cellulase genes

Notwithstanding the broad spectrum of enzymes that are able to digest cellulose and xylan, from an industrial perspective most of these enzymes are rather slow and inefficient in the breakdown of these substrates. To increase the effectiveness of these enzymes, especially on an industrial scale, researchers have used a variety of techniques to improve the cellulolytic and xylanolytic activity of various microorganisms (Table 8.9). These changes were made using either rational design or directed evolution of the target enzyme (see chapter 3). Rational design requires a detailed knowledge of the protein structure, identification of the amino acid sites to be altered, and site-directed mutagenesis to introduce the desired mutation(s). Directed evolution does not require any detailed knowledge of the target enzyme and does not preselect which amino acid residues should be changed. This strategy may be achieved by DNA shuffling, gene truncation, random insertion/deletion mutagenesis, or error-prone PCR. The altered enzymes listed in Table 8.9 provide just a small example of the types of useful changes that may be introduced into cellulases and xylanases.

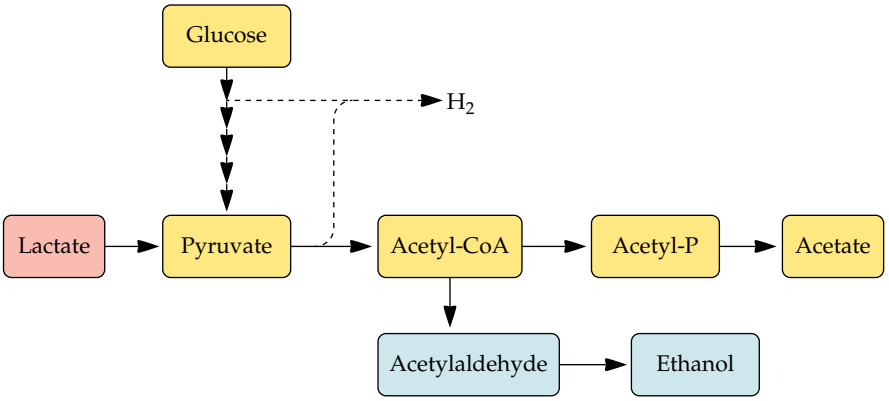
Table 8.9 Some engineered microbial cellulases and xylanases

Source organism	Enzyme	Changes observed
Fibrobacter succinogenes	1,3-1,4-α-D-glucanase	Increased turnover rate
Fibrobacter succinogenes	1,3-1,4-α-D-glucanase	Improved thermostability and specific activity
Trichoderma reesei	Endo-1,4-β-xylanase II	Improved alkali stability
Trichoderma reesei	β-glucosidase	Increased k_{cat}
Bacillus subtilis	Endo-β-1,4-xylanase	Improved acid stability
Bacillus subtilis	Endo β-glucanase Egl499	Increased thermostability
Trametes multicolor	Pyranose 2-oxidase	Increased selectivity for D-galactose and D-glucose
Geobacillus stearothermophilus	Xylanase XT6	52-fold increase in thermostability; increased catalytic efficiency
Geobacillus stearothermophilus	Xylanase XT6	Increased thermostability; increased catalytic efficiency
Hypocrea jecorina	Xylanase	Increased thermostability
Thermoascus aurantiacus	Endoglucanase	Increased k_{cat}
Clostridium thermocellum	Cellobiose phosphorylase	Increased thermostability
Clostridium thermocellum	Endoglucanase Cel8A	Increased thermostability
Clostridium thermocellum	Endoglucanase CelA	Increased thermostability
Thermobifida fusca	β-glucanase BglC	Increased thermostability

Direct Conversion of Biomass to Ethanol

Switchgrass (*Panicum virgatum*) is a perennial warm-season prairie grass that is native to North America, can grow up to 2.5 meters high on marginal farmland, and has recently been used as a feedstock for the production of biofuels. To produce ethanol from switchgrass, it is necessary to pretreat the plant material (with high temperature, acid, or alkali) to make the cellulose more readily available, to digest the cellulose to glucose by adding cellulolytic enzymes, and then to ferment the resultant glucose to ethanol. In some instances, following pretreatment, the saccharification (break down of cellulose) and fermentation steps may be combined using a cellulolytic bacterium such as *C. thermocellum*. However, the latter procedure does not produce ethanol in high yield. One way around this problem is to use *Caldicellulosiruptor bescii*, a thermophilic anaerobic Gram-positive cellulolytic bacterium that grows optimally at 75 to 80°C. This bacterium is able to break down a wide range of plant lignocellulosic biomass. *C. bescii* secretes a number of free cellulases that each contain multiple catalytic domains that act synergistically to degrade lignocellulosic substrates in a manner that is similar to a cellulosome. During the hydrolysis of biomass, fragments of one of these cellulases, CelA, are released from the protein, thereby enabling smaller active enzyme fragments to reach parts of the lignocellulosic substrate that the very large CelA protein is not able to access. In addition, when β-glucosidase is added to CelA, the enzyme is able to relatively rapidly break down crystalline cellulose, a substrate that is generally broken down only slowly. Unfortunately, *C. bescii* synthesizes a large amount of lactate and hydrogen as a by-product of its metabolism and, also, does not produce alcohol (Fig. 8.64). To prevent lactate from being synthesized, the lactate dehydrogenase gene was deleted. This resulted in an increase in the synthesis of acetate and hydrogen. Next, a gene from a strain of *C. thermocellum* encoding the bifunctional

Figure 8.64 The conversion of glucose to ethanol in an engineered strain of *C. bescii*. The multiple arrows shown between glucose and pyruvate indicate the number of enzymatic steps. Metabolites shown in yellow were all synthesized in the wild-type strain of *C. bescii*, as was lactate. However, in the engineered strain, the gene encoding the enzyme lactate dehydrogenase that converts pyruvate to lactate was deleted, and lactate was no longer synthesized. In addition, the engineered strain contains a gene encoding acetaldehyde/alcohol dehydrogenase that converts acetyl-CoA to acetaldehyde and then to ethanol. The details of electron transfer and hydrogen synthesis are not shown.



enzyme acetaldehyde/alcohol dehydrogenase was used to transform the lactate dehydrogenase-minus mutant of *C. bescii*. The bacterial strain that was modified as described was now able to directly convert switchgrass biomass to ethanol without the need for any pretreatment or the addition of exogenous enzymes. Moreover, the engineered *C. bescii* strain produced a significantly lowered level of acetate and hydrogen, instead synthesizing ethanol. In the future it will be necessary to enhance the level of ethanol that is synthesized by this engineered bacterium; however, this represents an important step in the development of an industrial procedure for the production of alcohol from switchgrass.

Alcohol Production by *Zymomonas mobilis*

Although industrial fermentations that produce alcohol are performed almost exclusively with *S. cerevisiae*, the bacterium *Z. mobilis* is a potentially useful organism for this purpose. *Zymomonas* is a Gram-negative, rod-shaped facultative anaerobe that can ferment glucose, fructose, and sucrose and produce a relatively high yield of alcohol (Table 8.10). This high yield of ethanol is probably related to the fact that *Zymomonas* does not proliferate extensively (i.e., produce biomass) during fermentation. Since *Zymomonas* uses less substrate for biomass formation, more is available for ethanol production. In this regard, yeast produces 2 mol of ATP per mol of glucose, whereas *Zymomonas* uses a different pathway and produces only 1 mol of ATP per mol of glucose. Historically, *Zymomonas* has been used in tropical regions as a fermentative agent for the production of alcoholic beverages (such as pulque, which is made from the sap of the agave plant).

Although *Zymomonas* produces alcohol at a much higher rate than does *S. cerevisiae* (Table 8.10), it can use only a limited number of carbon substrates (i.e., glucose, fructose, and sucrose) for the production of alcohol. In addition, broad-host-range cloning vectors and foreign genes are difficult to maintain in this organism, and *Zymomonas* is naturally resistant to many of the more commonly used antibiotics, which precludes using many standard antibiotic resistance marker systems for cloning experiments. Nevertheless, a number of foreign genes have been successfully introduced into and expressed in *Zymomonas*, expanding the range of substrates that *Zymomonas* can utilize. For example, genes encoding enzymes that hydrolyze lactose, starch, cellulose, xylose, arabinose, and cellobiose have all been expressed

Table 8.10 Comparison of *Z. mobilis* and *S. cerevisiae* as alcohol producers[a]

Attribute	Value for:	
	Z. mobilis	Yeast
Conversion of sugar to ethanol (%)	96	96
Maximum ethanol concentration (%)	12	12
Ethanol productivity rate (g g^{-1} h^{-1})	5.67	0.67
Volumetric ethanol productivity rate (g $liter^{-1}$ h^{-1})	200	29
Sugar tolerance (%)	>40	>40
pH range for ethanol production	3.5–7.5	2–6.5
Optimum temperature (°C)	25–30	30–38

[a]The ethanol productivity rate was measured under batch fermentation conditions. The volumetric ethanol productivity rate was measured during continuous culture. Both strains yielded the same ethanol concentration (12%) and had the same sugar tolerance (>40%).

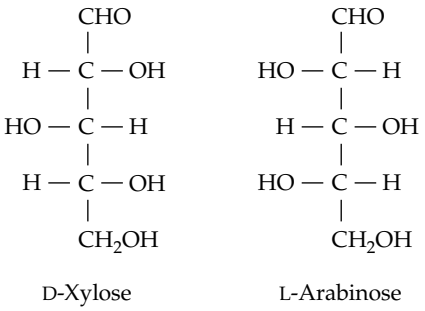


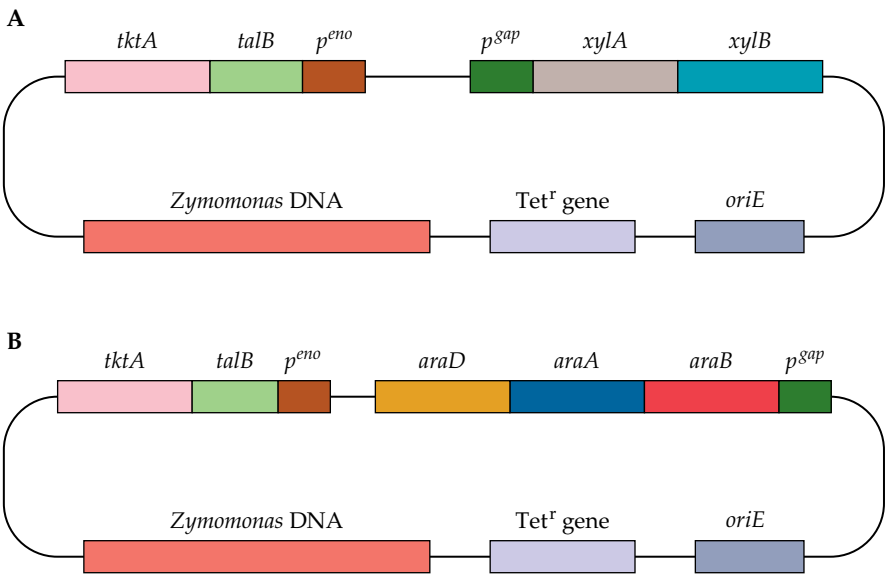
Figure 8.65 Chemical structures of D-xylose and L-arabinose.

in *Zymomonas*. However, in many of these cases, the transformed bacterium was unable to utilize the novel substrate as the sole carbon source.

The five-carbon sugar xylose is the predominant pentose sugar in hardwoods, while its isomer arabinose is present in large amounts in various agricultural and other herbaceous plants (Fig. 8.65). Some arabinose-containing plants, such as switchgrass, have been considered for use as dedicated energy crops, i.e., plants grown solely for use as sources of energy. Thus, it would be useful if *Z. mobilis* were able to convert xylose and arabinose to ethanol.

To develop strains of *Z. mobilis* capable of growth and ethanol production with xylose or arabinose as a substrate, the bacterium was transformed with genes encoding either xylose or arabinose assimilation. In both cases, the transformed *Z. mobilis* cells were unable to further metabolize pentoses (xylulose-5-phosphate, ribulose-5-phosphate, and ribose-5-phosphate) that are formed after xylose or arabinose was assimilated. To remedy this situation, *Z. mobilis* cells were simultaneously transformed with two pentose metabolism genes on an *E. coli–Z. mobilis* shuttle vector (Fig. 8.66). The pentose metabolism genes encode the enzymes transketolase and transaldolase, both under the control of the *Z. mobilis* enolase promoter. As expected, the transformants could now assimilate and metabolize either xylose or arabinose, with these sugars being efficiently converted to ethanol

Figure 8.66 (A) *Zymomonas–E. coli* shuttle vector carrying one operon with genes encoding enzymes used for xylose assimilation (*xylA* and *xylB*) and another with genes encoding enzymes involved in pentose metabolism (*tktA* and *talB*). p^{eno}, enolase promoter; p^{gap}, glyceraldehyde-3-phosphate dehydrogenase promoter; *xylA*, xylose isomerase gene; *xylB*, xylulokinase gene; *tktA*, transketolase gene; *talB*, transaldolase gene; Tet^r, tetracycline resistance gene; *oriE*, *E. coli* origin of replication. **(B)** *Zymomonas–E. coli* shuttle vector carrying one operon with genes encoding enzymes used for arabinose assimilation (*araB*, *araA*, and *araD*) and another with genes encoding enzymes involved in pentose metabolism (*tktA* and *talB*). *araB*, L-ribulokinase gene; *araA*, L-arabinose isomerase gene; *araD*, L-ribulose-5-phosphate-4-epimerase gene.



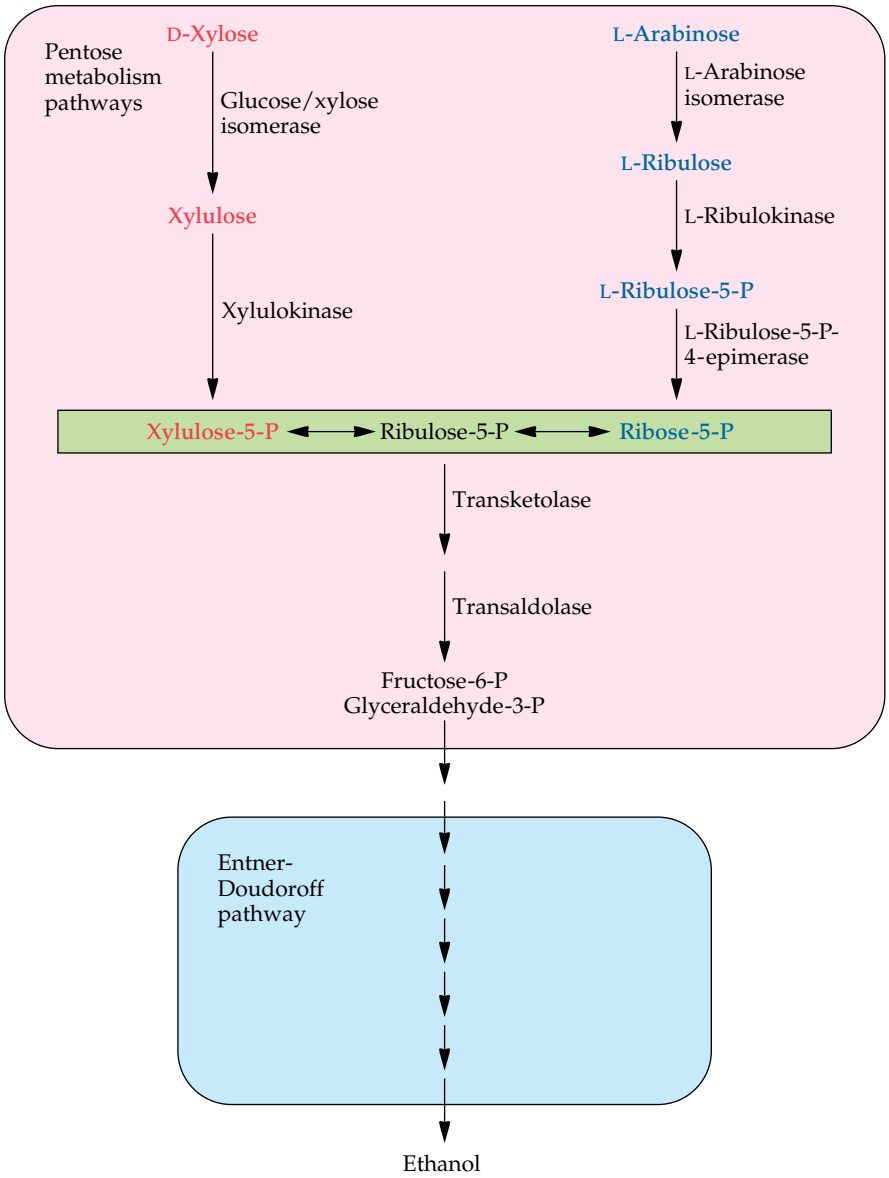


Figure 8.67 Schematic representation of the engineered assimilation of either xylose or arabinose by *Z. mobilis* and the engineered conversion of the resultant pentoses (shown in the green box) to ethanol.

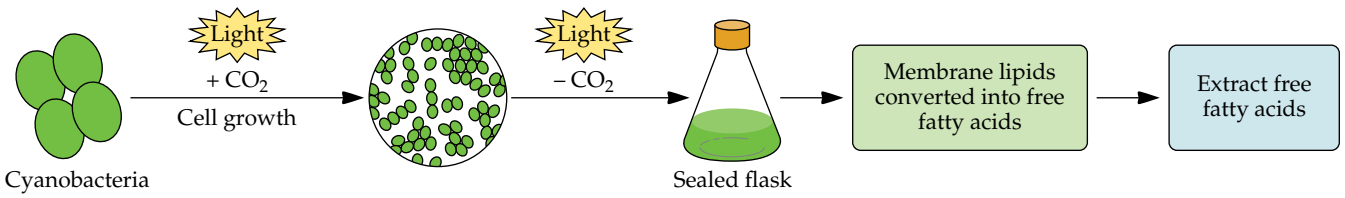
(Fig. 8.67). This work demonstrates the feasibility of metabolically engineering *Z. mobilis* as an ethanol producer by using xylose or arabinose as a carbon source. Moreover, the xylose- and arabinose-fermenting strains of *Z. mobilis* might be used together in a mixed bacterial culture for the conversion of the major sugars from certain agricultural residues into ethanol. To efficiently and economically convert lignocellulosic wastes to ethanol, it is necessary to convert all of the sugars, the pentoses as well as the hexoses, to ethanol. *Z. mobilis* lacks the pathways needed for the metabolism of mannose and galactose, which constitute a significant fraction of the hexoses in lignocellulosic material. Whether *Z. mobilis* can be engineered to be the organism of choice as part of a process of this sort is still an open question.

Instead of genetically engineering *Z. mobilis* to utilize xylose and arabinose and convert them into ethanol, some scientists have engineered *E. coli* to express some *Z. mobilis* genes so that it can produce ethanol. In this case, xylose and arabinose are converted to pyruvate by endogenous *E. coli* enzymes. Different *E. coli* genes were mutated to prevent the pyruvate from being converted into other, unwanted metabolites. Instead, the pyruvate is converted to ethanol by the enzymes pyruvate decarboxylase and alcohol dehydrogenase, with the genes for both of these enzymes coming from *Z. mobilis*. This sort of genetic modification has also been used to change the bacterium *Klebsiella oxytoca* into an ethanologenic organism. With both *E. coli* and *K. oxytoca*, the introduction of *Z. mobilis* genes yielded recombinant bacteria that were quite efficient in the laboratory at converting various sugars, both pentoses and hexoses, into alcohol. It now remains to be demonstrated whether any of these bacteria are effective on a large scale with an industrial substrate.

Lipids from Cyanobacteria

Numerous scientists have proposed using cyanobacteria and other photosynthetic microorganisms as a source of lipids and other biofuel compounds that is more cost effective than production by terrestrial plants. The photosynthetic efficiency of cyanobacteria is typically much higher than that of terrestrial plants, they do not require (nonproductive) support tissues such as roots and stems, they are readily manipulated genetically, and they may be cultivated on inexpensive media in locations where it is not possible to grow traditional agricultural crops. In addition, cyanobacteria produce a number of potentially useful lipids, including numerous diacylglycerols as their membrane components. Unfortunately, to recover those compounds, it is necessary to lyse the cells and then solvent-extract the lipids, a process that is estimated to be around three quarters of the cost of biofuel production from cyanobacteria. To get around these expensive and difficult steps, one group of scientists developed a unique strategy that does not require a mechanical cell lysis step. Briefly, they genetically engineered cyanobacteria to synthesize three lipolytic enzymes (galactolipase and two phospholipases) whose synthesis was readily induced following cell growth (Fig. 8.68). The lipolytic enzymes were not synthesized in the presence of the CO_2 that was required for cyanobacterial growth; however, when the atmosphere in the growth chamber was changed to air

Figure 8.68 Flow chart depicting the production of free fatty acids from the cyanobacterium *Synechocystis* sp. strain PCC6803. Cyanobacterial cells were grown with shaking under continuous light and CO_2 until they reached a cell density of ~10^9 cells/ml. At that point, cells were transferred to sealed flasks, and while they continued to be shaken and illuminated, the CO_2 supply was stopped. The absence of CO_2 caused the induction of several recombinant lipolytic enzymes that broke down the cyanobacterial membranes into free fatty acids that were eventually isolated by extraction into hexane.



that did not contain any CO_2, the synthesis of the lipolytic enzymes was rapidly induced. The lipolytic enzymes then degraded the cell membranes, effectively lysing the cells and converting the membranes into free fatty acids that were readily extracted. While the efficacy of this "CO_2-limitation-inducible" system has been demonstrated on a small scale, it remains to be seen whether a large-scale version of this procedure will yield the hoped-for cost savings in the production of biofuel free fatty acids.

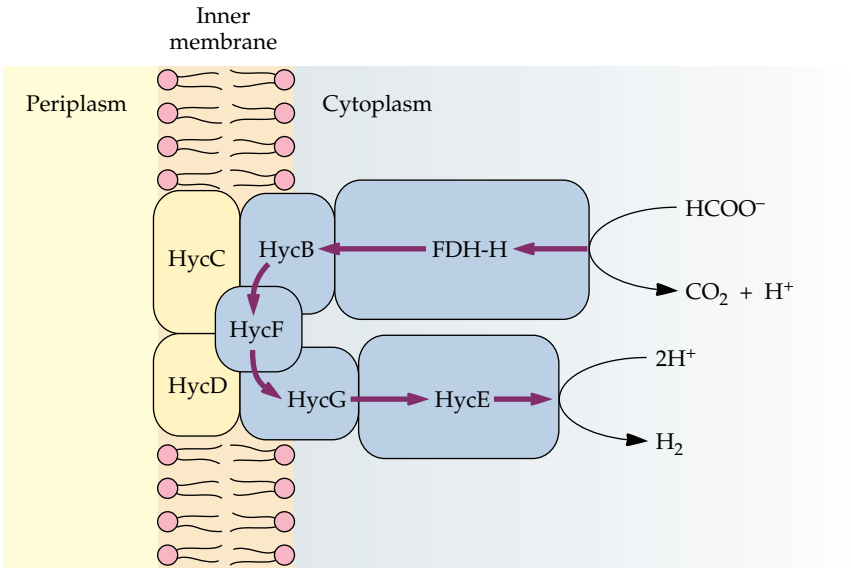
Hydrogen Production

Biologically produced hydrogen (biohydrogen) is a clean fuel that can be produced from various types of biomass and yields only water when it is burned. Hydrogen has been touted as the nonpolluting fuel of the future; however, it is both highly combustible and difficult to transport and to store. Nevertheless, it is expected that hydrogen will become a key component of our energy economy during the next 10 to 20 years.

It has been known for some time that formic acid (formate) can be produced inexpensively, often as a by-product of the synthesis of other chemicals, such as acetic acid. Moreover, some types of bacteria are able to convert formate into hydrogen and carbon dioxide via the formate hydrogen lyase system. If the bacterial system that is responsible for hydrogen synthesis could be optimized, it might be possible to develop a practical system for the synthesis of hydrogen from biomass.

The *E. coli* formate hydrogen lyase system consists of a large number of different proteins, including those shown in Fig. 8.69, as well as others that

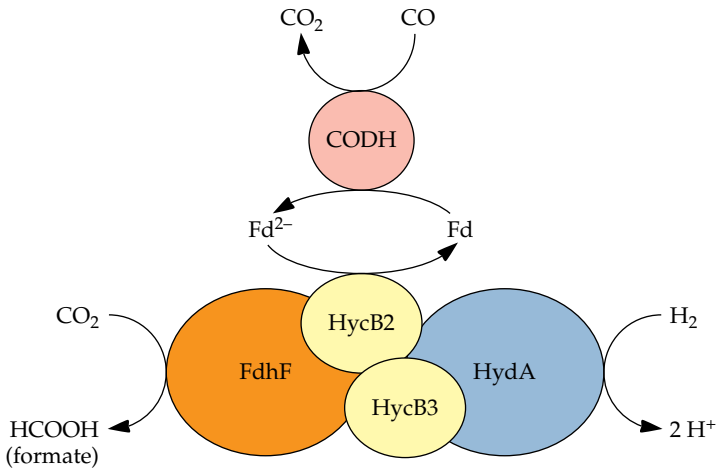
Figure 8.69 The major components of the *E. coli* formate hydrogen lyase. The purple arrows show the direction of the electron flux and the proteins through which the electrons flow (shown in blue). The Hyc proteins are the components of the formate hydrogen lyase system, which accept electrons from FDH-H (the large subunit of formate dehydrogenase). HycB is the small subunit of formate dehydrogenase and acts as the membrane docking target of FDH-H. HycC and HycD (shown in yellow) are integral membrane proteins and are thought to anchor the rest of the complex. HycE is the large subunit of hydrogenase 3, and HycG is the small subunit of hydrogenase 3. HycF is an electron transport protein.



specifically regulate the synthesis and maturation of these proteins. To overproduce the *E. coli* formate hydrogen lyase system and hence the amount of hydrogen, (i) the formate hydrogen lyase repressor gene, *hycA*, was inactivated and (ii) the formate hydrogen lyase activator gene, *fhlA*, was overexpressed. These manipulations resulted in the large subunit of formate dehydrogenase (FDH-H) and the large subunit of hydrogenase 3 (HycE) being overexpressed 6.5- and 7.0-fold, respectively, compared to the wild type. These changes resulted in a nearly 3-fold increase in hydrogen productivity in the mutant strain compared to the wild-type strain. Additional enhancement of the amount of hydrogen produced was obtained by culturing the engineered *E. coli* cells under anaerobic conditions in a bioreactor at a very high density (96 g [dry weight] per liter). When the formate concentration was maintained below 25 mM, continuous hydrogen synthesis of 23.6 g of hydrogen per hour per liter was realized. This level of hydrogen production is sufficient for this system to be considered potentially commercially viable.

One of the problems with renewable sources of electricity such as wind power is that the demand for electricity fluctuates and is not always equal to supply levels. Hence, it is necessary to find a way to store electric power at a large scale. Some scientists believe that electric power might be stored as hydrogen if there were a safe and economical means of storing and transporting hydrogen. Moreover, converting hydrogen into formate may provide a convenient and practical means of storing the energy from hydrogen. Recently, scientists discovered that the bacterium *Acetobacterium woodii* (a Gram-positive anaerobe) contains a simple enzyme system that can directly reduce CO_2 and H_2 to produce formate and that this enzyme system did not require either NADPH or any additional proteins (Fig. 8.70). Thus, it may be possible to convert the bacterium into a whole-cell catalyst for formate production. This may be done by removing sodium ions (which are essential for bacterial cell energy metabolism) from the cell growth medium so that the flux of reducing equivalents is directed away from cell respiration and toward formate production. It now remains to be

Figure 8.70 Schematic representation of the *Acetobacterium woodii* formate synthesis system. The catalytic subunits include a hydrogenase (HydA) and a formate dehydrogenase (FdhF). There are also two electron transfer subunits (HycB2 and HycB3). The system also uses ferredoxin (Fd) as an electron donor, where ferredoxin can be reduced (Fd^{2-}) by carbon monoxide dehydrogenase (CODH).



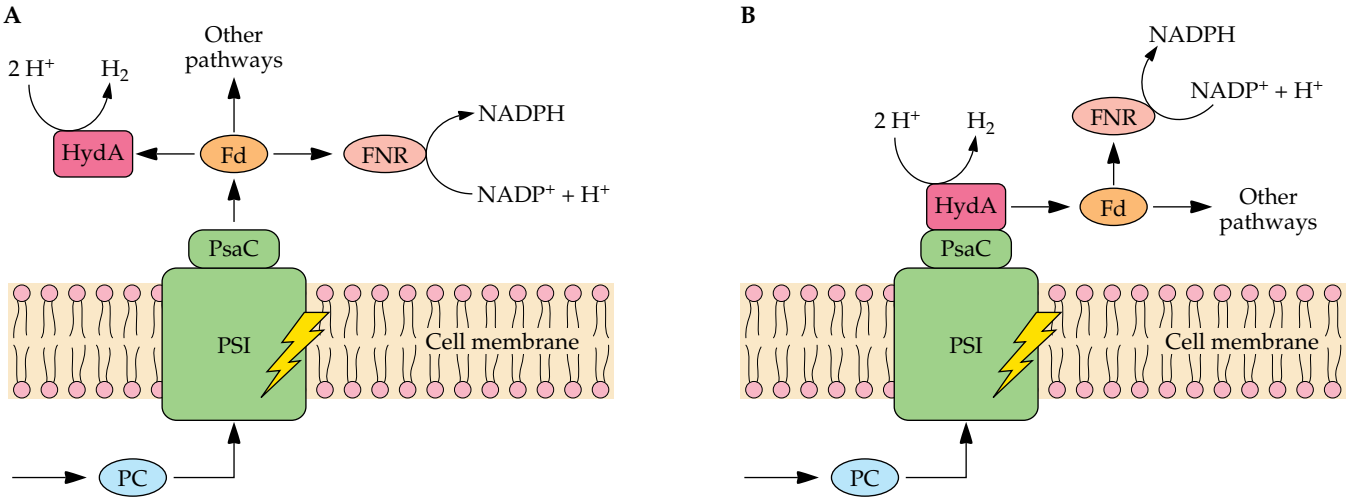


Figure 8.71 A portion of the photosynthetic electron transport chain in the green alga *Chlamydomonas reinhardtii*. **(A)** shows wild-type cells, while **(B)** shows the genetically modified cells. In **(A)**, very little hydrogen is produced, and a large amount of H^+ and $NADP^+$ is converted to NADPH. In **(B)**, a large amount of hydrogen is produced, while very little H^+ and $NADP^+$ are converted to NADPH. The yellow lightning bolt represents the light that drives photosynthesis. Abbreviations: PSI, photosystem I; Fd, ferredoxin; FNR, Fd:$NADP^+$ oxidoreductase; HydA, hydrogenase; PsaC, photosystem I subunit VII; PC, plastocyanin.

seen whether this interesting lab experiment can be scaled up and made into a useful energy storage technology.

One tantalizing means of biologically producing hydrogen involves engineering algae to convert water into hydrogen and oxygen with the process powered by sunlight. Although researchers have been working on this for many years, only recently have they reported efficiently utilizing the power of photosynthesis to produce hydrogen (Fig. 8.71). In green algae (aquatic members of the plant kingdom without roots, stems, or leaves), such as *Chlamydomonas reinhardtii*, in the thylakoid membranes of the chloroplast (organelles where photosynthesis takes place), the photosynthetic electron transport chain drives electron transport from water to ferredoxin. This system pumps protons across the membrane, providing energy in the form of ATP and NADPH to power CO_2 fixation. One of the main problems in using green algae to produce hydrogen is the competition for reduced ferredoxin with other biological processes. Most relevant in this context, algal hydrogenases compete poorly with Fd:$NADP^+$ oxidoreductase (FNR) for reduced ferredoxin (Fd). This is because FNR is typically present at levels of about 70 times that of hydrogenase and also binds Fd somewhat more tightly. To get around this problem, researchers genetically modified the gene for the *C. reinhardtii* hydrogenase so that it was expressed in chloroplasts instead of the cytoplasm and directly fused to a photosystem I protein. This genetic manipulation resulted in a much greater flux of electrons through the hydrogenase instead of through FNR, resulting in a significant increase in hydrogen production and a concomitant decrease in CO_2 fixation. Compared to the unmodified strain, the genetically modified alga constitutively produced the hydrogenase fusion protein, directed the majority of electrons from water oxidation to the

synthesis of hydrogen, produced lower levels of oxygen over a period of several days, and preserved the normal proton translocation and ATP synthesis over several days, thereby maintaining cell viability. Although this system is still in a rudimentary form, it should be possible to further improve it to the point where it can be used on a large and practical scale because all of the relevant genes are encoded within the bacteria-like algal chloroplast genome.

summary

In addition to using bacteria as factories for the production of proteins, such as restriction enzymes, it is possible to modify the metabolic pathways of organisms, either by introducing new genes or by altering existing ones. In this way, various organisms can be genetically engineered for the production of a range of low-molecular-weight compounds, such as L-ascorbic acid, indigo, amino acids, antibiotics, lycopene, succinic acid, and the monomeric subunits of various biopolymers, such as xanthan gum and polyhydroxyalkanoates. Here, the strategy is to insert the genes for one or more specific enzymes into the host organism. When they are expressed, the inserted genes encode a new pathway or augment a preexisting pathway for the synthesis of a specific compound. In addition, the biosynthesis of a desired compound may often be significantly increased by modulating the metabolic flux of the organism by turning on some pathways and blocking others. Several studies have shown that the creation of such unusual enzymatic pathways is technically feasible. Moreover, recombinant DNA technology has led to the development of new and more efficient synthetic routes for a variety of important compounds.

Some microorganisms may be used to clear the environment of contaminating substances. Many members of the bacterial genus *Pseudomonas* naturally carry plasmids that encode enzymes capable of degrading aromatic and halogenated organic compounds. In most cases, a single plasmid carries the genes encoding enzymes for a specific degradative pathway. By combining plasmids from different pseudomonad strains within a single host, it is possible to create an organism with multiple degradation capabilities. In addition, by genetic manipulation, the range of substrates degraded by a particular enzymatic pathway can be extended.

Raw biological material is called biomass and is often used as a starting material in industrial processes. The use of plant material as the source of carbohydrates for the production of alcohol or fructose requires a number of enzymatic steps. The enzymes that these processes catalyze are often used only once and then discarded. To enhance enzymatic conversions and decrease costs, bacterial genes encoding enzymes that are thermostable, highly efficient catalytically, or confer tolerance to alcohol have been cloned, characterized, and tested. To improve the commercial production of alcohol, researchers have genetically transformed the bacterium *Z. mobilis* with genes that allow it to utilize a broad range of compounds as carbon sources.

Often, as a consequence of processing plant material, large amounts of lignocellulose remain. This material generally has been treated as a waste product. However, there is now interest in using lignocellulose as a resource for carbon-containing compounds, especially glucose, that can be used in other processes such as biofuel production. Retrieving glucose from lignocellulose is not an easy matter. Lignocellulose is a complex of lignin, hemicellulose, and cellulose; without harsh and expensive pretreatment, it is generally refractory to enzymatic degradation. Recent research has focused on characterizing the mechanism of the breakdown of cellulose to glucose. The genes for endoglucanases, exoglucanases, and β-glucosidases from a variety of organisms have been cloned and characterized, but to date there has been little success in formulating a set of enzymes that efficiently degrades cellulose *in vitro* on a large scale.

REFERENCES

Alper H, Moxley J, Nevoigt E, Fink GR, Stephanopoulos G. 2006. Engineering yeast transcription machinery for improved ethanol tolerance and production. *Science* **314:**1565–1568.

Amaro TMMM, Rosa D, Comi G, Iacumin L. 2019. Prospects for the use of whey for polyhydroxyalkanoate (PHA) production. *Front Microbiol* **10:**992.

Anderson S, Marks CB, Lazarus R, Miller J, Stafford K, Seymour J, Light D, Rastetter W, Estell D. 1985. Production of 2-keto-l-gulonate, an intermediate in l-ascorbate synthesis, by a genetically modified *Erwinia herbicola*. *Science* **230:**144–149.

Austin HP, Allen MD, Donohoe BS, Rorrer NA, Kearns FL, Silveira RL, Pollard BC, Dominick G, Duman R, El Omari K, Mykhaylyk V, Wagner A, Michener WE, Amore A, Skaf MS, Crowley MF, Thorne AW, Johnson CW, Woodcock HL, McGeehan JE, Beckham GT. 2018. Characterization and engineering of a plastic-degrading aromatic polyesterase. *Proc Natl Acad Sci USA* **115:**E4350–E4357.

Bajaj P, Mahajan R. 2019. Cellulase and xylanase synergism in industrial biotechnology. *Appl Microbiol Biotechnol* **103:**8711–8724.

Banta S, Swanson BA, Wu S, Jarnagin A, Anderson S. 2002. Alteration of the specificity of the cofactor-binding pocket of *Corynebacterium* 2,5-diketo-d-gluconic acid reductase A. *Protein Eng* **15:**131–140.

Banta S, Swanson BA, Wu S, Jarnagin A, Anderson S. 2002. Optimizing an artificial metabolic pathway: engineering the cofactor specificity of *Corynebacterium* 2,5-diketo-d-gluconic acid reductase for use in vitamin C biosynthesis. *Biochemistry* **41:**6226–6236.

Barnett CC, Berka RM, Fowler T. 1991. Cloning and amplification of the gene encoding an extracellular β-glucosidase from *Trichoderma reesei*: evidence for improved rates of saccharification of cellulosic substrates. *Biotechnology (N Y)* **9:**562–567.

Baumgart M, Unthan S, Rückert C, Sivalingam J, Grünberger A, Kalinowski J, Bott M, Noack S, Frunzke J. 2013. Construction of a prophage-free variant of *Corynebacterium glutamicum* ATCC 13032 for use as a platform strain for basic research and industrial biotechnology. *Appl Environ Microbiol* **79:**6006–6015.

Bayer EA, Shimon LJW, Shoham Y, Lamed R. 1998. Cellulosomes-structure and ultrastructure. *J Struct Biol* **124:** 221–234.

Béguin P. 1990. Molecular biology of cellulose degradation. *Annu Rev Microbiol* **44:**219–248.

Berini F, Marinelli F, Binda E. 2020. Streptomycetes: attractive hosts for recombinant protein production. *Front Microbiol* **11:**1958.

Brunecky R, Alahuhta M, Xu Q, Donohoe BS, Crowley MF, Kataeva IA, Yang SJ, Resch MG, Adams MWW, Lunin VV, Himmel ME, Bomble YJ. 2013. Revealing nature's cellulase diversity: the digestion mechanism of *Caldicellulosiruptor bescii* CelA. *Science* **342:**1513–1516.

Buchholz SE, Dooley MM, Eveleigh DE. 1987. *Zymomonas*—an alcoholic enigma. *Trends Biotechnol* **5:**199–204.

Buchholz SE, Eveleigh DE. 1990. Genetic modification of *Zymomonas mobilis*. *Biotechnol Adv* **8:**547–581.

Budde CF, Riedel SL, Willis LB, Rha C, Sinskey AJ. 2011. Production of poly(3-hydroxybutyrate-co-3-hydroxyhexanoate) from plant oil by engineered *Ralstonia eutropha* strains. *Appl Environ Microbiol* **77:**2847–2854.

Carpine R, Olivieri G, Hellingwerf KJ, Pollio A, Mazocchella A. 2020. Industrial production of poly-β-hydroxybutyrate from CO_2: can cyanobacteria meet this challenge? *Processes (Basel)* **8:**323 http://dx.doi.org/10.3390/pr8030323.

Chakrabarty AM. 1981. Microorganisms having multiple compatible degradative energy-generating plasmids and preparation thereof. U.S. patent 4,259,444.

Chen G-Q, Hajnal I. 2015. The 'PHAome'. *Trends Biotechnol* **33:**559–564.

Chen G-Q, Hajnal I, Wu H, Lv L, Ye J. 2015. Engineering biosynthesis mechanisms for diversifying polyhydroxyalkanoates. *Trends Biotechnol* **33:**565–574.

Chen H, Du L. 2016. Iterative polyketide biosynthesis by modular polyketide synthases in bacteria. *Appl Microbiol Biotechnol* **100:**541–557.

Choi KR, Jang WD, Yang D, Cho JS, Park D, Lee SY. 2019. Systems metabolic engineering strategies: integrating systems and synthetic biology with metabolic engineering. *Trends Biotechnol* **37:**817–837.

Chung D, Cha M, Guss AM, Westpheling J. 2014. Direct conversion of plant biomass to ethanol by engineered *Caldicellulosiruptor bescii*. *Proc Natl Acad Sci USA* **111:**8931–8936.

Cole GE, McCabe PC, Inlow D, Gelfand DH, Ben-Bassat A, Innis MA. 1988. Stable expression of *Aspergillus awamori* glucoamylase in distiller's yeast. *Bio/Technology* **6:**417–421.

Daletos G, Katsimpouras C, Stephanopoulos G. 2020. Novel strategies and platforms for industrial isoprenoid engineering. *Trends Biotechnol* **38:**811–822.

Davidi L, Moraïs S, Artzi L, Knop D, Hadar Y, Arfi Y, Bayer EA. 2016. Toward combined delignification and saccharification of wheat straw by a laccase-containing designer cellulosome. *Proc Natl Acad Sci USA* **113:**10854–10859.

Deanda K, Zhang M, Eddy C, Picataggio S. 1996. Development of an arabinose-fermenting *Zymomonas mobilis* strain by metabolic pathway engineering. *Appl Environ Microbiol* **62:**4465–4470.

Dekker K, Sugiura A, Yamagata H, Sakaguchi K, Udaka S. 1992. Efficient production of thermostable *Thermus thermophilus* xylose isomerase in *Escherichia coli* and *Bacillus brevis*. *Appl Microbiol Biotechnol* **36:**727–732.

Eberhardt D, Jensen JVK, Wendisch VF. 2014. l-citrulline production by metabolically engineered *Corynebacterium glutamicum* from glucose and alternative carbon sources. *AMB Express* **4:**85.

Ensley BD, Ratzkin BJ, Osslund TD, Simon MJ, Wackett LP, Gibson DT. 1983. Expression of naphthalene oxidation genes in *Escherichia coli* results in the biosynthesis of indigo. *Science* **222:**167–169.

Fabara AN, Fraaije MW. 2020. An overview of microbial indigo-forming enzymes. *Appl Microbiol Biotechnol* **104:** 925–933.

Fierobe H-P, Mingardon F, Mechaly A, Bélaïch A, Rincon MT, Pagès S, Lamed R, Tardif C, Bélaïch J-P, Bayer EA. 2005. Action of designer cellulosomes on homogeneous versus complex substrates: controlled incorporation of three distinct enzymes into a defined trifunctional scaffoldin. *J Biol Chem* **280:**16325–16334.

Fu J-F, Tseng Y-H. 1990. Construction of lactose-utilizing *Xanthomonas campestris* and production of xanthan gum from whey. *Appl Environ Microbiol* **56:**919–923.

Galera-Prat A, Vera AM, Moraïs S, Vazana Y, Bayer EA, Carrión-Vázquez M. 2020. Impact of scaffoldin mechanostability on cellulosomal activity. *Biomater Sci* **8:**3601–3610.

Gao C, Hindra D, Mulder D, Yin C, Elliot MA. 2012. Crp is a global regulator of antibiotic production in streptomyces. *MBio* **3:**e00407–e00412.

Hahn-Hägerdal B, Galbe M, Gorwa-Grauslund MF, Lidén G, Zacchi G. 2006. Bio-ethanol: the fuel of tomorrow from the residues of today. *Trends Biotechnol* **24:**549–556.

Hahn-Hägerdal B, Karhumaa K, Fonseca C, Spencer-Martins I, Gorwa-Grauslund MF. 2007. Towards industrial pentose-fermenting yeast strains. *Appl Microbiol Biotechnol* **74:**937–953.

Hazer B, Steinbüchel A. 2007. Increased diversification of polyhydroxyalkanoates by modification reactions for industrial and medical applications. *Appl Microbiol Biotechnol* **74:**1–12.

Hopwood DA, Bibb MJ, Bruton CJ, Chater KF, Feitelson JS, Gil JA. 1983. Cloning *Streptomyces* genes for antibiotic production. *Trends Biotechnol* **1:**42–48 http://dx.doi.org/10.1016/0167-7799(83)90068-9.

Howard KA, Card C, Benner JS, Callahan HL, Maunus R, Silber K, Wilson G, Brooks JE. 1986. Cloning the *Dde*I restriction-modification system using a two-step method. *Nucleic Acids Res* **14:**7939–7951.

Howe JA, Wang H, Fischmann TO, Balibar CJ, Xiao L, Galgoci AM, Malinverni JC, Mayhood T, Villafania A, Nahvi A, Murgolo N, Barbieri CM, Mann PA, Carr D, Xia E, Zuck P, Riley D, Painter RE, Walker SS, Sherborne B, de **Jesus R, Pan W, Plotkin MA, Wu J, Rindgen D, Cummings J, Garlisi CG, Zhang R, Sheth PR, Gill CJ, Tang H, Roemer T.** 2015. Selective small-molecule inhibition of an RNA structural element. *Nature* **526:**672–677.

Hutchinson CR, Decker H, Madduri K, Otten SL, Tang L. 1993-1994. Genetic control of polyketide biosynthesis in the genus *Streptomyces*. *Antonie van Leeuwenhoek* **64:**165–176.

Kahn A, Moraïs S, Galanopoulou AP, Chung D, Sarai NS, Hengge N, Hatzinikolaou DG, Himmel ME, Bomble YJ, Bayer EA. 2019. Creation of a functional hyperthermostable designer cellulosome. *Biotechnol Biofuels* **12:**44.

Kallio P, Palva A, Palva I. 1987. Enhancement of α-amylase production by integrating and amplifying the α-amylase gene of *Bacillus amyloliquefaciens* in the genome of *Bacillus subtilis*. *Appl Microbiol Biotechnol* **27:**64–71.

Kanygin A. Milrad Y, Thummala C, Reifschneider K, Baker P, Marco P, Yacoby I, Redding KE. 2020. Rewiring photosynthesis: a photosystem I-hydrogenase chimera that makes H_2 *in vivo*. *Energy Environ Sci* **13:**2903–2914.

Nordberg Karlsson E, Schmitz E, Linares-Pastén JA, Adlercreutz P. 2018. Endo-xylanases as tools for production of substituted xylooligosaccharides with prebiotic properties. *Appl Microbiol Biotechnol* **102:**9081–9088.

Krämer R. 1996. Genetic and physiological approaches for the production of amino acids. *J Biotechnol* **45:**1–21 http://dx.doi.org/10.1016/0168-1656(95)00146-8.

Kennedy JF, Cabalda VM, White CA. 1988. Enzymic starch utilization and genetic engineering. *Trends Biotechnol* **6:**184–189.

Knott BC, Erickson E, Allen MD, Gado JE, Graham R, Kearns FL, Pardo I, Topuzlu E, Anderson JJ, Austin HP, Dominick G, Johnson CW, Rorrer NA, Szostkiewicz CJ, Copié V, Payne CM, Woodcock HL, Donohoe BS, Beckham GT, McGeehan JE. 2020. Characterization and engineering of a two-enzyme system for plastics depolymerization. *Proc Natl Acad Sci USA* **117:**25476–25485.

Knowles J, Lehtovaara P, Teeri T. 1987. Cellulase families and their genes. *Trends Biotechnol* **5:**255–261.

Kusumawardhani H, Hosseini R, de Winde JH. 2018. Solvent tolerance in bacteria: fulfilling the promise of the biotech era? *Trends Biotechnol* **36:**1025–1039.

Latorre-García L, Adam AC, Manzanares P, Polaina J. 2005. Improving the amylolytic activity of *Saccharomyces cerevisiae* glucoamylase by the addition of a starch binding domain. *J Biotechnol* **118:**167–176.

Ladner CC, Williams GJ. 2016. Harnessing natural product assembly lines: structure, promiscuity, and engineering. *J Ind Microbiol Biotechnol* **43:**371–387.

Lam FH, Ghaderi A, Fink GR, Stephanopoulos G. 2014. Biofuels. Engineering alcohol tolerance in yeast. *Science* **346:**71–75.

Lee SY, Chang HN, Chang YK. 1994. Production of poly(β-hydroxybutyric acid) by recombinant *Escherichia coli*. *Ann N Y Acad Sci* **721**(1 Recombinant D)**:**43–53.

Lee SY, Chang HN. 1995. Production of poly(3-hydroxybutyric acid) by recombinant *Escherichia coli* strains: genetic and fermentation studies. *Can J Microbiol* **41**(Suppl 1)**:**207–215.

Lee SY, Kim HU. 2015. Systems strategies for developing industrial microbial strains. *Nat Biotechnol* **33:**1061–1072.

Leungsakul T, Johnson GR, Wood TK. 2006. Protein engineering of the 4-methyl-5-nitrocatechol monooxygenase from *Burkholderia* sp. strain DNT for enhanced degradation of nitroaromatics. *Appl Environ Microbiol* **72:**3933–3939.

Li L, Liu Z, Jiang H, Mao X. 2020. Biotechnological production of lycopene by microorganisms. *Appl Microbiol Biotechnol* **104:**10307–10324.

Liu C-I, Liu GY, Song Y, Yin F, Hensler ME, Jeng W-Y, Nizet V, Wang AH-J, Oldfield E. 2008. A cholesterol biosynthesis inhibitor blocks *Staphylococcus aureus* virulence. *Science* **319:**1391–1394.

Liu X, Fallon S, Sheng J, Curtiss R III. 2011. CO_2-limitation-inducible green recovery of fatty acids from cyanobacterial biomass. *Proc Natl Acad Sci USA* **108:**6905–6908.

Matthäus F, Ketelhot M, Gatter M, Barth G. 2014. Production of lycopene in the non-carotenoid-producing yeast *Yarrowia lipolytica*. *Appl Environ Microbiol* **80:**1660–1669.

Meng M, Lee C, Bagdasarian M, Zeikus JG. 1991. Switching substrate preference of thermophilic xylose isomerase from d-xylose to d-glucose by redesigning the substrate binding pocket. *Proc Natl Acad Sci USA* **88:**4015–4019.

Mermod N, Harayama S, Timmis KN. 1986. New route to bacterial production of indigo. *Bio/Technology* **4:**321–324.

Mingardon F, Chanal A, López-Contreras AM, Dray C, Bayer EA, Fierobe H-P. 2007. Incorporation of fungal cellulases in bacterial minicellulosomes yields viable, synergistically acting cellulolytic complexes. *Appl Environ Microbiol* **73:**3822–3832.

Mohanan N, Montazer Z, Sharma PK, Levin DB. 2020. Microbial and enzymatic degradation of synthetic plastics. *Front Microbiol* **11:**580709.

Motika SE, Ulrich RJ, Geddes EJ, Lee HY, Lau GW, Hergenrother PJ. 2020. Gram-negative antibiotic active through inhibition of an essential riboswitch. *J Am Chem Soc* **142:**10856–10862.

Morya R, Salvachúa D, Thakur IS. 2020. *Burkholderia*: an untapped but promising bacterial genus for the conversion of aromatic compounds. *Trends Biotechnol* **38:**963–975.

Murakami T, Burian J, Yanai K, Bibb MJ, Thompson CJ. 2011. A system for the targeted amplification of bacterial gene clusters multiplies antibiotic yield in *Streptomyces coelicolor*. *Proc Natl Acad Sci USA* **108:**16020–16025.

Nakamori S, Kobayashi S-I, Kobayashi C, Takagi H. 1998. Overproduction of l-cysteine and l-cystine by *Escherichia coli* strains with a genetically altered serine acetyltransferase. *Appl Environ Microbiol* **64:**1607–1611.

Nikel PI, de Almeida A, Melillo EC, Galvagno MA, Pettinari MJ. 2006. New recombinant *Escherichia coli* strain tailored for the production of poly(3-hydroxybutyrate) from agroindustrial by-products. *Appl Environ Microbiol* **72:**3949–3954.

Panesar PS, Marwaha SS, Kennedy JF. 2006. *Zymomonas mobilis*: an alternative ethanol producer. *J Chem Technol Biotechnol* **81:**623–635 http://dx.doi.org/10.1002/jctb.1448.

Park JH, Lee KH, Kim TY, Lee SY. 2007. Metabolic engineering of *Escherichia coli* for the production of l-valine based on transcriptome analysis and *in silico* gene knockout simulation. *Proc Natl Acad Sci USA* **104:**7797–7802.

Pfeifer BA, Admiraal SJ, Gramajo H, Cane DE, Khosla C. 2001. Biosynthesis of complex polyketides in a metabolically engineered strain of *E. coli*. *Science* **291:**1790–1792.

Piekarowicz A, Yuan R, Stein DC. 1991. A new method for the rapid identification of genes encoding restriction and modification enzymes. *Nucleic Acids Res* **19:**1831–1835.

Prajapati AS, Panchal KJ, Pawar VA, Noronha MJ, Patel DH, Subramanian RB. 2018. Review on cellulase and xylanase engineering for biofuel production. *Ind Biotechnol (New Rochelle NY)* **14:**38–44.

Quax WJ, Mrabet NT, Luiten RGM, Schuurhuizen PW, Stanssens P, Lasters I. 1991. Enhancing the thermostability of glucose isomerase by protein engineering. *Biotechnology (N Y)* **9:**738–742.

Ramos JL, Wasserfallen A, Rose K, Timmis KN. 1987. Redesigning metabolic routes: manipulation of TOL plasmid pathway for catabolism of alkylbenzoates. *Science* **235:**593–596.

Ren Q, Sierro N, Kellerhals M, Kessler B, Witholt B. 2000. Properties of engineered poly-3-hydroxyalkanoates produced in recombinant *Escherichia coli* strains. *Appl Environ Microbiol* **66:**1311–1320.

Rhie HG, Dennis D. 1995. Role of *fadR* and *ato*C(Con) mutations in poly(3-hydroxybutyrate-co-3-hydroxyvalerate) synthesis in recombinant *pha*$^+$ *Escherichia coli*. *Appl Environ Microbiol* **61:**2487–2492.

Richins RD, Kaneva I, Mulchandani A, Chen W. 1997. Biodegradation of organophosphorus pesticides by surface-expressed organophosphorus hydrolase. *Nat Biotechnol* **15:**984–987.

Schuchmann K, Müller V. 2013. Direct and reversible hydrogenation of CO_2 to formate by a bacterial carbon dioxide reductase. *Science* **342:**1382–1385.

Suenaga H, Nonaka K, Fujihara H, Goto M, Furukawa K. 2010. Hybrid pseudomonads engineered by two-step homologous recombination acquire novel degradation abilities toward aromatics and polychlorinated biphenyls. *Appl Microbiol Biotechnol* **88:**915–923.

Sun J, Alper HS. 2015. Metabolic engineering of strains: from industrial-scale to lab-scale chemical production. *J Ind Microbiol Biotechnol* **42:**423–436.

Takagi H, Awano N, Kobayashi S, Noji M, Saito K, Nakamori S. 1999. Overproduction of l-cysteine and l-cystine by expression of genes for feedback inhibition-insensitive serine acetyltransferase from *Arabidopsis thaliana* in *Escherichia coli*. *FEMS Microbiol Lett* **179:**453–459.

Tao H, Gonzalez R, Martinez A, Rodriguez M, Ingram LO, Preston JF, Shanmugam KT. 2001. Engineering a homoethanol pathway in *Escherichia coli*: increased glycolytic flux and levels of expression of glycolytic genes during xylose fermentation. *J Bacteriol* **183:**2979–2988.

Timmis KN, Steffan RJ, Unterman R. 1994. Designing microorganisms for the treatment of toxic wastes. *Annu Rev Microbiol* **48:**525–557.

Unde V, Mutnuri S. 2014. Bio-catalytic production of indigo using biphasic organic solvent. *Discov Biotechnol* **5:**3–10.

van Rooyen R, Hahn-Hägerdal B, La Grange DC, van **Zyl WH.** 2005. Construction of cellobiose-growing and fermenting *Saccharomyces cerevisiae* strains. *J Biotechnol* **120:**284–295.

Verdoes JC, van Diepeningen AD, Punt PJ, Debets AJM, Stouthamer AH, van den Hondel CAMJJ. 1994. Evaluation of molecular and genetic approaches to generate glucoamylase overproducing strains of *Aspergillus niger*. *J Biotechnol* **36:**165–175.

Wang L, Zhao S, Chen X-X, Deng Q-P, Li C-X, Feng J-X. 2018. Secretory overproduction of a raw starch-degrading glucoamylase in *Penicillium oxalicum* using strong promoter and signal peptide. *Appl Microbiol Biotechnol* **102:**9291–9301.

Wang Q, Luan Y, Cheng X, Zhuang Q, Qi Q. 2015. Engineering of *Escherichia coli* for the biosynthesis of poly(3-hydroxybutyrate-*co*-3-hydroxyhexanoate) from glucose. *Appl Microbiol Biotechnol* **99:**2593–2602.

Wayman M, Chen S, Doan K. 1992. Bioconversion of waste paper to ethanol. *Process Biochem* **27:**239–245.

Weber JM, Leung JO, Swanson SJ, Idler KB, McAlpine JB. 1991. An erythromycin derivative produced by targeted gene disruption in *Saccharopolyspora erythraea*. *Science* **252:**114–117.

Weber T, Charusanti P, Musiol-Kroll EM, Jiang X, Tong Y, Kim HU, Lee SY. 2015. Metabolic engineering of antibiotic factories: new tools for antibiotic production in actinomycetes. *Trends Biotechnol* **33:**15–26.

Weimer A, Kohlstedt M, Volke DC, Nikel PI, Wittmann C. 2020. Industrial biotechnology of *Pseudomonas putida*: advances and prospects. *Appl Microbiol Biotechnol* **104:**7745–7766.

Yang D, Park SY, Park YS, Eun H, Lee SY. 2020. Metabolic engineering of *Escherichia coli* for natural product biosynthesis. *Trends Biotechnol* **38:**745–765.

Yim KS, Lee SY, Chang HN. 1996. Synthesis of poly-(3-hydroxybutyrate-co-3-hydroxyvalerate) by recombinant *Escherichia coli*. *Biotechnol Bioeng* **49:**495–503.

Yin S, Li Z, Wang X, Wang H, Jia X, Ai G, Bai Z, Shi M, Yuan F, Liu T, Wang W, Yang K. 2016. Heterologous expression of oxytetracycline biosynthetic gene cluster in *Streptomyces venezuelae* WVR2006 to improve production level and to alter fermentation process. *Appl Microbiol Biotechnol* **100:**10563–10572.

Yoon S-H, Kim J-E, Lee S-H, Park H-M, Choi M-S, Kim J-Y, Lee S-H, Shin Y-C, Keasling J-D, Kim S-W. 2007. Engineering the lycopene synthetic pathway in *E. coli* by comparison of the carotenoid genes of *Pantoea agglomerans* and *Pantoea ananatis*. *Appl Microbiol Biotechnol* **74:**131–139.

Yoshida S, Hiraga K, Takehana T, Taniguchi I, Yamaji H, Maeda Y, Toyohara K, Miyamoto K, Kimura Y, Oda K. 2016. A bacterium that degrades and assimilates poly(ethylene terephthalate). *Science* **351:**1196–1199.

Yoshida A, Nishimura T, Kawaguchi H, Inui M, Yukawa H. 2005. Enhanced hydrogen production from formic acid by formate hydrogen lyase-overexpressing *Escherichia coli* strains. *Appl Environ Microbiol* **71:**6762–6768.

Zhang M, Eddy C, Deanda K, Finkelstein M, Picataggio S. 1995. Metabolic engineering of a pentose metabolism pathway in ethanologenic *Zymomonas mobilis*. *Science* **267:**240–243.

review questions

1. Describe a strategy for isolating the gene for the restriction endonuclease EcoRI.
2. Outline a strategy for cloning the gene for 2,5-diketo-D-gluconic acid reductase from *Corynebacterium* into *Pantoea*. Why is this useful?
3. Suggest a strategy for improving the commercial utility of a cloned 2,5-diketo-D-gluconic acid reductase gene.
4. How can indigo be produced in *E. coli?*
5. How can genetic manipulation be used to increase the bacterial synthesis of L-cysteine? L-valine? L-citrulline?
6. How can the genome of *Corynebacterium glutamicum* be modified to make it a more effective host bacterium for the overproduction of various amino acids?
7. How can *E. coli* be engineered to produce lycopene?
8. Why is it difficult to genetically transform various *Streptomyces* spp.? How can this difficulty be overcome?
9. Suggest a strategy for isolating some of the genes that are involved in the biosynthesis of the red-colored antibiotic undecylprodigiosin, which is normally synthesized by *S. coelicolor*.
10. Suggest an approach for producing modified versions of polyketide antibiotics, such as erythromycin.
11. How can a drug that lowers cholesterol levels decrease *Staphylococcus aureus* infections?
12. What is a riboswitch? How can compounds that bind to riboswitches be used as antibiotics?
13. What effect does the cyclic AMP receptor protein have on antibiotic biosynthesis in various *Streptomyces* strains?
14. How can *Streptomyces* strains be engineered to produce tandem copies of a cluster of antibiotic biosynthesis genes in order to increase the level of antibiotic synthesized?
15. How would you significantly increase the production of the antibiotic oxytetracycline starting with a strain of *Streptomyces rimosus* that synthesizes this antibiotic?
16. What is whey? How can it be used to produce industrially important compounds?
17. How can polyhydroxyalkanoates be produced from whey?
18. Suggest a scheme for producing poly(3-hydroxybutyric acid) in *E. coli*.
19. How can a wide variety of variants of polyhydroxyalkanoates be synthesized?
20. How can a wild-type bacterium be modified to increase its solvent tolerance?
21. Briefly, outline the steps that could be used to significantly increase the synthesis of a desired metabolite from a host microorganism.
22. What is a "superbug"?

23. How can pesticide-degrading enzymes be expressed on the surface of a bacterium?

24. Explain how a *Pseudomonas* strain that carries plasmid pWWO and does not normally degrade 4-ethylbenzoate can be genetically manipulated to hydrolyze this compound.

25. How would you expand the substrate range of a strain of *Burkholderia* sp. that normally degrades 2,4-dinitrotoluene?

26. How do bacteria break down some plastics in the environment?

27. Outline a protocol that you would use to clone fungal cellulase genes.

28. Suggest schemes for the isolation of prokaryotic endoglucanase and β-glucosidase genes.

29. Delineate the role of α-amylase and glucoamylase in the industrial production of alcohol. How might genetic manipulation of the genes encoding these enzymes be used to improve this process?

30. How would you use an unpurified enzyme to digest various raw starches?

31. How would you alter cellular ATP-dependent pumps in order to increase alcohol tolerance in yeast strains?

32. What is glucose isomerase? Why is it important? How and why would you modify the gene encoding this enzyme?

33. Elaborate some of the advantages and disadvantages of using *Z. mobilis* instead of *S. cerevisiae* for alcohol production. How would you improve the industrial performance of *Z. mobilis*?

34. What is a designer cellulosome? How is it produced?

35. How would you create a designer cellulosome to degrade all three components of plant cell walls, i.e., cellulose, hemicellulose, and lignin?

36. How would you engineer yeast strains to more efficiently convert glucose into ethanol?

37. How would you engineer a bacterium to directly convert biomass into ethanol?

38. How can *Z. mobilis* be engineered to produce ethanol from xylose and arabinose?

39. How would you engineer glucoamylase to be more efficient in degrading starch?

40. How would you genetically engineer a cyanobacterium to produce large amounts of free fatty acids?

41. How would you engineer *E. coli* to produce hydrogen gas from formic acid?

42. How can a bacterium be used to convert hydrogen into formic acid?

43. How would you modify the green alga *Chlamydomonas reinhardtii* to more efficiently produce hydrogen from sunlight and water?

Principles of Microbial Growth
Batch Fermentation
Fed-Batch Fermentation
Continuous Fermentation
Maximizing the Efficiency of the Fermentation Process
High-Density Cell Cultures
Increasing Plasmid Stability
Quiescent *E. coli* Cells
Protein Secretion
Reducing Acetate
Improving Antibody Production in *E. coli*
Bioreactors
Typical Large-Scale Fermentation Systems
Two-Stage Fermentation in Tandem Airlift Reactors
Two-Stage Fermentation in a Single Stirred-Tank Reactor
Batch versus Fed-Batch Fermentation
Harvesting Microbial Cells
Disrupting Microbial Cells
Downstream Processing
Inclusion Bodies
Utilizing an Immobilized Enzyme
Magnetic Separation of Proteins
Large-Scale Production of DNA and RNA
Plasmid DNA
mRNA
SUMMARY
REFERENCES
REVIEW QUESTIONS

9 Large-Scale Production of Proteins and Nucleic Acids from Recombinant Microorganisms

In the early days of molecular biotechnology, many biologists naively thought that scale-up was simply a matter of multiplication; i.e., they believed that whatever conditions were found to be effective for microbial growth on a small scale would be equally effective on a large scale and that to achieve this it was merely necessary to use a larger reaction vessel with a correspondingly larger volume of growth medium. However, this view is far from reality. For example, good growth of aerobic microorganisms can usually be achieved in a standard 200-ml laboratory flask that is aerated with a mixer driven by a 300-watt motor. If the system were directly scaled up, a single 10,000-liter container would require a mixer with a 15-megawatt motor. Such a motor would be as large as a house, and the heat generated during stirring would boil the microorganisms. Although biochemical engineers may quibble about some aspects of this specific example, they all know that the industrial production of microorganisms is not merely a multiplication of bench-scale conditions. For a start, increasing the size of the reaction vessel (bioreactor or fermenter) is required for the large-scale growth of microorganisms, because it would be impractical to set up 50,000 individual culture flasks, each containing 200 ml, to obtain 10,000 liters of cell suspension.

A number of parameters must be precisely regulated to obtain maximum yields of cells from either small (1- to 10-liter) or large (>1,000-liter) bioreactors. These parameters include the temperature, pH, rate and nature of mixing of the growing cells, and with aerobic organisms, oxygen demand.

Molecular Biotechnology: Principles and Applications of Recombinant DNA,
Sixth Edition. Bernard R. Glick and Cheryl L. Patten.
© 2022 American Society for Microbiology.

Expression of Intracellular Hemoglobin Improves Protein Synthesis in Oxygen-Limited *Escherichia coli*

C. Khosla, J. E. Curtis, J. DeModena, U. Rinas, and J. E. Bailey
Bio/Technology **8:**849–853, 1990

milestone

Because of the low solubility of oxygen in water, the growth of aerobic bacteria often becomes limited by the amount of dissolved oxygen in the fermentation broth. This problem is especially acute at high cell densities or when fermentations are performed on a large scale. To address this problem, chemical engineers have attempted to increase the rate of transfer of introduced oxygen to the liquid of the growth medium. Their approaches have included (i) sparging the growth medium with pure oxygen rather than with air, (ii) introducing the air (or oxygen) under pressure, (iii) adding chemicals, such as perfluorocarbons, to the fermentation broth to increase the solubility of oxygen, and (iv) modifying the configuration of the fermentation vessel to optimize the aeration or agitation of the fermentation broth. While all of these solutions to the "oxygen problem" are somewhat effective, they are subject to a threshold effect beyond which it is impossible to introduce a sufficient amount of oxygen to improve the final yield of the fermentation.

As an alternative to these "hardware" solutions, Bailey and coworkers designed a biological system in which the host organism was modified so that it would be more efficient at using the low levels of oxygen that are normally present in the growth medium. They cloned a gene encoding a hemoglobin-like molecule from the Gram-negative bacterium *Vitreoscilla* and introduced it into several different recombinant bacteria. The introduced bacterial hemoglobin bound oxygen from the environment and created a higher level of available oxygen within the cells, which resulted in an increase in growth and foreign-gene expression. This approach provided a clever biological solution to what at first glance seemed to be an almost insurmountable engineering problem.

Moreover, the optimal conditions generally change with each 10-fold increase in the volume of a bioreactor.

There are also other technical considerations. The design of the bioreactor is important. It should ensure adequate sterility and provide appropriate levels of containment of genetically engineered microorganisms. The reactor should also include monitoring devices that permit the accurate and continuous online measurement of as many critical reaction parameters as possible so that adjustments can be made rapidly and easily throughout the course of the fermentation reaction (i.e., the growth of the microorganism). In addition, because sterilization may alter the composition of the medium (e.g., by destroying vitamins), it is important to ascertain that the medium composition is still optimal for maximal microbial growth following sterilization.

Generally, large-scale fermentation and product purification are stepwise processes (Fig. 9.1). A typical procedure begins with formulation and sterilization of the growth medium and sterilization of the fermentation equipment. Pure cultures of recombinant bacterial cells are typically stored at cold temperatures of approximately –80°C (conditions under which these organisms are extremely stable) for long periods of time. The cells are grown first as a stock culture (5 to 10 ml), then in a shake flask (200 to 1,000 ml), and then in a seed fermenter (10 to 100 liters). Finally, the production fermenter (1,000 to 100,000 liters) is inoculated. After the fermentation step is completed, the cells are separated from the culture fluid by either centrifugation or filtration. If the product is intracellular, the cells are disrupted, the cell debris is removed, and the product is recovered from the debris-free fluid. If the product is extracellular, it is purified from the cell-free culture medium.

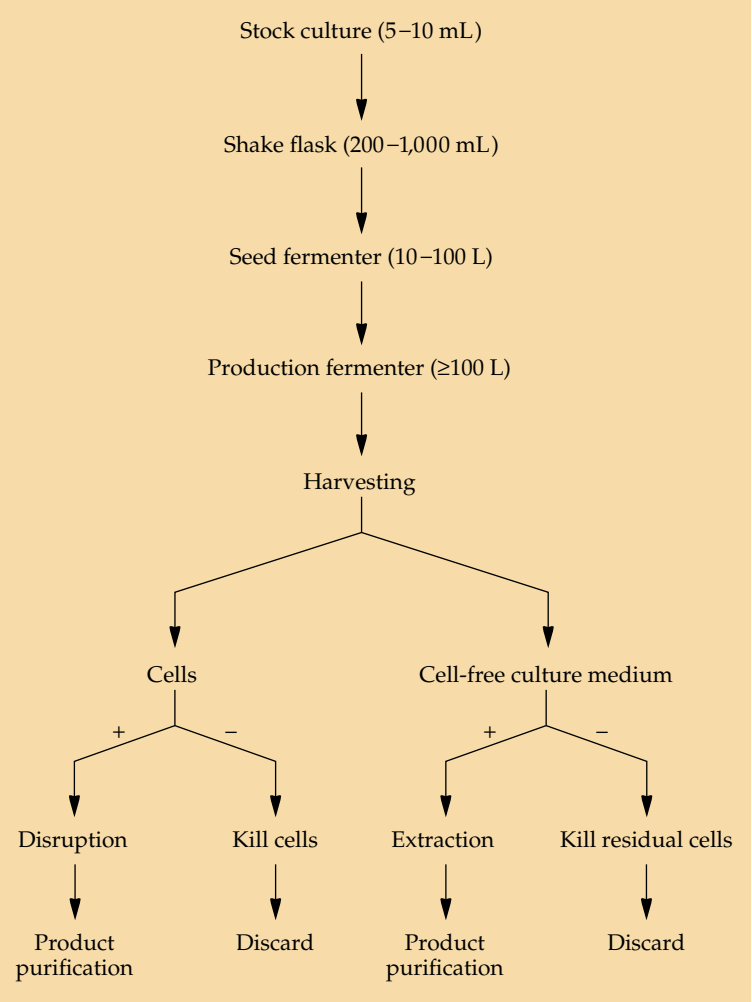


Figure 9.1 Generalized scheme for a large-scale fermentation process. The commercial product is usually in either the cell or cell-free fraction, but not in both; consequently, one or the other of these fractions will be processed further (+) or discarded (–).

Principles of Microbial Growth

Microorganisms can be grown in batch, fed-batch, or continuous culture. In batch fermentation, the sterile growth medium is inoculated with the appropriate microorganisms, and the fermentation proceeds without the addition of fresh growth medium. In fed-batch fermentation, nutrients are added incrementally at various times during the fermentation reaction; no growth medium is removed until the end of the process. In the continuous fermentation process, fresh growth medium is added continuously during fermentation, but there is also concomitant removal of an equal volume of spent medium containing suspended microorganisms. For each type of fermentation, oxygen (which is usually provided in the form of sterile air), an antifoaming agent, and if required, acid or base are injected into the bioreactor as needed.

Batch Fermentation

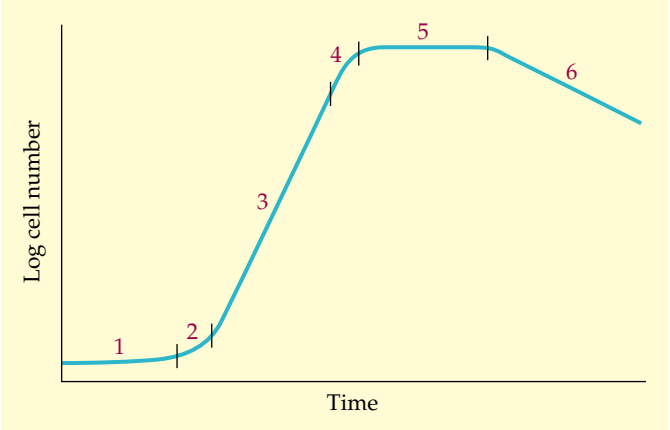
During a batch fermentation, the composition of the culture medium, the concentration of microorganisms (biomass concentration), the internal chemical composition of the microorganisms, and the amount of either target protein or metabolite all change as a consequence of the state of cell growth, cellular metabolism, and availability of nutrients. Under these conditions, six typical phases of growth are usually observed: lag phase, acceleration phase, logarithmic (log) or exponential phase, deceleration phase, stationary phase, and death phase (Fig. 9.2).

Typically, there is no immediate increase in the numbers of cells after the inoculation into sterilized growth medium. This initial period is called the lag phase. During the lag period, the microbial cells adapt to the new environmental conditions. The cells may have to adjust to a different pH or to a new level of available nutrients. As part of an adaptive response, previously unexpressed metabolic pathways may be induced. A lag phase generally occurs whenever the cells of the inoculum are derived from a culture that has stopped growing (i.e., has entered stationary phase) because of substrate limitation or product inhibition. These cells need time to reset their gene expression to adjust to the new medium and conditions. The length of the lag phase corresponds to how long the inoculated cells were in stationary phase and the extent to which the previous growth medium of the starting cells differed from the new, fresh culture medium. Conversely, when the inoculum is a cell culture from a growing cell population in log phase, a discernible lag phase may not occur, and growth may begin immediately. Following the lag phase, the brief period when the rate of cell growth increases until log-phase growth is attained is called the acceleration phase.

During the log phase of growth, the cell mass undergoes several doublings, and the specific growth rate of the culture remains constant. With excess substrate (nutrient supply) and no inhibition of growth by a compound that is present in the growth medium, the specific growth rate is independent of the substrate concentration. Changes to cell growth can be represented in mathematical form, making it possible for biochemical engineers to precisely model and then more easily scale up this process.

Sometimes, scientists refer to the doubling or generation time, t, of a culture rather than to its specific growth rate, μ, where $t = \ln 2/\mu$. The

Figure 9.2 Pattern of microbial cell growth in a batch fermenter. The six phases of the growth cycle are (1) lag, (2) acceleration, (3) log (exponential), (4) deceleration, (5) stationary, and (6) death.



generation time of a culture is the length of time that it takes, under defined conditions, for the number of cells or the cell biomass to double. For the majority of the microorganisms that are grown in culture, the value of $\mu_{max,}$ the maximal specific growth rate, varies from about 2.1 to 0.086 h^{-1} (reciprocal hours), which corresponds to doubling times of approximately 20 minutes to 8 hours.

The maximal rate of log-phase growth of the culture occurs when there is an excess of substrate, that is, when $S >> Ks$, where S is the concentration of substrate and Ks is a substrate-specific equilibrium constant. In practice, the value of Ks is often so low that substrate levels equivalent to Ks are rarely encountered during log-phase growth. For example, for *Escherichia coli*, while the Ks for glucose is approximately 1 mg/liter, the initial glucose level is usually around 10,000 mg/liter. However, as the culture nears the end of the log phase, the concentration of the remaining substrate, S, is depleted and may even fall below the value of Ks. Under conditions where $S < Ks$, the microorganisms rapidly enter the deceleration phase. However, because of the large cell population at the end of the log phase, the substrate may be so rapidly assimilated that the deceleration phase is short-lived and not observable.

After either the depletion of a critical growth substance, such as the carbon source, from the medium or the accumulation of metabolic end products that inhibit growth, the increase in cell mass eventually ceases and the cells enter the stationary phase. During this phase, although the amount of biomass remains constant, cellular metabolism often changes dramatically; in some instances, compounds (secondary metabolites) that are of considerable commercial interest are synthesized. For example, antibiotics are usually produced during the stationary phase of the microbial growth cycle. The duration of the stationary phase depends on the particular organism and the conditions of growth.

In the death phase, the energy reserves of the cell are virtually exhausted, and metabolic activity ceases. For most commercial processes, the fermentation reaction is halted, and the cells are harvested before the death phase begins.

Fed-Batch Fermentation

In fed-batch fermentations, substrate is added in increments at various times throughout the course of the reaction. Addition of nutrients is required for maintenance of the host cell metabolism and continued cell growth and to ensure that the synthesis and stability of recombinant protein are not impaired. The fermentation is monitored, and carbon and nitrogen sources together with trace elements may be added. These additions may prolong both the log and stationary phases, thereby increasing the biomass and the amount of synthesis of stationary-phase metabolites, such as antibiotics. However, microorganisms in stationary phase often produce proteolytic enzymes (proteases), and these enzymes can degrade foreign proteins synthesized by a genetically engineered microorganism. Therefore, when proteins are produced from a recombinant microorganism, it is important that the fermentation reaction not be allowed to reach this part of the growth cycle. Because it is often difficult to measure the substrate concentration directly during the fermentation reaction, other indicators that are correlated with the consumption of substrate, such as the production of organic acids, changes in the pH, or the production of CO_2, can be used to estimate when additional substrate is needed. Fed-batch fermentations require

more monitoring and greater control than batch fermentations, and many traditional industrial fermentations have been based on batch fermentation; therefore, batch protocols are used to a much greater extent. However, since fed-batch fermentations may be advantageous in the development of systems for the production of proteins from recombinant microorganisms, they are becoming increasingly popular. Depending upon the particular microorganism, its genetic background, and the nature of the recombinant protein, a fed-batch fermentation strategy can increase the yield from 25% to more than 1,000% compared with batch fermentation.

Fed-batch processes are not limited to microbial cells but are also used with mammalian and insect cells in culture. This is important because (i) these cell culture systems are increasingly being used for the production of human therapeutic proteins and (ii) in the absence of fed-batch strategies, animal cells in batch culture are not very efficient in producing foreign proteins.

Continuous Fermentation

In a continuous fermentation, a steady-state condition is attained when the total number of cells and the total volume in the bioreactor remain constant. In other words, under these conditions, the loss of cells due to outflow (product removal) is exactly balanced by the gain in new cells by growth (division).

The commercial use of continuous fermentation processes is currently limited to certain special cases such as production of antibiotics and organic solvents. This is primarily because of the greater experience that scientists have with growing cells in batch mode. However, the cost of producing a specific amount of cell biomass by continuous culture is potentially much lower than producing the same amount by batch fermentation. The following factors account for the savings.

- Continuous fermentations use smaller, less expensive bioreactors than batch fermentations to produce the same amount of product.
- After a large-scale batch fermentation is completed, large-scale equipment is needed for cell harvesting, cell breakage, and subsequent downstream processing (purification) of the protein or metabolic product that is produced by the microorganism. Continuously grown cells, however, are produced "a little bit at a time," so that the equipment required for cell harvesting, cell breakage, and downstream processing can be much (5- to 10-fold) smaller and therefore less expensive.
- Continuous fermentation, by definition, avoids the "down time" between batch runs, during which the bioreactor is cleaned, sterilized, and otherwise prepared for reuse. A common hindrance to efficient industrial fermentation is the loss of productivity due to the down time of the bioreactor. Continuous fermentations have much less down time, because a single reaction can be maintained for a much longer period.
- The physiological state of the cells during continuous fermentation is more uniform, so that yields of product are more consistent. In batch fermentations, small differences in the timing of cell harvest, which generally coincides with the mid- to late log phase of growth, can lead to significant physiological differences, which in turn may affect the ability to efficiently harvest the target product.

Despite its merits, continuous fermentation has potential drawbacks that must be addressed before its use becomes more widespread.

- The duration of a continuous fermentation can be 500 to 1,000 hours (~20 to 40 days), and therefore, some cells might lose recombinant plasmid constructs. Cells that lack plasmids usually have a smaller metabolic load and divide faster than do plasmid-containing microorganisms, so that yields may decline with time because fewer cells are synthesizing the plasmid-encoded product protein. However, integration of the cloned gene into the genome of the host organism can avoid this problem (see chapter 3).
- Maintenance of sterile conditions on an industrial scale for long periods is difficult. Furthermore, continuous processes need sterile backup equipment, a requirement that can greatly increase capital costs.
- In an effort to keep the cost of the growth medium as low as possible, the composition of culture medium that is used for large-scale industrial fermentations is often not subject to the same level of quality assurance as that accorded laboratory medium components and therefore may vary from batch to batch. This variation can alter the physiology of the cells and decrease productivity.

Because batch fermentation has a proven history of reliability, and in most instances the equipment and protocols are well established and in place, there is reluctance to switch to another type of fermentation system, even though a continuous mode of operation may be the most efficient fermentation strategy. Nevertheless, a number of researchers have recently developed laboratory-scale (up to 10 liters) or pilot plant-scale (up to 1,000 liters) continuous or fed-batch processes for the production of proteins from recombinant microorganisms. Therefore, it is probably only a matter of time before the use of fed-batch and continuous fermentations become more widespread in industry.

Maximizing the Efficiency of the Fermentation Process

Regardless of the type of fermentation process that is used to grow cells, it is necessary to monitor and control culture parameters, such as the dissolved oxygen concentration, pH, temperature, and degree of mixing. Changes in any one of these parameters can have a dramatic effect on the yield of cells and the stability of the protein product.

Optimal growth of *E. coli* cells and many other microorganisms that are used as hosts for cloned genes usually requires large amounts of dissolved oxygen. Because oxygen is only sparingly soluble in water (0.0084 g/liter at 25°C), it must be supplied continuously—generally in the form of sterilized air—to a growing bacterial culture. However, the introduction of air into a bioreactor produces bubbles, and if the bubbles are too large, the rate of transfer of oxygen to the cells may be insufficient to support optimal growth. Thus, fermenter design should include provision for monitoring the dissolved-oxygen level of the culture, providing oxygen to the culture, and adequately mixing the culture to efficiently disperse the bubbles.

Most microorganisms grow optimally between pH 5.5 and 8.5. However, during growth in a bioreactor, cellular metabolites are released into the growth medium, a process that can change the pH of the medium.

Therefore, the pH of the medium must be monitored, and either acid or base must be added as needed to maintain a constant pH. Of course, the added acid or base must be well mixed into the fermentation broth so that the pH of the growth medium is the same throughout the entire reaction vessel.

Maintenance of the correct temperature is essential for the success of a fermentation reaction. Microorganisms grown at a temperature below the optimum grow slowly and have a reduced rate of cellular production (productivity). On the other hand, if the growth temperature is too high—but not lethal—there may be premature induction of the expression of the target protein, if it is under the control of a temperature-sensitive repressor, or induction of a heat shock (stress) response, which will produce cellular proteases that lower the yield of the protein product.

Adequate mixing of a microbial culture is essential for many aspects of a fermentation, including assurance of an adequate supply of nutrients to the cells and prevention of the accumulation of any toxic metabolic by-products in local, poorly mixed regions of the bioreactor. Effective mixing is relatively easily attained with small-scale cultures, but it is one of the major problems when the scale of fermentation is increased.

Agitation of the fermentation broth also affects other factors, such as the rate of transfer of oxygen from the gas bubbles to the liquid medium and then from the medium to the cells, efficient heat transfer, accurate measurement of specific metabolites in the culture fluid, and efficient dispersion of added solutions, such as acids, bases, nutrients, or antifoaming agents. On these grounds, it might be concluded that the more mixing there is, the better the growth. However, excessive agitation of a fermentation broth can cause hydromechanical stress (shear), which damages larger microbial or mammalian cells, and a temperature increase, which may also decrease cell viability. Thus, a balance must be struck between the need to provide thorough mixing and the need to avoid damage to the cells.

There is an additional consideration for scaled-up fermentations that has nothing to do with the technical aspects of the process but depends instead on whether a genetically engineered microorganism is being used. In most countries, specific rules and regulations must be followed when genetically engineered microorganisms are grown on a large scale. Although most recombinant microorganisms are not hazardous, it is nevertheless important to ensure that they are not inadvertently released into the environment. Therefore, fail-safe systems are used to prevent accidental spills of live recombinant organisms and to contain them if they occur. Furthermore, all recombinant microorganisms must be treated by an approved procedure to render them nonviable before they are discharged from the production facility. The spent culture medium must also be treated to ensure that it does not contain any viable organisms and that its disposal does not create an environmental problem.

High-Density Cell Cultures

A major objective of fermentation is to obtain the largest amount of product in a given volume in as short a time as possible. High cell densities are absolutely necessary for high productivity. Generally, when foreign proteins are produced by recombinant *E. coli*, the greater the final cell density, the greater the amount of product that is formed. In practice, cell

concentrations of more than 50 (and in a few cases more than 150) grams (dry weight) of cells per liter of culture have been obtained with fed-batch cultures of recombinant *E. coli*. The dry weight of *E. coli* cells is typically 20 to 25% of the wet weight.

One way to achieve a high density of *E. coli* cells is to optimize the growth medium. Some nutrients, including carbon and nitrogen sources, can inhibit cell growth if they are present at too high a concentration. Glucose (a common carbon source) is inhibitory above 50 grams per liter, ammonia (a common nitrogen source) is inhibitory above 3 grams per liter, iron is inhibitory above 1.15 grams per liter, magnesium is inhibitory above 8.7 grams per liter, phosphorus is inhibitory above 10 grams per liter, and zinc is inhibitory above 0.038 gram per liter. Therefore, merely increasing the amount of nutrients in the growth medium in a batch culture does not necessarily yield a high cell density. In addition, since the nutrients in complex media, such as peptone or yeast extract, can vary from one batch of medium to another, fermentations that use complex media are not always reproducible.

Acetate, which can be inhibitory to cell growth, is produced by *E. coli* both when the cells are grown under oxygen-limiting conditions and in the presence of excess glucose. The acetate problem can be minimized by using glycerol instead of glucose as a carbon source (glycerol costs ~$800 and glucose costs ~$300 U.S. a ton), lowering the culture temperature, or using an *E. coli* strain that has been genetically engineered to shunt acetate into less toxic compounds (see below).

E. coli and most other microorganisms that are used to express foreign proteins generally require oxygen for optimal growth. Unfortunately, oxygen has only a limited solubility in aqueous media. Thus, as the cell density of a growing culture increases, the cells rapidly deplete the growth medium of dissolved oxygen. When cells become oxygen limited, exponential growth slows, and the culture rapidly enters a stationary phase during which cellular metabolism changes. One consequence of the stationary phase is the production by the host cells of proteases that can degrade foreign proteins. Oxygen dissolves into the growth medium very slowly, so the amount of dissolved oxygen available to growing cells is often not increased fast enough. Researchers have tried to modify the fermenter (bioreactor) configuration to optimize the amount of dissolved oxygen. They have also increased the rate of introduction of air (sparging), increased the agitation of cells, and added chemicals to the growth medium to increase the solubility of oxygen. These efforts have all been tried in an effort to deal with the limited amount of dissolved oxygen. Also, pure oxygen rather than air, which is only approximately 20% oxygen, can be introduced into growing cell cultures. Cells can also be grown under pressure to increase the solubility of oxygen, which increases the rate of transfer of oxygen to the cells in the aqueous growth medium. However, all of these efforts have met with only limited success.

Alternatively, expression of the *Vitreoscilla* sp. hemoglobin gene in a number of different host organisms has been shown to significantly increase the uptake of oxygen by growing cells (Fig. 9.3) and to thereby increase the amount of product formed (Table 9.1). *Vitreoscilla* hemoglobin functions by tightly binding oxygen, especially when the concentration of oxygen is low, and providing it to the respiratory chain by directly interacting with the terminal respiratory cytochrome. In doing so, *Vitreoscilla* hemoglobin can enhance growth and heterologous protein production. For

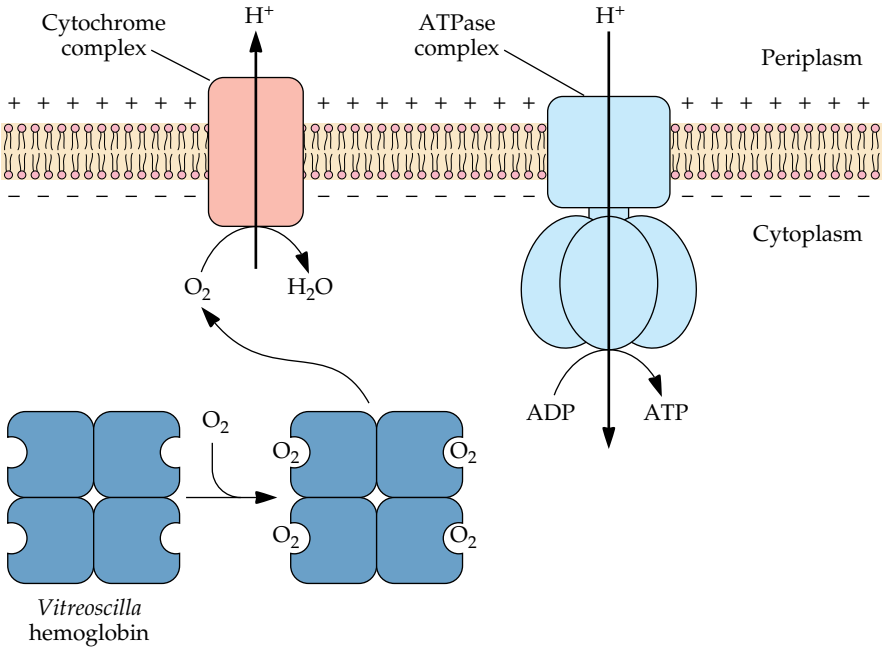


Figure 9.3 Schematic representation of the binding of O_2 by *Vitreoscilla* hemoglobin, the utilization of O_2 in pumping H^+ from the cytoplasm to the periplasm, and the subsequent coupling of H^+ uptake to ATP generation.

example, introduction of the *Vitreoscilla* hemoglobin gene increased erythromycin production by *Saccharopolyspora erythraea* and enhanced the production of cephalosporin C by *Acremonium chrysogenum*. It has been suggested that, in different heterologous hosts, the *Vitreoscilla* hemoglobin maintains bacterial cell metabolism at the energetically more efficient aerobic state by providing a steady supply of oxygen. More recently, *Vitreoscilla*

Table 9.1 Some examples of the use of *Vitreoscilla* hemoglobin to increase the synthesis of various bioproducts produced in different bacteria

Bacterial host	Bioproduct synthesized
Pseudomonas aeruginosa	Rhamnolipid
Saccharopolyspora spinosa	Spinosad
Bacillus amyloliquefaciens	Poly-γ-glutamic acid
Aeromonas hydrophila	Hydroxyalkanoates
Citrobacter freundii	Methionine γ-lyase
Corynebacterium crenatum	L-arginine
Sphingomonas elodea	Gellan gum
Streptomyces gilvosporeus	Natamycin
Streptomyces diastatochromogenes	Toyocamycin
Citrobacter freundii	Methionine γ-lyase
Streptomyces sp.	Candicidin D
Escherichia coli	Biomass
Escherichia coli	Ethanol
Saccharopolyspora erythraea	Erythromycin
Acremonium chrysogenum	Cephalosporin

hemoglobin has been engineered to increase the production of foreign protein in several yeasts and fungi as well as plant cells.

Finally, high-density cell cultures are most readily attained in fed-batch cultures. The addition of nutrients to replace consumed nutrients that become depleted may be constant, stepwise, or exponential. With stepwise feeding, increasing amounts of nutrients are added at increasingly higher cell concentrations. It is possible to automate the fed-batch addition of nutrients based on the measured concentration of a growth-limiting substrate, such as glucose, in the culture medium during the fermentation process.

Increasing Plasmid Stability

The loss of plasmids during the large-scale growth of recombinant *E. coli* cells is a major industrial problem. Plasmid loss limits the yield of plasmid-encoded recombinant proteins, especially when cells are grown in continuous culture. Plasmid instability in bacterial cultures is typically a consequence of the unequal distribution of plasmids to daughter cells during growth and cell division. Generally, once cells have lost a plasmid, they grow faster, with the result that cells lacking plasmids eventually dominate the culture. One approach to avoid this problem is to include an antibiotic resistance gene on the plasmid being used and then add that antibiotic to the culture medium. In addition to the obvious economic cost of the antibiotic, especially when dealing with large-scale cultures, disposal of spent growth medium is a potential environmental hazard in that both the antibiotic and antibiotic resistance genes may be released into the environment. One way to get around this problem is to delete an essential gene from the chromosomal DNA of the host bacterium and at the same time place this gene on the plasmid that is being stabilized. As a result, only plasmid-carrying cells can grow, making the bacterial strain totally dependent upon maintenance of the plasmid. In one example, the essential gene that was used encoded translation initiation factor 1 (Fig. 9.4). The target gene was placed on the plasmid under the transcriptional control of the strong and IPTG (isopropyl-β-D-thiogalactopyranoside)-inducible p^{trc} promoter. With this system, selection that utilizes antibiotics is no longer necessary, thereby decreasing both the cost and the environmental risk associated with large-scale fermentation.

Quiescent *E. coli* Cells

While it is possible to achieve high levels of foreign-gene expression in *E. coli* and other bacterial host cells, it is difficult to engineer recombinant bacteria to produce large amounts of a foreign protein and, at the same time, to grow to a high cell density. This is because a recombinant bacterial cell partitions its resources between production of the foreign protein and cell growth. Nevertheless, it is possible to first grow cells to a high density and then to shift the allocation of available resources from growth to foreign-protein production. With this in mind, one group of workers engineered a quiescent cell expression system in which a plasmid-encoded protein is expressed in nongrowing but metabolically active cells. The quiescent state is established by the overexpression of Rcd, a regulatory protein, in an *hns* mutant *E. coli* host cell. The *hns* gene codes for a histone-like nucleoid-structuring protein. Cultures of the *hns* mutant of *E. coli* in which the *rcd* gene is induced cause the *E. coli* chromosome to become condensed and to gradually cease synthesizing host proteins. However, plasmid

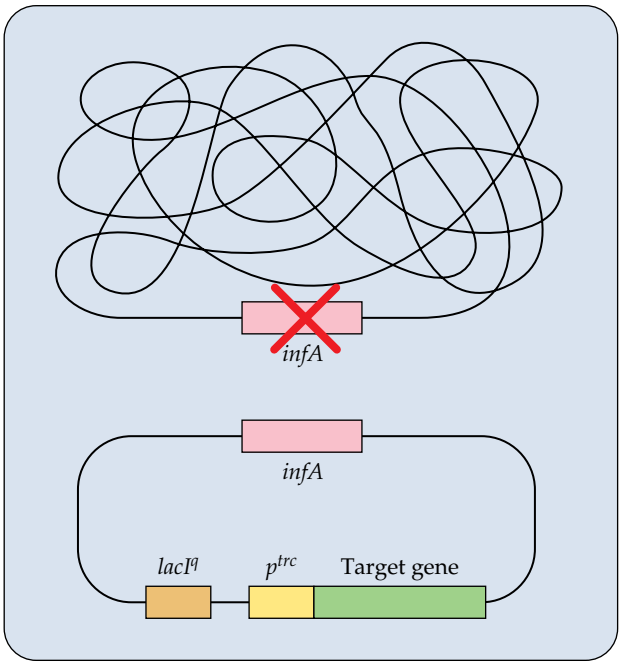


Figure 9.4 Schematic representation of a bacterial cell in which the essential protein synthesis initiator factor 1 gene (*infA*) was deleted from the chromosome and included on a plasmid. The target gene is under the transcriptional control of the strong Trc promoter (p^{trc}) that is controlled by the *lac* repressor encoded by *lacI*^{q}, which overproduces the protein.

genes are not affected, and there is, subsequently, continued synthesis of plasmid-encoded foreign proteins for many hours after induction; the quiescent state may be stable for up to 24 hours. Moreover, in many instances, the copy number of the plasmid cloning vector may be amplified in the nongrowing cells. In one study that utilized this system, the *rcd* gene was placed under the transcriptional control of the strong p^{R} promoter, while the recombinant protein gene (in this instance encoding a single-chain antibody variable fragment [scFv]) was controlled by the (equally strong) p^{L} promoter (Fig. 9.5). The activities of both the p^{R} and p^{L} promoters are repressed by a temperature-sensitive *c*I repressor protein; in this system, the gene for this protein is encoded in the host chromosomal DNA. When cells are grown at 30°C, the *c*I repressor prevents transcription from both p^{R} and p^{L}. When the temperature is shifted to 40 to 42°C, the temperature-sensitive *c*I repressor protein is inactivated, and transcription can proceed from both p^{R} and p^{L}. The temperature shift therefore causes the cells to simultaneously become quiescent and to synthesize the recombinant scFv protein. In this particular case, the scFv protein contained a leader peptide that directed the target protein to be secreted into the growth medium. As shown in Table 9.2, in both batch and fed-batch modes, the quiescent cells produce less biomass and secrete considerably more of the scFv protein into the growth medium than do control *E. coli* host cells engineered to express scFv under the control of the p^{L} promoter. Understanding the commercial potential of this unique system, the scientists who developed this technology have applied for and obtained patents to protect their intellectual property rights.

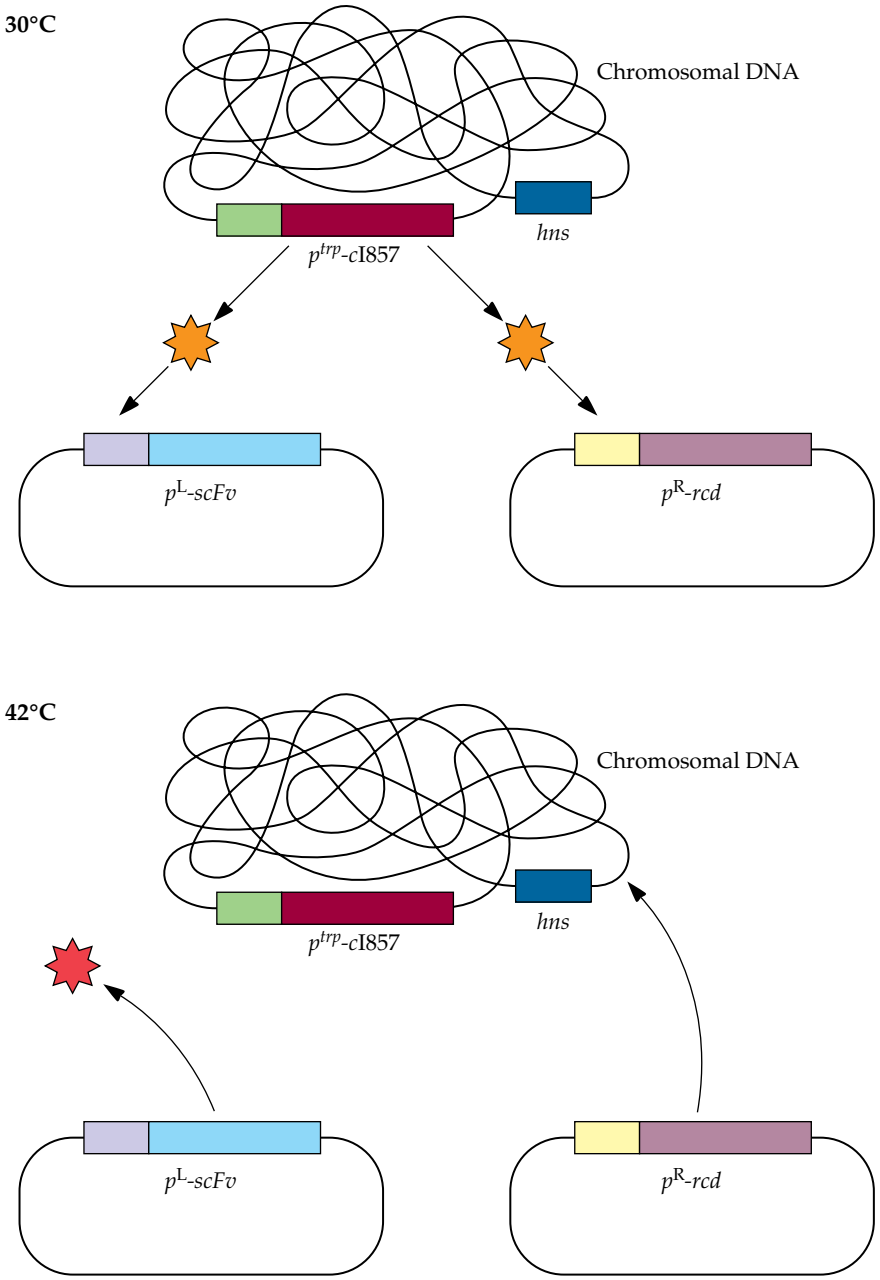


Figure 9.5 Synthesis of a single-chain antibody fragment (scFv) in "quiescent *E. coli* cells." At 30°C, the *c*I repressor protein (shown in orange), encoded by the *c*I857 gene (inserted into the chromosomal DNA), binds to the operators of the p^R and p^L promoters and prevents transcription of the plasmid-encoded *scFv* and *rcd* genes. At 42°C, the temperature-sensitive *c*I repressor protein is inactivated so that transcription directed by the p^R and p^L promoters can proceed. Turning on the p^R promoter causes the Rcd protein to be synthesized, thereby causing the cells to become quiescent because the Rcd protein interacts with the chromosomal mutant of the *hns* gene, resulting in the cessation of host cell protein synthesis. At the same time, turning on the p^L promoter activates transcription of the gene encoding scFv, resulting in the synthesis of the scFv protein (shown in red).

Table 9.2 Cell growth and foreign protein (scFv) secreted into the growth medium for *E. coli* cells with induced quiescence compared to control (wild-type) cells

Cell type	Batch culture		Fed-batch culture	
	Cell growth (optical density at 600 nm)	scFv secreted (mg/liter)	Cell growth (optical density at 600 nm)	scFv secreted (mg/liter)
Quiescent	3.5	37	20	150
Control	20	13	80	35

Protein Secretion

High-level expression of many different foreign proteins in *E. coli* results in the formation of cytoplasmic inclusion bodies consisting of insoluble improperly folded proteins. Even when the foreign protein is soluble, purifying it from a cytoplasmic extract can be a major undertaking. In addition, sometimes proteins that are secreted into the growth medium or the periplasm are readily produced at a much higher level than when they are expressed in the cytoplasm. While these considerations make only a small difference in laboratory-scale experiments, they are of critical importance when foreign proteins are produced on a large scale. Therefore, it may be preferable to engineer proteins for secretion into the growth medium or the periplasm. Note that periplasmic proteins are readily separated from the much larger number of cytoplasmic proteins.

One group of investigators observed that expression levels in *E. coli* were quite low for several different foreign proteins, i.e., human granulocyte-macrophage colony-stimulating factor, α-interferon 2b (IFN-α2b), or scFv, under the transcriptional control of the strong p^m/*xylS* promoter/regulator system. The yields of human granulocyte-macrophage colony-stimulating factor and scFv, but not IFN-α2b, increased dramatically when the genes encoding these proteins were fused to a translocation signal sequence (Fig. 9.6). Interestingly, different translocation signal sequences that were tested were optimally effective with each of the proteins tested. To obtain a high level of expression of IFN-α2b, before assembling the construct, it was necessary to chemically synthesize the gene in order to eliminate the use of codons that are rarely used in *E. coli*. While the use of translocation signal sequences significantly stimulated the levels of expression of these three human proteins, depending on the protein and the translocation signal sequence, from 20 to 50% of the protein that was produced was nevertheless found to be in an insoluble form. Thus, in order for this system to be used routinely for the large-scale production of human proteins in *E. coli*, a simple strategy (e.g., coexpression of a chaperonin protein) that minimizes the extent of insoluble protein formation must be employed (see chapter 3), or alternatively, the inclusion body must be efficiently solubilized (discussed later in this chapter).

Reducing Acetate

It is often difficult to achieve high levels of foreign-gene expression and a high cell density at the same time because of the accumulation of harmful waste products, especially acetate, which inhibits both cell growth and protein production and also wastes carbon and energy resources. One strategy to reduce the inhibitory effects of acetate is to remove the acetate from the culture during the course of the fermentation. This may be achieved

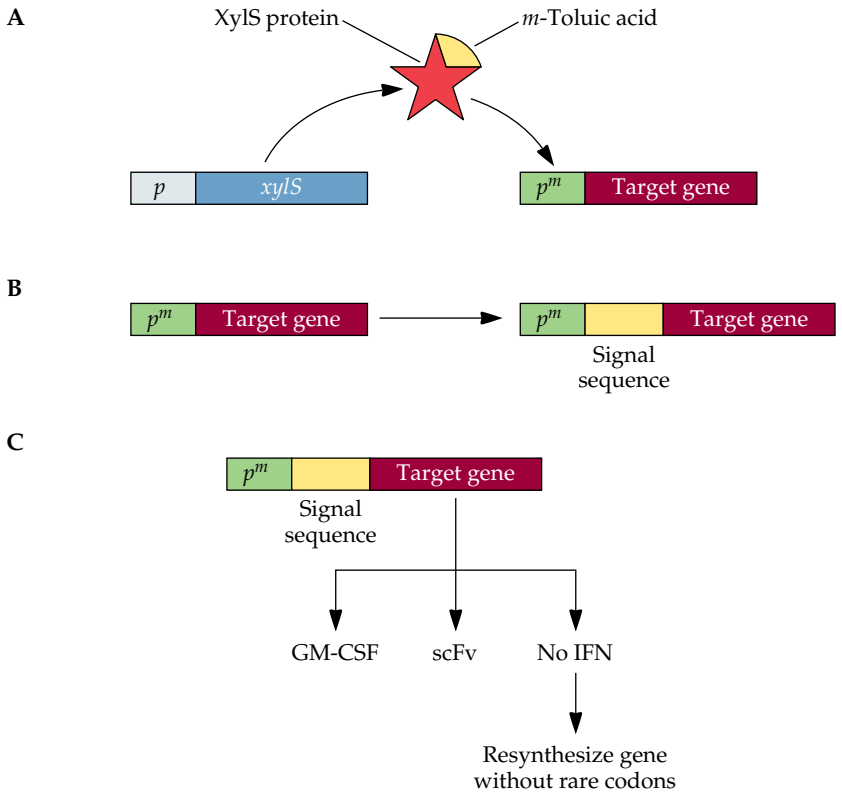


Figure 9.6 Engineering proteins for high-level expression following high-cell-density growth. **(A)** A target protein gene is placed under the transcriptional control of the strong and inducible (in this case by *m*-toluic acid) *pm* promoter. **(B)** The target gene is fused to any one of several different secretion signal sequences. **(C)** Rare codons are removed from protein-coding sequences that are not well expressed by chemically synthesizing the entire gene. GM-CSF, granulocyte-macrophage colony-stimulating factor; IFN, IFN-α2b.

by several different methods, including continuous dialysis and the use of macroporous ion-exchange resins. However, these methods may also remove nutrients necessary for cell growth along with the acetate.

Since acetate is often associated with the use of glucose as a carbon source, lower levels of acetate, and hence higher yields of protein, are generally obtained when fructose or mannose is used as a carbon source. Another strategy for reducing acetate accumulation in rich medium without impairing cell growth entails decreasing the glucose uptake rate of the cells by adding methyl α-glucoside, a glucose analogue, to the growing cells.

An alternative means of reducing acetate production in cells involves introducing a gene (from *Bacillus subtilis*) encoding the enzyme acetolactate synthase into *E. coli* host cells. This enzyme catalyzes the formation of acetolactate from pyruvate, thereby decreasing the flux through acetyl coenzyme A to acetate (Fig. 9.7A). The cells that were transformed with genes encoding the acetolactate synthase subunits produced 75% less acetate than the nontransformed cells and instead synthesized acetoin, which is approximately 50-fold less toxic to cells than acetate. The protein yield was also doubled.

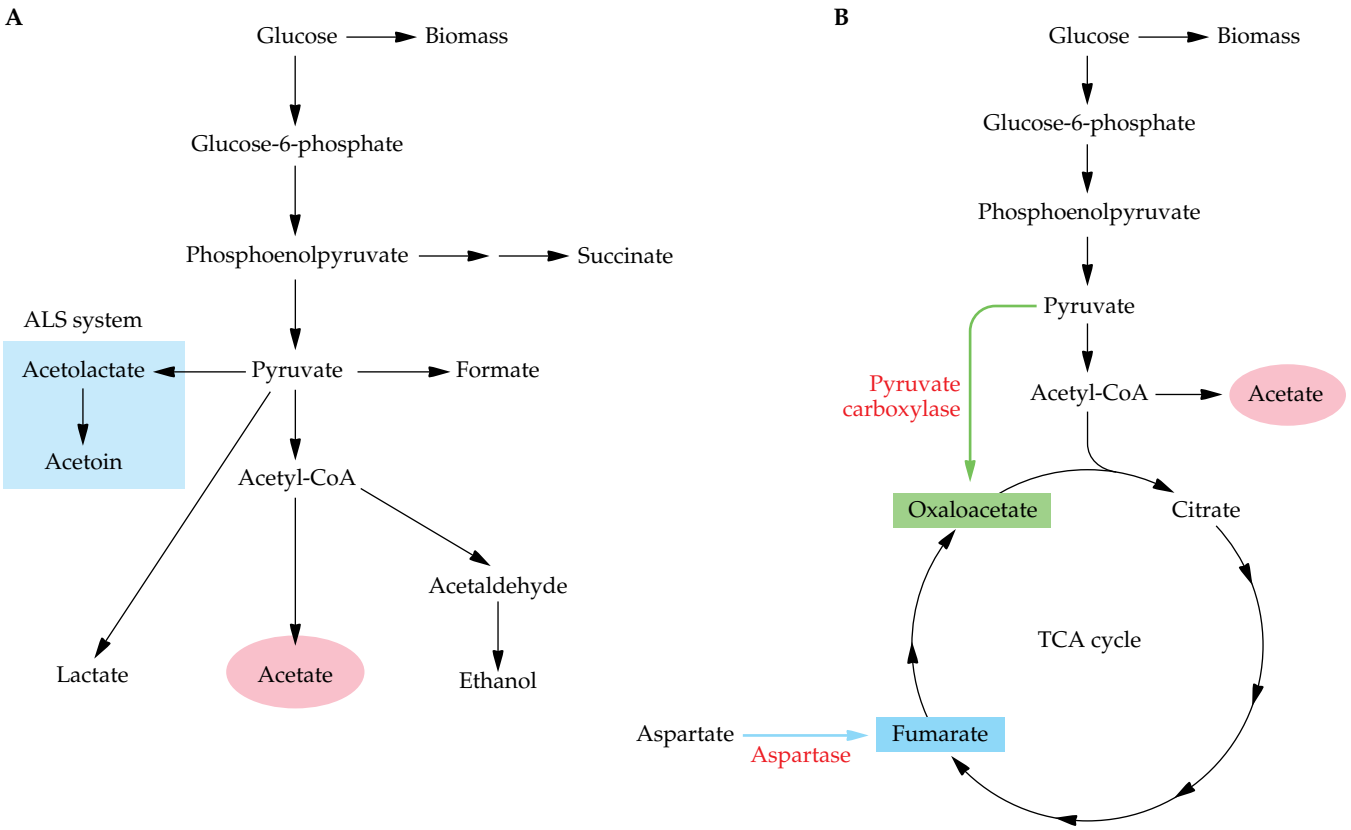


Figure 9.7 **(A)** Schematic representation of the pathways for glucose metabolism in an *E. coli* strain that has been transformed with a plasmid carrying the genes for the protein subunits of acetolactate synthase (ALS). Note that the conversion of glucose to biomass is a multistep process. CoA, coenzyme A. **(B)** Replenishment of the TCA cycle in *E. coli* by the introduction of a gene from *R. etli* encoding pyruvate carboxylase. This avoids the conversion of pyruvate to acetate. The TCA cycle may also be replenished by the introduction of a gene encoding aspartase, converting aspartate in the medium to fumarate. Note that the conversion of glucose to biomass is a multistep process. CoA, coenzyme A.

Another strategy to decrease acetate is to redirect carbon flow to the tricarboxylic acid (TCA) cycle (citric acid cycle, or Krebs cycle). This is a useful approach because recombinant protein production decreases carbon flow in the TCA cycle as a consequence of the withdrawal of the intermediates that serve as protein precursors (i.e., amino acids). One group of researchers transformed *E. coli* host cells with the gene for the enzyme pyruvate carboxylase, which converts pyruvate directly to oxaloacetate (Fig. 9.7B). Since *E. coli* does not normally contain pyruvate carboxylase, the gene was isolated from a strain of the Gram-negative bacterium *Rhizobium etli*. With the introduction of pyruvate carboxylase, acetate levels decreased, the cell yield increased, and the amount of foreign protein synthesized increased (Table 9.3). This result reflects the fact that the addition of pyruvate carboxylase allows *E. coli* cells to use the available carbon more efficiently, directing it away from acetate toward biomass and protein formation. Although it has not been tested extensively, it is thought that this strategy may be a generally effective method for increasing the level of expression of foreign proteins produced in *E. coli* and other bacterial host cells.

Table 9.3 Relative acetate levels, cell yields, and foreign-protein activities from *E. coli* cells transformed with an *R. etli* pyruvate carboxylase gene[a]

E. coli strain	Acetate concentration	Cell yield	Foreign-protein activity[b]
– Pyruvate carboxylase	1.0	1.0	1.0
+ Pyruvate carboxylase	0.43	1.41	1.68

[a]Data from March et al., *Appl Environ Microbiol* **68:**5620–5624, 2002.

[b]The foreign protein was β-galactosidase, whose activity is relatively easy to quantify. The data have been normalized to the values for the *E. coli* strain without (–) pyruvate carboxylase.

Similar to the strategy discussed above, the TCA cycle may also be replenished by converting aspartate to fumarate (Fig. 9.7B). To do this, *E. coli* host cells were transformed with the gene for L-aspartate ammonia lyase (aspartase) under the control of the strong *tac* promoter on a stable low-copy-number plasmid. Aspartase activity is induced by the addition of IPTG at the mid- to late log phase of growth. The target recombinant protein is introduced on a separate plasmid. Using this system in minimal medium containing aspartate, the production of different recombinant proteins could be increased up to 5-fold, with 30 to 40% more biomass production.

Improving Antibody Production in *E. coli*

The synthesis of whole antibody molecules in eukaryotic cell lines is relatively slow and expensive, requiring specialized media and equipment. Thus, it would be advantageous if antibody molecules could be produced inexpensively and rapidly in *E. coli* cells. The main factor limiting high levels of antibody molecule expression in bacterial cells is the ability to correctly fold and stoichiometrically assemble the two light and two heavy antibody peptide chains. Unfortunately, the cytoplasm of *E. coli* contains two disulfide bond-reducing pathways, which makes it difficult to efficiently assemble a complex protein such as a monoclonal antibody that contains several disulfide linkages between the protein subunits. The problem of forming disulfide bonds in complex proteins could be circumvented by decreasing the disulfide bond-reducing power by deleting the bacterial genes encoding thioredoxin reductase and glutathione reductase. Disulfide bond formation may be further improved by expressing a fusion of two endoplasmic reticulum proteins in the *E. coli* host, thiol peroxidase and protein disulfide isomerase. These altered *E. coli* host cells efficiently express full-length monoclonal antibodies; however, these antibodies are not glycosylated. With these alterations, the commercialized monoclonal antibody, Humira (which is used to treat rheumatoid arthritis), is expressed 4-fold better than in the parental *E. coli* strain. This preliminary study suggests that a wide range of complex proteins may be produced in this engineered *E. coli*.

Bioreactors

A cursory examination of the biochemical engineering literature may give the impression that there are a limitless number of bioreactor designs. However, closer inspection reveals that virtually all of these designs fall into three fundamental classes:

- Stirred-tank reactors (STRs), which have internal mechanical agitation (Fig. 9.8A)
- Bubble columns, which rely on the introduction of air or another gas (sparging) for agitation (Fig. 9.8B)

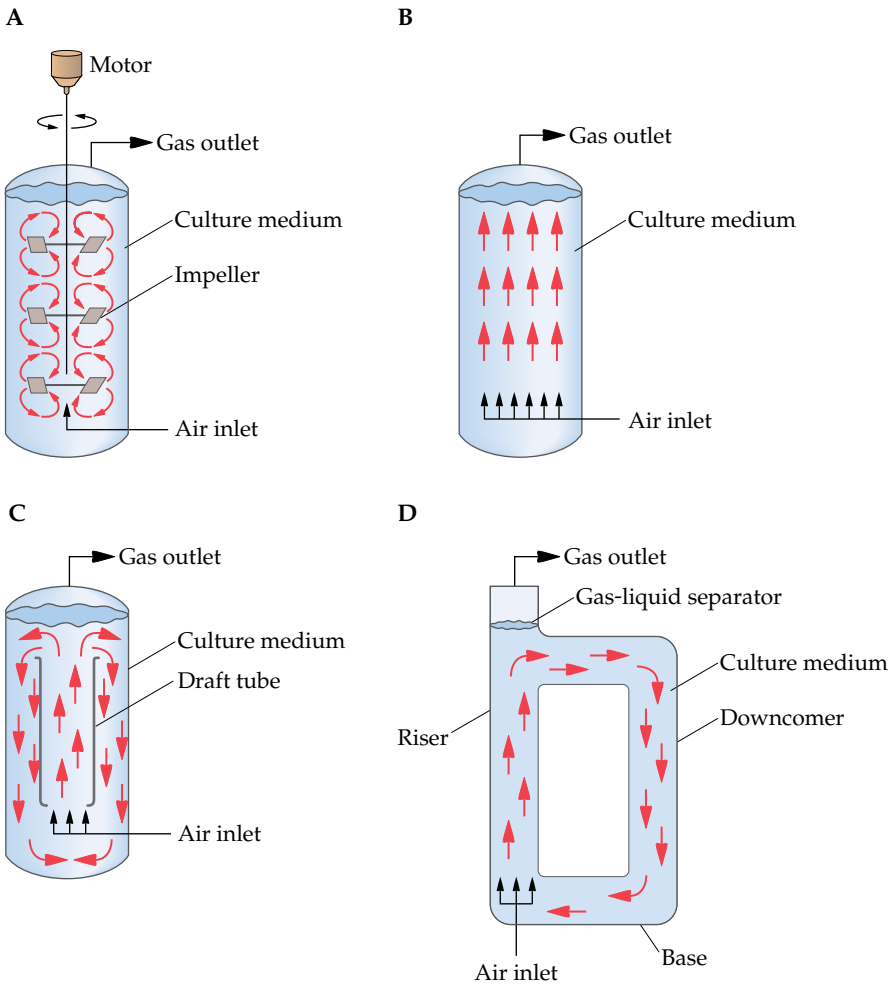


Figure 9.8 Simplified examples of bioreactor configurations. **(A)** STR; **(B)** bubble column; **(C)** internal-loop airlift reactor with a central draft tube; **(D)** external-loop airlift reactor. The arrows within the bioreactors indicate the direction of flow of the culture medium.

- Airlift reactors, which have either an internal (Fig. 9.8C) or an external (Fig. 9.8D) loop; the mixing and circulation of the culture fluid in these reactors are the results of the motion of an introduced gas (usually air), which causes density differences to be generated within the different parts of the bioreactor.

The traditional, and by far the most commonly used, bioreactor is the STR. This type of bioreactor has several advantages over other bioreactor configurations.

- It has highly flexible operating conditions.
- It is readily available commercially from several different manufacturers.
- It provides efficient gas transfer to the growing microbial cells, or in the words of fermentation engineers, the volumetric mass transfer coefficient, *kLa*, of STRs is high.

- It has been used extensively by fermentation engineers and microbiologists for growing a wide variety of microorganisms.

In an STR, gas, usually air, is added to the culture medium under pressure through a device called a sparger, which can be either a ring with many small holes or a tube with a single orifice. Although sparging rings generate smaller bubbles and consequently create better initial gas distribution, sparging tubes are often preferred in many small-scale applications (<20 liters) because they are less likely to become plugged. Thorough dispersion of the gas within the bioreactor requires one or more impellers (agitators) in addition to the sparger. Mechanical agitation of the culture medium by the impellers breaks larger bubbles into smaller ones, disperses the bubbles throughout the medium, and enhances the residence time of the bubbles in the bioreactor. At high levels of agitation, the mean size of the bubbles in large bioreactors is essentially independent of the size of the holes in the sparger. The type of impeller, its rotational speed, and the physicochemical properties of the liquid phase are important factors that give rise to efficient gas dispersion. In large bioreactors, however, if the initial gas distribution from the sparger is not uniform across the tank, even vigorous agitation may not create a homogeneous gas environment.

Because of the potentially corrosive or abrasive nature of many culture media and sterilization procedures, STRs are usually constructed from stainless steel or glass. The glass units are usually limited to laboratory-scale bioreactors that have a capacity of <50 liters.

One limitation on the size of a bioreactor is the ability of the system to efficiently remove heat that is generated as a consequence of either the metabolism of the growing cells or the energy input by agitation. Too much heat raises the temperature and alters the physiological state of the cells, potentially decreasing the product yield. Heat can be removed by using a cooling jacket around the reaction vessel or through the use of internal cooling coils. Although internal cooling coils are more effective than jackets in keeping the fermentation reaction close to the desired temperature, they can become fouled (coated) with microorganisms, which prevents cooling, and they sometimes interfere with the proper agitation of the fermentation broth.

Contamination of a fermentation with fungi or bacteria is usually disastrous. Therefore, bioreactors are designed so that they can be sterilized, usually with pressurized steam. There should be no internal dead spaces or surfaces that escape contact with the steam during sterilization. All seals, probes, and valves also must be readily steam-sterilizable. When a bioreactor is designed, a trade-off is often made between a full set of ports for probes, which enables monitoring of the fermentation parameters, and fewer ports, which makes the maintenance of sterility easier.

The high level of agitation of the culture medium during a fermentation reaction often causes considerable foaming. Excessive foam can wet the filter at the port through which the introduced air exits the bioreactor, thereby both decreasing the air flow and providing a potential pathway for the entry of contaminating cells. Either chemical antifoaming agents (such as some oils, silicones, alcohols, and stearates) or mechanical foam breakers can be used to control foaming. However, the chemical agents can diminish the extent of microbial growth by preventing oxygen transfer or, in some cases, by inhibiting cellular enzymes. Furthermore, if an antifoaming compound is not removed before downstream processing, it can contaminate the final product. Foaming can also be controlled to some extent

by providing sufficient "head space" within the bioreactor, i.e., the space above the liquid in which the bubbles can dissipate. In practice, then, the "working volume," or actual volume of the culture, in an STR is typically only about 75% of the total volume of the bioreactor.

Many of the considerations that apply to STRs also apply to pneumatic reactors, such as bubble columns and airlift bioreactors. Thus, for example, sterility, constant pH, and constant temperature are key components of any fermentation, regardless of the precise configuration of the bioreactor.

The configurations of bubble columns and airlift bioreactors give them some distinct advantages over STRs. These pneumatic reactors are more energy efficient than STRs because agitation is provided by the injection of a stream of air—or another gas if anaerobic microorganisms are being grown—rather than by a mechanical stirrer. Also, with the elimination of the mixer shaft in these units, there is one less potential site of entry for contaminating organisms.

Pneumatic reactors generate a lower-shear environment than do STRs. Also, in airlift reactors, the shear stress is more evenly distributed throughout the vessel than in STRs. The reduction of shear forces is important for the following reasons.

- Genetically engineered microorganisms are often more susceptible to lysis when exposed to shear stress than are unmodified organisms, because the extra metabolic burden of synthesizing a foreign protein often causes genetically engineered microorganisms to form weakened cell walls.
- A frequent cellular response to the shear forces is decreased synthesis of all cellular proteins, including the recombinant protein.
- Shear stress can alter the physical and chemical properties of the cells so that the downstream processing steps become more difficult to perform. For example, the fermentation conditions can inadvertently increase the amount of surface polysaccharides that a microorganism produces and, as a consequence, can change the conditions for effective harvesting and lysis of these cells, making it more difficult to purify the target protein.

In bubble columns, the air is introduced under high pressure near the bottom, but the smaller bubbles coalesce into larger ones as they rise through the column, leading to uneven gas distribution. In addition, the use of high-pressure air tends to cause excessive foaming of the medium. These disadvantages restrict the flexibility or effective range of operating conditions, as well as the potential size of bubble columns.

Airlift bioreactors, however, can be readily adapted for either pilot plant or large-scale fermentation processes. In an airlift reactor, the gas is introduced into the bottom of a vertical channel (riser). Both the gas and liquid flow up the riser until they reach an open space at the top (gas–liquid separator), where the gas is at least partially disengaged from the liquid. The degassed liquid, which is denser than the gassed liquid, descends in a separate vertical channel (downcomer) and moves along the base of the reactor until it reaches the bottom of the riser. In this way, the culture fluid and cells are continuously being circulated around the bioreactor.

There are two main types of airlift bioreactors: those that have a single container with internal baffles that create interior liquid circulation channels (internal-loop reactors) (Fig. 9.8C) and those that have an external

loop so that the culture liquid circulates through separate, independent channels (external-loop reactors) (Fig. 9.8D). Internal-loop airlift reactors are simple in design, but once they are constructed, both the volume and the circulation rate are fixed for all fermentation processes. In contrast, the external loop of external-loop airlift reactors can be easily changed or modified, e.g., by altering its volume, to suit the requirements of different fermentations.

Airlift bioreactors are generally more efficient than bubble columns, especially for denser or more viscous suspensions of microorganisms. In airlift reactors, mixing is generally better, and bubble coalescence is not as big a problem as it is in bubble columns. In extremely large airlift fermenters, such as the 1,500,000-liter fermenter built by Imperial Chemical Industries, Ltd., in England for the production of single-cell protein, it takes a considerable amount of time for a cell to complete a full cycle through the reactor. For such very large fermenters, to prevent the cells from becoming substrate depleted while they are traversing the bioreactor, there are multiple injection points for the introduction of substrate along the length of the unit.

Typical Large-Scale Fermentation Systems

When recombinant microorganisms are used to overproduce protein products, such as pharmaceuticals like insulin, cells are typically grown to the mid- to late log phase of the growth cycle so that the target protein levels will be optimal. On the other hand, when recombinant microorganisms are used as "factories" to synthesize useful metabolites, such as antibiotics, host cells are commonly grown to the deceleration or stationary phase, where the synthesis of secondary metabolites is often optimal. Clearly, these sorts of differences must be considered when a large-scale fermentation process is being developed.

For maximal protein production, it is generally preferable to use cloned genes that are under the control of strong promoters that can be regulated (see chapter 3). Initially, it was thought that constitutive expression of a cloned gene would be sufficient to obtain reasonable quantities of the product. However, experience has shown that continuous transcription and translation of a cloned gene drains energy from essential cell functions and slows cell growth. With an inducible system, the expression of a cloned gene can be confined to a specific period in the growth cycle of the microorganism. On these grounds, for optimal protein production, the process should consist of two separate stages. First, the cells are grown under optimum conditions to a relatively high cell density. Second, depending on the nature of the promoter that drives the cloned gene, transcription is induced either by shifting the temperature or by adding a chemical inducer, such as IPTG, to the medium.

A two-stage system is not easy to implement in a large bioreactor (>100 liters) because it is technically very difficult either to raise the temperature quickly, typically from 30 to 42°C, or to ensure that the chemical inducer is rapidly and evenly mixed in a large vessel. Moreover, many chemical inducers, such as IPTG, are too expensive to be used on a large scale. However, as discussed below, this problem can be overcome by using two connected bioreactors (two-stage fermentation) so that the cells are grown in the first vessel and the induction of recombinant protein

expression is carried out in the second. Under these conditions, growth and induction are separately optimized, thereby increasing the overall amount of product formed per unit of time (productivity) of the fermentation.

Two-Stage Fermentation in Tandem Airlift Reactors

E. coli strain NM989, which carries the heterologous gene encoding the enzyme T4 DNA ligase under the transcriptional control of the bacteriophage λ p^L promoter and a temperature-sensitive *c*I repressor, was grown and induced in a two-stage airlift bioreactor. In this bacterial strain, the T4 DNA ligase gene was integrated in the chromosomal DNA, a location obviating any potential problems of plasmid instability that might occur during extended fermentation. The growth stage was carried out at 30°C in an external-loop airlift bioreactor that had a 10-liter working volume. The T4 DNA ligase gene was not expressed under these conditions. A second external-loop airlift bioreactor, with a working volume of approximately 5 liters at 42°C, was used for the induction stage (Fig. 9.9). In contrast to other expression systems, an increase in temperature, to heat-induce the p^L promoter, triggers the bacterial host cell heat shock response that is controlled by an alternative sigma factor, σ32, encoded by the *rpoH* gene. The activation of this expression system dramatically increases the need for cellular ATP as a result of both protein overexpression and the heat shock response. Among other things, this results in a decreased bacterial growth rate and plasmid instability. To avoid or lessen the effect of these problems, researchers may employ either fed-batch or continuous fermentation strategies.

With this two-stage fermentation in tandem airlift reactors, and growing *E. coli* strain NM989, the two bioreactors were linked by a transfer tube with a pump that controlled the continuous flow of cell suspension from the growth stage bioreactor (first) into the induction stage vessel (second). In addition, cells suspended in the culture medium were removed from the induction stage bioreactor at a specific rate and prepared for downstream processing.

The maximal specific growth rate (μ_{max}) of the microbial culture was approximately 0.66 h^{-1} in the first bioreactor and 0.54 h^{-1} in the second. These values correspond to cell doubling times of 63 and 77 minutes, respectively. Fresh medium was continuously added at a rate of 2 liters per hour to the growth stage fermenter, and cell suspension (effluent) was simultaneously removed from the induction stage fermenter at precisely the same rate. As a consequence of the different liquid volumes of the two bioreactors and the different growth rates at the two different temperatures, an average cell spent about 5 hours in the growth stage bioreactor and 2 hours in the induction stage bioreactor. The different residence times in the two phases of this fermentation process were necessary to optimize both the number of cells produced and the yield and stability of the T4 DNA ligase. Generally, residence times can be altered as required by adjusting the relative working volumes in the two fermenters of a two-stage system and by adjusting the volumetric rate of input of nutrients into the first bioreactor.

The double-external-loop design of the airlift fermenter (Fig. 9.9) used in this work facilitated the adjustment of the working volumes of the two fermenters relative to one another. It also added versatility to the system, so that it was possible to obtain a variety of different growth conditions for different populations of recombinant cells. For T4 DNA ligase production,

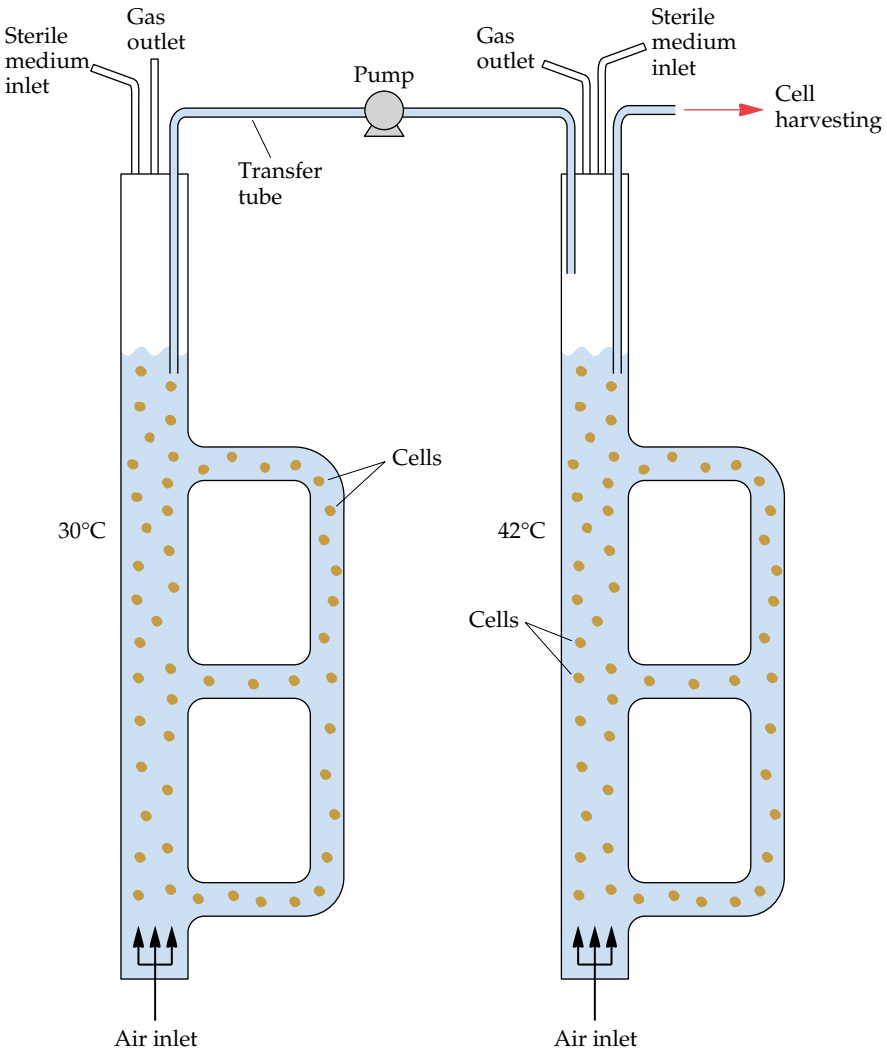


Figure 9.9 Two-stage airlift reactor used for the temperature-dependent induction of a protein product. Cells from the growth stage (left) fermenter, which is at 30°C, are pumped into the induction stage fermenter (right), which is at 42°C. Each bioreactor has a double external loop that is fitted with valves. By changing the valve settings, working volumes best suited for different fermentations can be created.

the best results were obtained when approximately 33 ml of cell suspension was transferred every minute from the growth stage bioreactor to the induction stage bioreactor. Because this amount of cell suspension was equivalent to only 0.67% of the volume of the induction stage bioreactor, the incoming cells underwent a virtually instantaneous temperature shift from 30 to 42°C. Nutrients in a concentrated form were added at a specific rate to the induction stage bioreactor throughout the fermentation to keep the cells in this bioreactor in log phase. This action prevented the T4 DNA ligase from being degraded by the proteolytic enzymes that are normally synthesized during the deceleration and stationary phases.

With this continuous two-stage bioreactor, it was found that induced *E. coli* NM989 could be grown to a density of approximately 4 grams (dry

weight) of cells per liter of culture. After induction, about 4% of the cell protein is T4 DNA ligase, an amount that corresponds to approximately 25,000 units of enzyme activity per gram (dry weight). This process can produce approximately 100,000 units of enzyme activity per liter of culture, or about 4,800,000 units per day. Assuming that it is possible to recover about 40% of the initial activity following the complete purification of the enzyme, and that the enzyme sells (retail) for about $0.005 per unit, then the final yield of purified enzyme in 1 day will be worth about $9,600. Although these calculations do not consider all of the costs that go into the production of recombinant proteins from a genetically engineered microorganism, it is clear that for highly valued products, such as T4 DNA ligase, small to moderate-size continuous fermentation systems can generate significant returns on the initial capital investment. It should be noted that this work was done several years ago when the retail cost of this enzyme was approximately 100 times greater than its current cost.

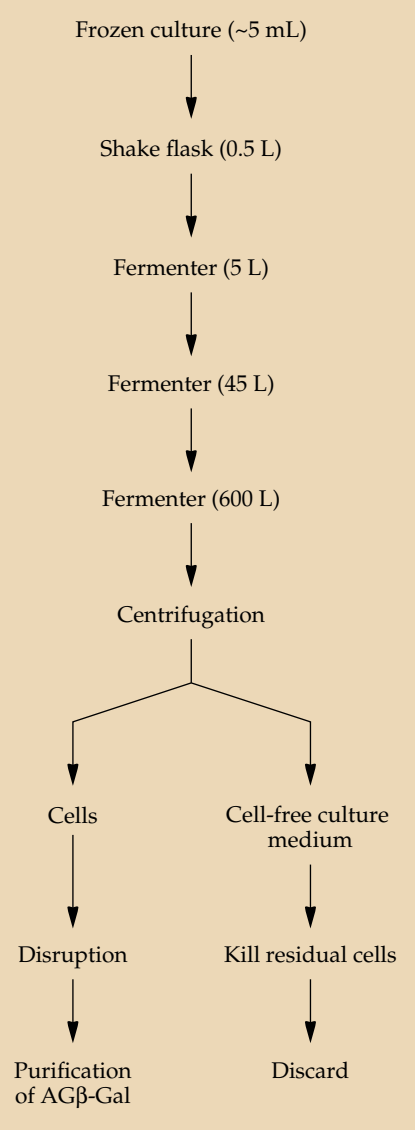


Figure 9.10 Scheme for the large-scale production of the protein AGβ-Gal. The volumes in parentheses indicate how much culture medium was used at each step of the process. The culture medium occupied about 60 to 75% of the total volume of each of the bioreactors (fermenters).

Two-Stage Fermentation in a Single Stirred-Tank Reactor

The tripartite fusion protein, AGβ-Gal, which is used for immunological assays, was produced on a large scale in a single STR. The gene encoding AGβ-Gal was constructed by recombinant DNA techniques and encodes the five immunoglobulin G-binding regions of *Staphylococcus aureus* protein A, two immunoglobulin G-binding regions from *Streptococcus* sp. strain G148 protein G, and β-galactosidase from *E. coli*. The synthetic AGβ-Gal gene was placed under the transcriptional control of the bacteriophage λ p^R promoter, which is negatively regulated by the temperature-sensitive *c*I repressor in the same manner as the p^L promoter, cloned into a plasmid that carries the gene for ampicillin resistance, and introduced by transformation into *E. coli*. The strain with the plasmid that encodes AGβ-Gal carries a second plasmid that has the gene for the *c*I repressor protein and a kanamycin resistance gene.

A 5-liter volume of these cells was grown in an STR at 30°C in the presence of both ampicillin and kanamycin—to provide selective pressure for the retention of both plasmids—and then used to inoculate a growth medium without the presence of antibiotics at 30°C in a 45-liter STR. The cell suspension from the 45-liter fermenter in turn served as an inoculum for a 600-liter STR, where the cells were grown at 30°C without any added antibiotics (Fig. 9.10). In general, to keep the cost of the process to a minimum, antibiotics are not added to large-scale microbial cultures. In addition, to gain regulatory approval for proteins that will be used as human therapeutic agents, the purified protein should not contain even small amounts of antibiotics. When the cell density in the 600-liter bioreactor reached the equivalent of about 4 grams (dry weight) per liter of growth medium, the temperature was shifted from 30 to 40°C to induce the expression of the AGβ-Gal protein. Under these conditions, it takes about 1 hour to reach 40°C. A temperature of 40°C, rather than 42°C, was chosen because the lower temperature was found to yield the same level of AGβ-Gal protein while allowing the cells to grow for a longer time. In other words, the slightly lower induction temperature (40°C) condition yielded a larger amount of protein product.

The production of the AGβ-Gal protein (measured as specific β-galactosidase activity) increased for 2 hours after the initiation of the

induction and then decreased. This decrease in enzyme activity was probably due to the synthesis of proteases by cells that had entered the deceleration and stationary phases of the growth cycle. In addition, after growth for 4 hours at 40°C, 50% of the cells had lost their plasmids. These problems notwithstanding, after 4 hours at 40°C, the AGβ-Gal protein was approximately 20% of the dry weight of the total biomass. Considering the very high yield of the target protein that was produced by this strategy, it is probably not necessary to integrate the genes for the AGβ-Gal protein and the *c*I repressor into the chromosomal DNA of the *E. coli* host cell in an effort to increase the final yield.

Batch versus Fed-Batch Fermentation

In some instances, a simple fed-batch strategy can be used to produce both a high cell density and a high level of expression of a target protein (Fig. 9.11). For example, a plasmid carrying a gene encoding a hybrid protein that includes the insulin B peptide under the control of the *E. coli trp* promoter was introduced into a *trp*-minus mutant strain of *E. coli* that cannot synthesize tryptophan, and the transformant was grown in media containing various amounts of tryptophan. At high levels of tryptophan, the synthesis of the target protein was repressed. However, after consumption of the tryptophan in the medium by the growing cells, the tryptophan repressor was released from the *trp* promoter, and synthesis of the target protein was induced. With this system, the addition of tryptophan to the medium resulted, in batch cultures, in increases in the amounts of both biomass and

Figure 9.11 Schematic representation of the amount of foreign protein produced as a function of time by recombinant *E. coli* following induction (arrow) in the mid-log phase of growth. Prior to induction, no foreign protein is synthesized. After induction in batch mode, in the absence of additional nutrients, the cells soon enter stationary phase and synthesize proteases that degrade the foreign protein product. After induction in fed-batch mode, the added nutrients ensure that the cells remain in log phase for an extended time and do not produce any proteases until 1.5 to 2 hours later than the cells in batch mode; therefore, the foreign protein is more stable and is easier to recover. In addition, the provision of nutrients in fed-batch mode makes it less likely that plasmids carrying foreign genes will be lost than in batch mode. The time represents the number of hours from the start of the fermentation.

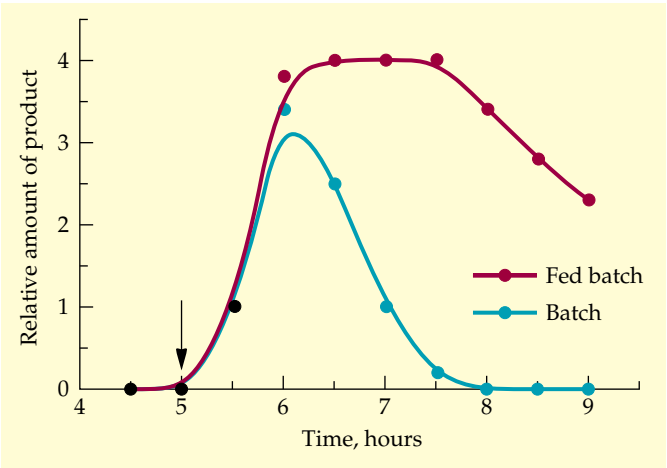


Table 9.4 Comparison of batch and fed-batch fermentations for the production of a fusion protein including the insulin B peptide[a]

Product[b]	Yield in fermentation system:		
	Batch	Batch + Trp	Fed batch
Biomass (g [DW]/liter)	6.7	12	20
Fusion protein (% of total protein)	4.6	7.9	11
Total amount of fusion protein (g/liter)	0.17	0.53	1.21
Plasmid-bearing cells (%)	86	62	90

[a]Data from Gosset et al., *Appl Microbiol Biotechnol* **39:**541–546, 1993.

[b]DW, dry weight. In the "Batch + Trp" fermentation, 0.1 g of tryptophan was added. In the fed-batch fermentation, 0.1 g of tryptophan was added every 2 hours, for a total of five times during the course of the 10-hour fermentation. A larger amount of tryptophan added to the batch fermentation did not increase the amount of either biomass or target protein produced.

target protein produced. However, fed-batch fermentation was more effective than batch fermentation with or without added tryptophan (Table 9.4).

In another experiment, IFN-γ was produced in *E. coli*. Expression of the IFN-γ gene was controlled by the p^L promoter regulated by the temperature-sensitive *c*I repressor. Cells that contained this construct on a plasmid were grown in either batch or fed-batch mode (with both stepwise and constant-rate medium-feeding strategies). In the fed-batch mode, the addition of growth medium was carried out simultaneously with the temperature induction of the p^L promoter in the late exponential growth phase. Use of the fed-batch mode resulted in a significant increase in the length of the cell growth phase following induction. This fed-batch strategy enabled researchers studying this system to achieve a cell biomass that was 5-fold higher and a final IFN-γ concentration that was 23-fold higher than they were able to achieve in batch culture.

Another group studied the fermentation conditions that yielded the optimum expression of a monoclonal antibody (Fab) fragment directed against tetanus toxoid under the transcriptional control of the *E. coli lac* promoter. The plasmid construct included signal sequences that were inserted immediately upstream of the antibody (light and heavy) genes to target the antibody fragments to the *E. coli* periplasm. In this case, inexpensive lactose, rather than the considerably more expensive IPTG, could be used to induce the expression of the monoclonal antibody fragment gene, provided that the host strain of *E. coli* carried the *lac* operon (i.e., was Lac^+) and therefore produced lactose permease to take up lactose from the medium and β-galactosidase to convert it to glucose and galactose. In fact, allolactose, an isomer of D-lactose (and not D-lactose itself), is the actual inducer of the *E. coli lac* operon and is formed only if the host strain of *E. coli* contains a small amount of β-galactosidase. An important facet of these experiments is that lactose not only acts as an inducer, but also is metabolized, providing an additional carbon source for the Fab-producing *E. coli* cells, thereby supporting both cell growth and product accumulation (Fig. 9.12). Again, in these experiments, fed-batch fermentation was clearly superior to batch fermentation, yielding both a larger amount of cell biomass and a greater concentration of Fab fragment.

In another study, researchers reported that instead of a single introduction of lactose in a fed-batch fermentation, to produce active protein fragments of human apolipoprotein(a) in *E. coli*, a strategy that employed continuous induction with lactose beneficially influenced the expression of the

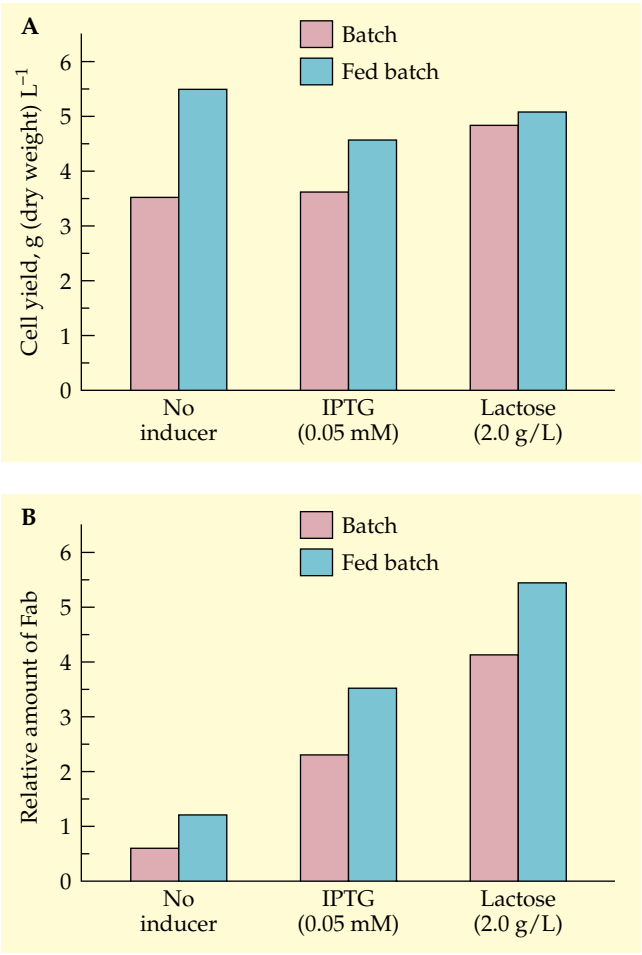


Figure 9.12 Induced recombinant protein production with batch and fed-batch *E. coli* cells grown to late logarithmic phase. The cells were either not induced, induced with 0.05 mM IPTG, or induced with 2.0 grams of lactose per liter. **(A)** In batch culture, the cells that were induced with lactose, which acts as both an inducer and a growth substrate, grew to a significantly greater extent than either the noninduced or the IPTG-induced cells. With fed-batch cells, the smaller extent of growth of induced versus noninduced cells probably reflects the resources that the induced cells direct to the synthesis of the Fab antibody fragment. **(B)** Under both batch and fed-batch conditions, more of the Fab antibody fragment was produced when the cells were induced with lactose than with IPTG. Fab antibody fragment synthesis in the absence of inducer represents incomplete repression of the *lac* promoter.

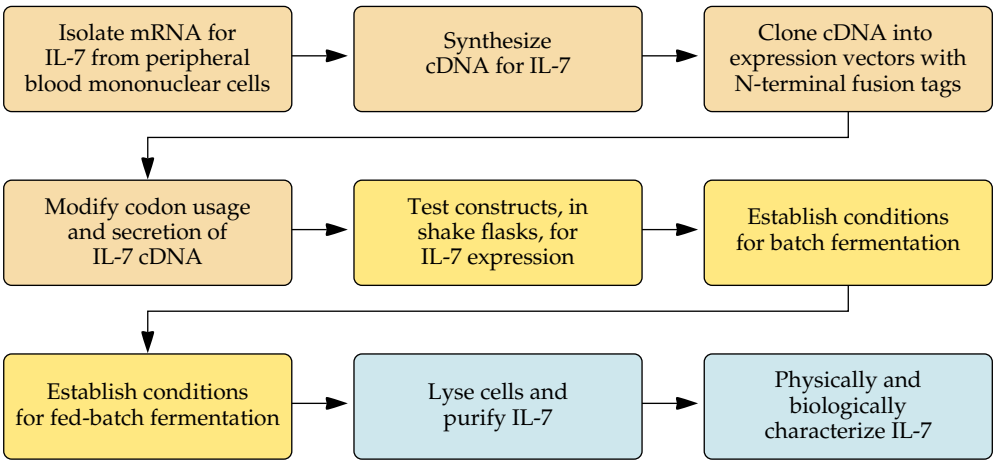
target protein. In a 75-liter fermenter, using lactose as the sole feed was not efficient for cell growth, presumably because the host strain of *E. coli* was unable to metabolize galactose. However, with a 1:50 ratio of lactose to glycerol, the target protein reached 16% of the total cellular protein. It remains to be seen whether continuous lactose induction will benefit the expression of proteins other than human apolipoprotein(a) and to what extent this process can be scaled up.

Human interleukin-7 stimulates the differentiation of hematopoietic stem cells into lymphoid progenitor cells. It is also important for B and T cell development. This cytokine protein has been used as an immunotherapy

agent in the treatment of various types of cancer and HIV infections and in certain types of stem cell transplants. Recombinant interleukin-7 has been produced in *E. coli*, the yeast *Pichia pastoris*, and baculovirus, with expression in each of these systems having significant limitations. To significantly improve the production of interleukin-7 in *E. coli*, the preferred host for recombinant protein expression, a number of both molecular and fermentation steps needed to be optimized. *E. coli* was chosen as the expression system because of its rapid generation time, well-characterized genetics, and the possibility of high-cell-density fermentation strategies. The initial steps of this process (Fig. 9.13) include isolating mRNA for interleukin-7, using this mRNA as a template for the synthesis of interleukin-7 cDNA, cloning the cDNA into standard vectors, and expressing the cDNA in *E. coli* cells (see chapter 3). Upon observing that the level of cDNA expression in *E. coli* was relatively low, the codon usage of this cDNA was changed to better reflect the codon usage in the *E. coli* host cells. In addition, to prevent interleukin-7 from being synthesized as an insoluble inclusion body, the N terminus of the cDNA was fused with the gene encoding the small ubiquitin-like modifier protein, with the result that >90% of the protein was obtained in a soluble form. Next, the fermentation of *E. coli* cells carrying the modified interleukin cDNA was optimized, first in batch and then in fed-batch culture. However, to prevent plasmid instability, it was necessary to include high levels of selective antibiotic. Finally, the recombinant *E. coli* cells were recovered, and the protein was purified in high yield to near homogeneity (>90% pure), where it showed similar properties and biological activity compared to IL-7 prepared by other methods. While it is likely that this system can be optimized even further, this work nevertheless provides a useful overview of the many steps involved in producing human proteins on a commercial scale.

Porcine parvovirus is the cause of a global disease of swine resulting in embryonic and fetal infection and death. At present, there is no treatment for infected animals; however, both inactivated and modified live virus vaccines have been developed, with all of the licensed vaccines in the United States being inactivated. Porcine parvovirus is a nonenveloped

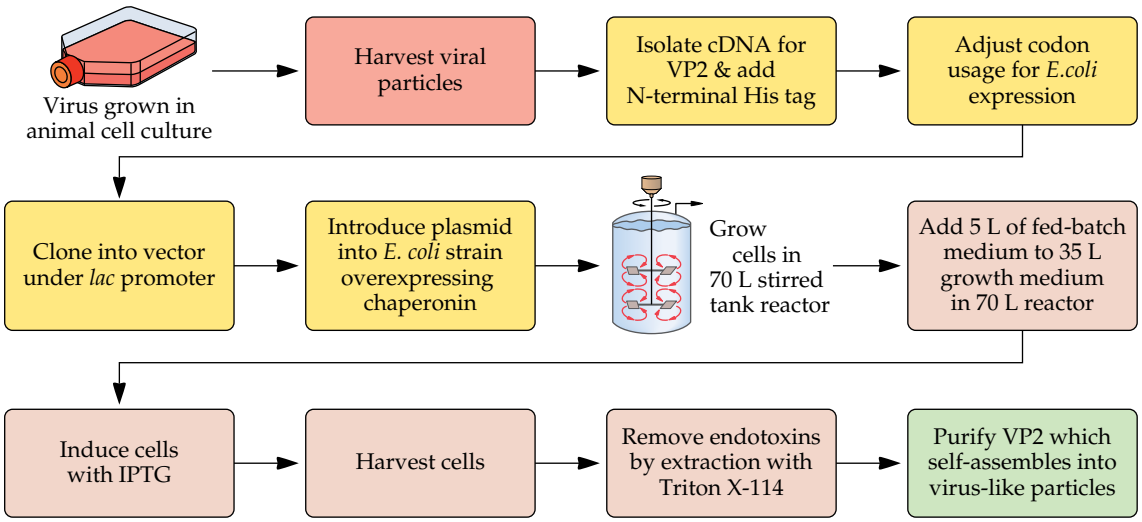
Figure 9.13 Flow chart showing the steps involved in the development of an industrial process for the efficient large-scale production of human interleukin-7 (IL-7).



virus with a single-stranded DNA genome. The viral capsid includes two structural proteins, viral proteins 1 and 2 (VP1 and VP2), with VP2 being the major structural component and having the ability to self-assemble into virus-like particles *in vitro*. These virus-like particles are responsible for eliciting neutralizing antibodies in immunized animals. Human papillomavirus (see chapter 7) can also self-assemble into immunogenic virus-like particles. Recently, a group of scientists described isolating the gene for VP2 from porcine parvovirus, optimizing its expression in *E. coli*, growing the recombinant *E. coli* in 35-liter batches (by fed-batch fermentation), inducing the expression of VP2, removing any endotoxins present, and finally, purifying VP2 so that it could be used as a vaccine (Fig. 9.14). Initially, the cDNA encoding VP2 was isolated and then fused with an N-terminal histidine tag that could later be used to facilitate this protein's purification. To increase the level of expression of this eukaryotic protein in *E. coli*, the codon usage of the cDNA was altered to avoid the use of any rare *E. coli* codons. Then the entire construct was inserted into an *E. coli* expression vector under the transcriptional control of the *E. coli lac* promoter. Next, the engineered plasmid was introduced into an *E. coli* strain that contained an overexpressed chaperonin protein; this ensures that the VP2 protein will fold correctly and not form an inclusion body. The transformed *E. coli* strain was grown to late log phase in 35 liters of medium in a 70-liter stirred tank reactor with the addition of 4 aliquots of concentrated growth medium (5 liters in total) so that the reaction was performed as a fed-batch fermentation. IPTG was added to induce the expression of the VP2 protein. After approximately 1.5 additional hours in the fermenter, the cells were harvested and extracted with detergent (Triton X-114) to remove any endotoxins before the VP2 protein was purified. Preliminary tests have indicated that this protein self-assembles into a virus-like particle that induces neutralizing antibodies in pigs as expected.

Fed-batch fermentation strategies have also been successfully employed in the production of nonprotein products, such as poly(3-hydroxybutyrate) and poly(3-hydroxybutyrate-co-3-hydroxyvalerate), biopolymers with

Figure 9.14 Flow chart showing the major steps involved in the development of an industrial process for the large-scale production of a vaccine against porcine parvovirus.



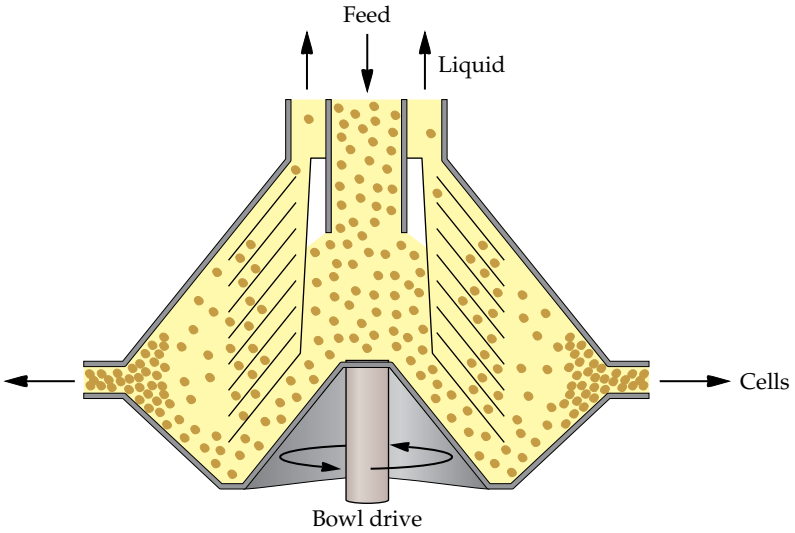
plastic-like properties. In this case, the *E. coli* host cells carried a plasmid that contained polyhydroxyalkanoate biosynthesis genes from the bacterium *Alcaligenes latus*. By using an inexpensive growth medium, such as whey (a waste by-product of cheese-making that consists mainly of lactose), to produce these polymers, it is hoped that a commercially viable product can be produced on a large scale.

Harvesting Microbial Cells

The first step in the process of purification of a product that is synthesized during microbial fermentation is the separation of the cells from the culture medium. Recombinant and native microbial cells can both be harvested with the same type of equipment. However, as a consequence of physiological changes, such as alterations in cell size and the production of extracellular polysaccharides, conditions that have been established for nontransformed cells may not be optimal for recombinant cells expressing a relatively large amount of a foreign protein.

For large volumes, either high-speed centrifugation, which is the current method of choice, or membrane microfiltration is used to separate cells from the growth medium. High-speed semicontinuous centrifuges (Fig. 9.15) have been specifically designed for the harvesting of microbial cells. The cell suspension may be continuously fed into a running centrifuge, and the cells are concentrated within it while the clarified medium is collected in an external container. When the centrifuge chamber is full of packed cells, the run is stopped and the cells are removed. The need to stop and then restart the procedure periodically, especially when large volumes are being processed, can be a major inconvenience. Alternatively, some high-speed centrifuges have been designed to run continuously with the cells being collected outside the centrifuge chamber. Unfortunately, the cost of both the equipment and the power to run it, the release of the microorganisms into the air as aerosols during the harvesting procedure, and

Figure 9.15 Schematic representation of a high-speed semicontinuous centrifuge that has been specifically designed for the harvesting of microbial cells.



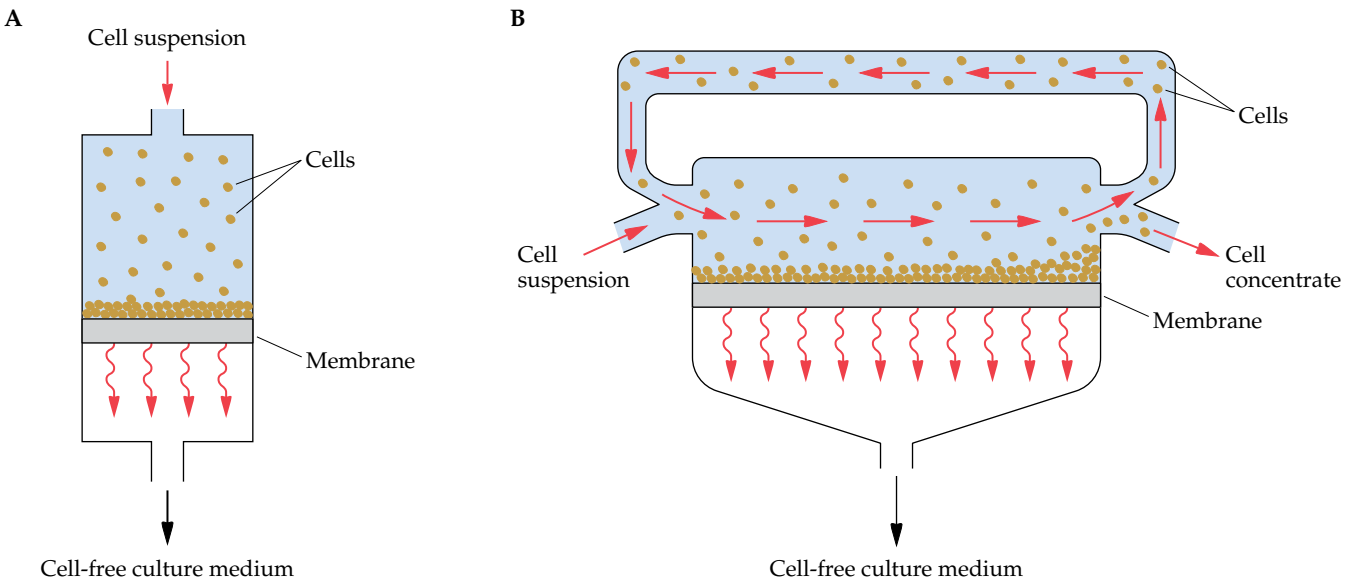
the difficulty in removing all the microbial cells from the spent medium all limit the use of this separation procedure.

Membrane filtration is an alternative method of separating the cells from the culture medium. Unfortunately, with traditional (dead-end) filtration, the microbial cells accumulate on the surface of the polymeric membrane filter. Consequently, the flow rate of spent medium through the membrane decreases rapidly over time (Fig. 9.16A). Increasing the pressure on the membrane enhances the flow for a short time; however, the cells still accumulate on the surface of the membrane and may eventually form a more compact and less permeable layer as a result of the pressure.

An alternative filtration technique entails passing the cell suspension at a high speed across the surface of the membrane (cross-flow or tangential-flow filtration) (Fig. 9.16B). Under these conditions, only a very small fraction of the circulating liquid actually goes through the membrane in any one pass. The remaining cell suspension acts to sweep the membrane clean of accumulated cells, so that the rate of liquid flow through the membrane does not decrease as rapidly as it does in dead-end filtration. In both dead-end and cross-flow filtration, the average pore size of the membranes is 0.2 to 0.45 μm. After many cycles in a cross-flow filtration system, almost all of the culture medium will have passed through the membrane. The use of cross-flow filtration is currently primarily used in laboratory-scale operations. Most industrial-scale operations still rely on continuous centrifugation.

At present, to optimize the production of a recombinant protein, each step in the process, including recombinant DNA, fermentation, cell harvesting, cell disruption, and downstream processing, is optimized separately. Recently, scientists have attempted, using both experimental and computational approaches, to optimize the overall process. However, this approach is still at a very early stage of development,

Figure 9.16 Membrane filtration systems for concentrating microbial cells. **(A)** Dead-end filtration; **(B)** cross-flow filtration. The arrows within each unit show the direction of the liquid flow.



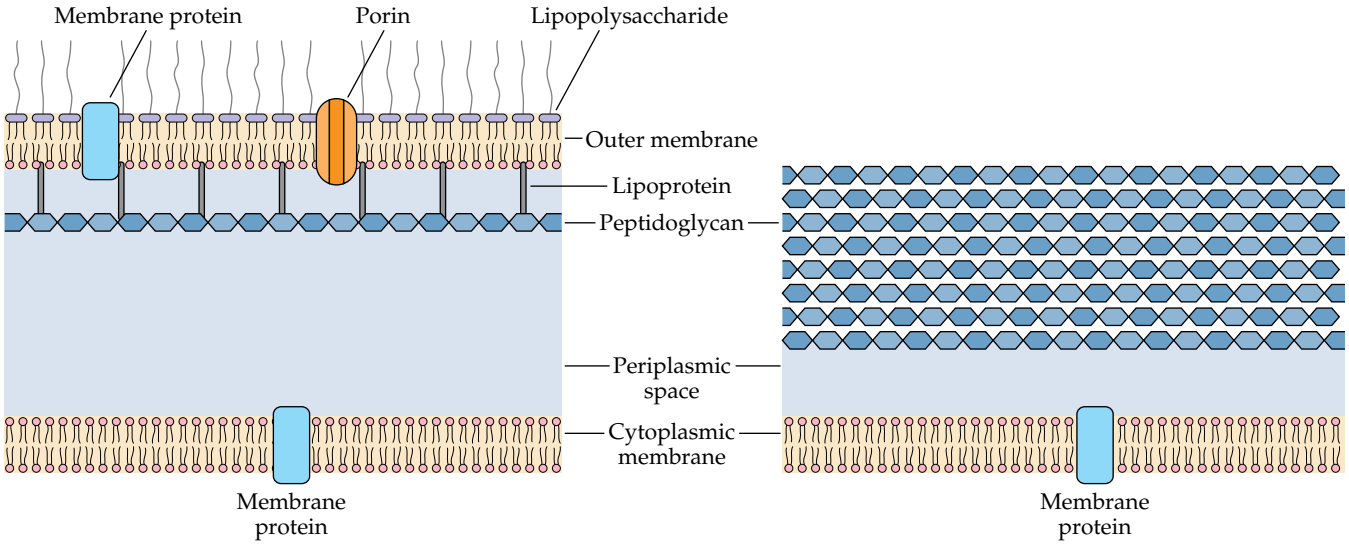
Disrupting Microbial Cells

A large number of chemical, biological, and physical methods have been developed for disrupting microbial cells. All of these procedures represent a compromise, because they must be vigorous enough to break the microbial cell walls yet gentle enough to ensure that the protein product is not denatured. There is no single set of conditions for cell wall lysis because the cell walls of diverse microbial species are composed of different polymers (Fig. 9.17). For example, in Gram-positive bacteria, the cell wall is external to the cytoplasmic membrane and consists of a thick peptidoglycan layer of *N*-acetylglucosamine and *N*-acetylmuraminic acid residues cross-linked by oligopeptides. The cell wall of Gram-negative bacteria has an outer membrane, a thin peptidoglycan layer, and a cytoplasmic membrane. The yeast cell wall is composed of a thick layer of partially phosphorylated mannans and β-glucan, which is external to a cell membrane.

Cell wall composition and strength depend on culture conditions, the cellular growth rate, the phase of the growth cycle when the cells are harvested, how the concentrated cells are stored, and whether the isolated microorganism was expressing a cloned gene. All of these factors affect the cells' susceptibility to disruption.

The chemical methods that disrupt microbial cell walls include treatment with alkali, organic solvents, or detergents. If the protein product is stable at pH values from about 10.5 to 12.5, then bacterial cell lysis can easily be carried out on a large scale at low cost. For example, recombinant human growth hormone is efficiently released from *E. coli* by treatment with sodium hydroxide at pH 11. Few, if any, viable cells remain after alkali treatment, which obviates concerns about the inadvertent release of a genetically engineered microorganism from a production facility. Treatment with an organic solvent is a simple and inexpensive way to disrupt cells and has been used for the isolation of enzymes from yeasts. However, preliminary tests must be run to make sure that the proposed treatment does not denature the target protein. Detergents permeabilize

Figure 9.17 Schematic representation of the cell walls and membranes of Gram-negative (on the left) and Gram-positive (on the right) bacterial cells.



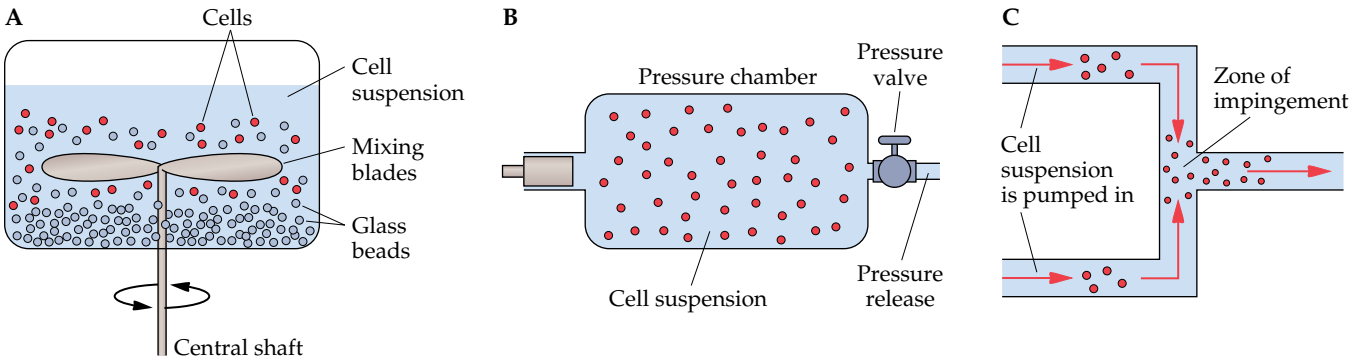
bacterial cells by solubilizing cell membranes and membrane proteins. As a consequence of this activity, holes are formed, and proteins and other molecules are released from the cells. Unfortunately, on a large scale, detergents are relatively expensive, frequently denature the protein product, and are often retained as contaminants throughout the purification process.

The major biological method for disrupting microbial cells is enzymatic lysis. For example, peptidoglycan in the cell walls of Gram-positive bacteria is readily hydrolyzed by the enzyme lysozyme, which is isolated from egg whites; the cell walls of Gram-negative bacteria are hydrolyzed by lysozyme together with the metal-chelating agent ethylenediaminetetraacetic acid (EDTA); and the glucans in the cell walls of yeasts are hydrolyzed by combinations of one or more of the following enzymes: β-1,3-glucanase, β-1,6-glucanase, mannanase, and chitinase. Enzymatic treatments are highly specific, and the conditions for lysis are mild. However, cost considerations currently limit the use of enzymes as cell lysis agents following large-scale fermentations.

Microbial cells can be physically disrupted either by nonmechanical methods, which include osmotic shock and repeated cycles of freezing and thawing, or by mechanical procedures, such as sonication, wet milling, high-pressure homogenization, and impingement. Generally, after treatment by a nonmechanical method, many of the cells remain intact. In contrast, mechanical disruption is highly efficient, which makes it the preferred choice. A sonicator that generates high-pressure sound waves that cause cell disruption by shear and cavitation (production of internal holes) is generally useful for small volumes.

Wet milling is quite commonly used for disrupting large quantities of cells (Fig. 9.18A). A concentrated cell suspension is pumped into the chamber of a high-speed agitator bead mill that is filled with an inert abrasive material, such as small glass beads (<1 mm in diameter), and is fitted with a central shaft that has a number of attached blades. When the device is turned on and the blades are put in motion, most of the cell disruption occurs as a consequence of the shear forces generated by the high-speed motion of the glass beads. Optimized cell disruption depends on both the number and configuration of the agitator disks, the agitator speed, the size of the glass beads, the number of glass beads, the cell concentration, the geometry of the grinding chamber, and the temperature. Bead mills have been successfully

Figure 9.18 Schematic representation of three methods of mechanical cell disruption of microbial cells. **(A)** Wet milling; **(B)** high-pressure homogenization; **(C)** impingement.



used to disrupt a wide range of different kinds of microbial cell types and can readily break recombinant, as well as nonrecombinant, cells.

In the high-pressure homogenization process (Fig. 9.18B), concentrated cells are pumped into a valve assembly under high pressure, and the pressure is then rapidly decreased, causing the cells to lyse. This process can be customized for different microorganisms and protein products by changing the operating pressure, the design of the valve, the temperature of the cell suspension, or the number of times the cell mass is treated.

Impingement (Fig. 9.18C) is a cell disruption procedure in which a high-velocity stream of suspended cells under pressure hits either a stationary surface or a second fluid stream of suspended cells. The forces that are created at the point of contact disrupt the cells. With a device called a microfluidizer, for example, two parallel streams of *E. coli* cells in suspension are directed toward one another. With this device, a high percentage of the cells are disrupted by a single passage through the unit. However, additional passages may be required for complete breakage of other cell types. Unlike high-pressure homogenizers and high-speed agitator bead mills, which generally require highly concentrated cell suspensions, this device can be used to disrupt cells in either dilute or concentrated preparations. The activities of cellular proteins are not significantly impaired by the technique. When the cell suspension is pretreated with low levels of lysozyme and then disrupted in the microfluidizer at much lower than normal pressure and fluid velocity, the activities of labile proteins, which might be otherwise inactivated by the high pressure, are retained.

Downstream Processing

The next step in the purification process depends on both the nature and the location of the product. If the final product is a protein and it is secreted to the culture medium, the medium is concentrated, often by ultrafiltration, and the target protein may be purified by column chromatography or other standard procedures. If the product is a low-molecular-weight compound in the culture medium, it can be purified by the appropriate extraction procedures. Finally, if the product is in the cellular fraction, the cells must be disrupted (as described above) and the cell debris removed by either low-speed, high-capacity centrifugation or membrane microfiltration before the steps leading to product purification are initiated.

It has been estimated that downstream processing represents approximately 45 to 90% of the total cost of manufacturing a recombinant protein. Therefore, efficient and economical purification strategies are key to bringing a recombinant protein to the marketplace.

The protein product is precipitated from the crude lysate, the clarified lysate, or the cell-free culture medium with organic solvents (alcohol or acetone) or ammonium sulfate. Under these conditions not all proteins are precipitated from the crude lysate so that the target protein is usually enriched approximately 2- to 5-fold. Unfortunately, the cost of the precipitant can add significantly to the cost of the process. Alternatively, the crude protein mixtures are concentrated and fractionated by cross-flow ultrafiltration through membranes that have a smaller average pore size than those used for either cell concentration or debris removal (Fig. 9.16B). This approach can be used with volumes ranging from 1 liter

to several thousand liters and can be performed continuously, which means that large-volume systems are not required. Depending on the size and properties of the target protein, this method can yield 10- to 100-fold enrichment.

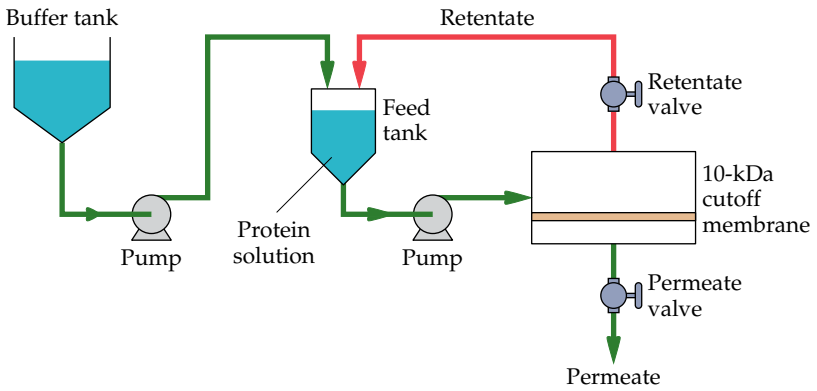
On a large scale, it is impractical to remove small molecules, such as salts or organic solvents, from protein solutions by conventional laboratory procedures. Consequently, the same apparatus that is used to concentrate proteins by ultrafiltration has been developed for the large-scale dialysis of proteins (Fig. 9.19). In addition, by using two different-size membranes in tandem, large volumes of target proteins with different molecular masses may be fractionated by ultrafiltration (Fig. 9.20).

The required degree of purity of the final protein product depends on its end use. In some cases, such as enzymes for use with laundry detergent, crude preparations are satisfactory, but for other products, such as pharmaceutical proteins, additional purification procedures are absolutely necessary.

Inclusion Bodies

A number of proteins that are overproduced intracellularly are confined to insoluble particles (inclusion bodies) within the bacterium. Following cell harvesting and after cell disruption, such inclusion bodies can be readily separated from the bulk of the remaining cell components; however, it is not always straightforward to recover active soluble proteins from inclusion bodies. Initially, researchers found it difficult to solubilize

Figure 9.19 Schematic representation of large-scale ultrafiltration dialysis of a protein solution. The solution from the feed tank is pumped across the ultrafiltration membrane, with only a small fraction of the solution actually passing through the membrane and the rest being used to sweep the membrane clean of protein. The volume that passes through the membrane is matched by an equal volume of buffer added to the system. A membrane with a nominal molecular mass cutoff below the size of the target protein is used. For example, a 10-kDa cutoff membrane might be used to retain a 30-kDa protein while removing salt from the solution. Following dialysis, the target protein is found, in a dilute solution, in the feed tank. This solution may be concentrated by ultrafiltration using an identical setup except that no buffer tank is used. The arrows indicate the direction of liquid flow. The relative volume of permeate (liquid that passes through the membrane) compared with the retentate (retained liquid) is controlled by adjusting the permeate and retentate valves; this helps to keep the membrane relatively free of protein that might otherwise clog its pores.



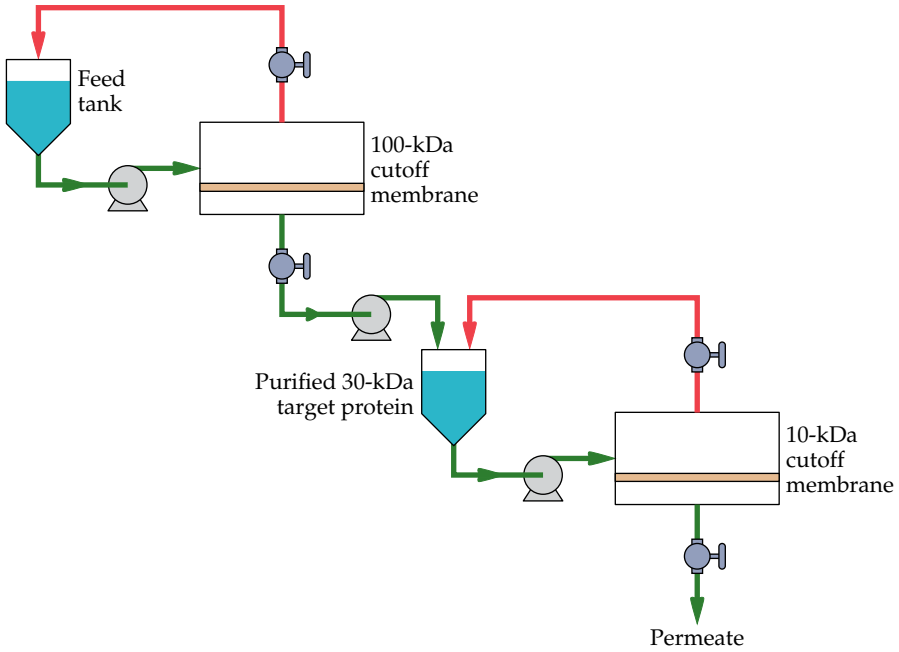


Figure 9.20 Schematic representation of large-scale protein purification using ultrafiltration membranes. In the example shown, two membranes are used sequentially, a 100-kDa cutoff membrane and a 10-kDa cutoff membrane, in order to purify a 30-kDa protein. The solution, which contains a protein mixture, is pumped through the 100-kDa cutoff membrane, with the larger proteins being retained and the smaller proteins, including the 30-kDa target protein, passing through the membrane. This protein solution is next pumped through a 10-kDa cutoff membrane, with the 30-kDa target protein being retained. The arrows indicate the direction of the liquid flow. The symbols are identical to those in Fig. 9.19.

inclusion bodies without irreversibly denaturing the protein, but protocols have now been devised to renature the proteins found in inclusion bodies (Fig. 9.21). These additional steps to isolate, solubilize, and protein-refold the proteins in the inclusion bodies increase the complexity and cost of the purification process.

When producing target proteins on a laboratory scale, it is often possible to fuse the target protein to another peptide or protein, with the result that the overproduced fusion protein is mostly soluble and readily purified using standard procedures. Unfortunately, once the fusion protein has been purified, the fusion partner must be removed from the target protein by either chemical or enzymatic means (see chapter 3), significantly increasing the cost of producing the target protein. In addition, expending a portion of the cell's resources on the production of the fusion partner often decreases the yield of the target protein.

In instances where the target protein is relatively stable, the inclusion body containing it and other proteins may be solubilized following treatment with relatively mild levels of denaturing agents. The drawback of this approach is that under these conditions not all of the target protein becomes solubilized.

Employing high levels of denaturing compounds almost always completely denatures and solubilizes all of the proteins in an inclusion body.

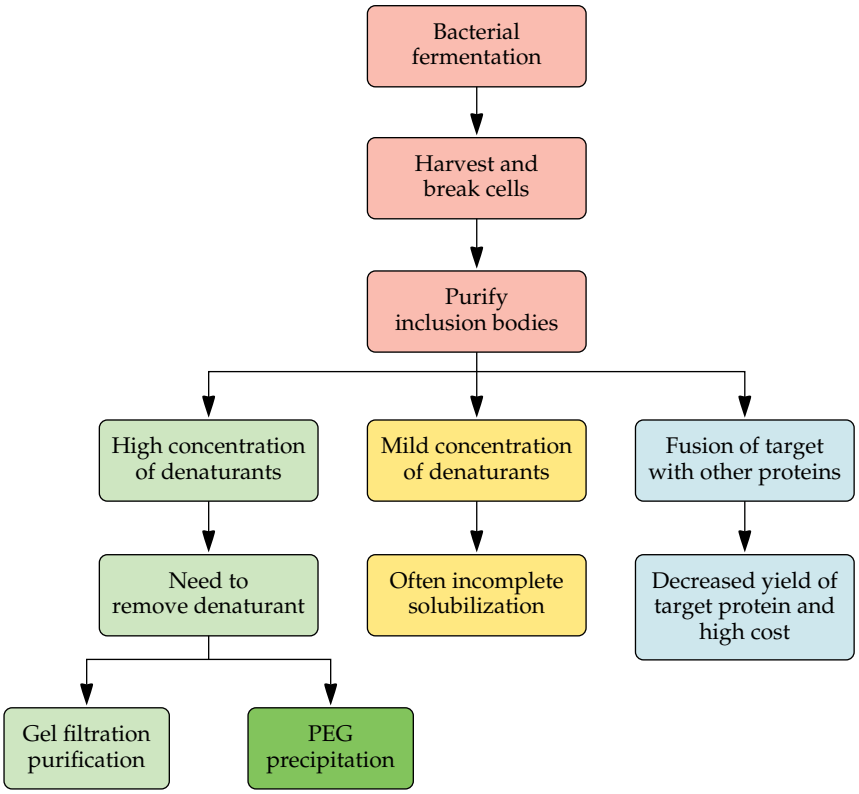


Figure 9.21 Schematic representation of different approaches to recover overexpressed target proteins from bacterial inclusion bodies.

However, in this case the problem becomes removing all of the denaturant so that the target protein may be correctly refolded and purified.

In one recent report scientists expressed a cDNA encoding human interleukin-15 (IL-15) in *E. coli* cells under the transcriptional control of the *lac* promoter. Cells were grown at 37°C to mid- to late log phase before 1 mM IPTG (see chapter 2) was added for 3 hours to induce IL-15 expression. Then, the 450 grams of cells from 15 liters of growth medium was harvested, the cells were disrupted, and the inclusion bodies were purified. Washed inclusion bodies were solubilized in a solution of 7 M guanidine hydrochloride (a strong protein denaturant) and 100 mM dithiothreitol (which reduces disulfide bridges). Next, the solubilized inclusion body proteins were purified by either gel filtration chromatography or polyethylene glycol precipitation. Efficient gel filtration is slow and requires either a very large column or multiple runs on a medium-sized column, making this step quite difficult to scale up. When 10% polyethylene glycol was added to the solubilized inclusion bodies and left at 4°C for 2 hours, the protein pellet was readily recovered. While both polyethylene glycol precipitation and gel filtration removed most of the protein contaminants, polyethylene glycol precipitation took a little less than half as much time and cost about 90% as much as the gel filtration method. Thus, on a relatively large scale, following solubilization of inclusion bodies, polyethylene glycol precipitation was found to be an effective method to help efficiently recover the target protein.

Utilizing an Immobilized Enzyme

In the purification of protein products or the processing of cellular metabolites, it is sometimes useful to employ an immobilized enzyme that either captures or acts upon the protein or metabolite in question. For example, the majority of the human population is unable to digest lactose, and those individuals typically suffer from gastrointestinal distress when lactose is ingested. It has been estimated that 75% or more of individuals whose forbearers are of African, Jewish, Native American, or East Asian descent are lactose intolerant. To overcome this problem, milk may be predigested (hydrolyzed) with the addition of purified β-galactosidase (also called lactase) to glucose and galactose before being consumed or converted into other products such as cheese or sour cream. For greater efficiency, lactose hydrolysis may be achieved by immobilized β-galactosidase. This approach has the potential to reduce processing costs through the repeated utilization of the same enzyme, achieve a high rate of lactose conversion, and avoid inclusion of the enzyme in the lactose-free milk, which obviates the need for further purification. β-Galactosidase is typically immobilized by either covalent bonds between the carrier (often alginate beads) and the enzyme or by the enzyme's absorption onto an insoluble carrier. The immobilization procedure requires the enzyme to be highly purified. In addition, the immobilization process may partially inactivate the enzyme so that the use of a more stable version of β-galactosidase, possibly from a thermophilic microorganism, may be advantageous in facilitating the rapid and efficient hydrolysis of lactose.

In one study, researchers genetically engineered a fusion protein consisting of a thermostable β-galactosidase and a thermostable cellulose-binding domain from a cellulose-degrading protein (Fig. 9.22). Following the growth of *E. coli* cells that expressed the fusion protein, the *E. coli* host cells were gently lysed, and the fusion protein was immobilized onto cellulose fibers. The immobilization was performed in a single step, with no indication that other (unwanted) proteins had bound to the cellulose. The immobilized fusion protein was then used in a columnar continuous-flow system for the hydrolysis of lactose in solution (Fig. 9.22). When a column with immobilized fusion protein was run at 70°C at a flow rate of 0.64 column volumes per minute, approximately 90% of the lactose in cow's milk was hydrolyzed. Moreover, the passage of around 10,000 column volumes through the column did not decrease the efficiency of hydrolysis, and the column could be stored at room temperature for a period of at least 6 months without any loss in enzyme activity. This work is a first step in the development of a continuous large-scale process for the production of lactose-free milk.

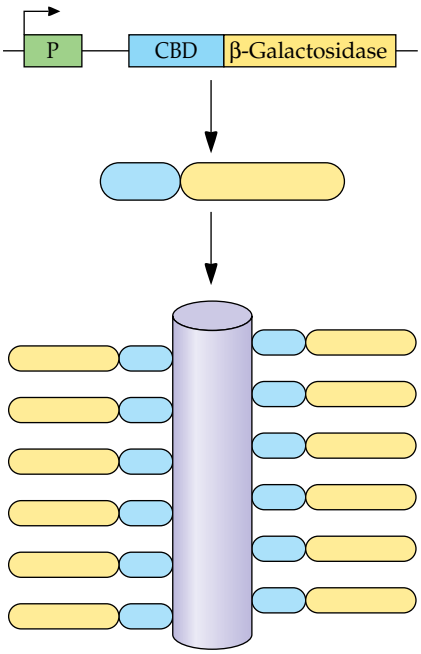


Figure 9.22 Schematic representation of the construction of a fusion gene encoding a thermostable cellulose-binding domain (CBD) and a thermostable β-galactosidase under the transcriptional control of a bacterial promoter (P). The fusion protein is expressed and, following cell lysis, is captured and immobilized on a cellulose fiber through the cellulose-binding domain of the fusion protein.

Magnetic Separation of Proteins

A simple procedure for purifying proteins from crude protein mixtures that may be used on a large scale utilizes metallic particles conjugated to a ligand such as an antibody (Fig. 9.23). The antibody component of the antibody-metallic bead conjugate binds tightly to the targeted protein, and a magnet is then used to capture the bound target proteins via the metal component. Unbound proteins are removed by simple filtration or aspiration, and finally, the bound target proteins are released

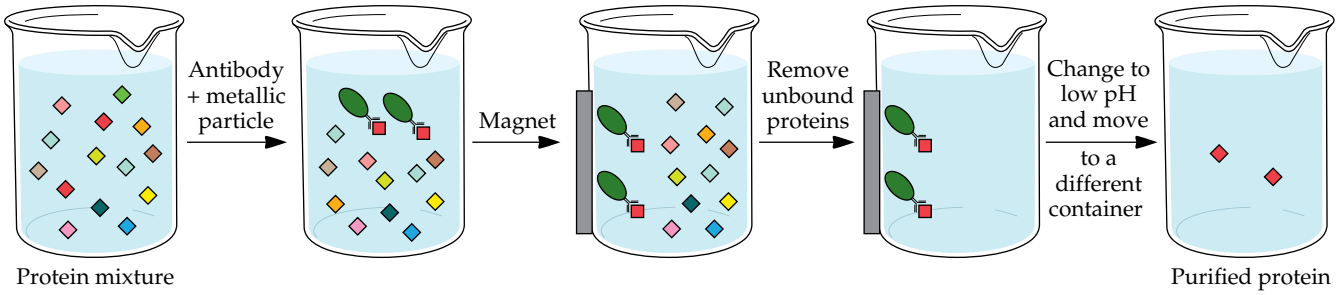


Figure 9.23 Flow chart depicting the purification of a target recombinant protein from a protein mixture using metallic beads coupled to antibodies that bind tightly to the target protein.

from the complex, generally by lowering the pH of the buffer solution. If necessary, the (largely) purified target proteins may then be further purified by conventional techniques such as ion exchange or size exclusion chromatography.

Large-Scale Production of DNA and RNA

Plasmid DNA

Gene therapy and genetic immunization protocols are being tested in a large number of clinical trials. Many of these trials require plasmids as vectors to deliver the remedial DNA to the patient. In fact, of all of the nonviral vectors that are currently being used in clinical trials, plasmid DNA is the most common, accounting for ~18.7% of the total. As these procedures become more routine and are extended to many more patients, it will be necessary to produce plasmid DNA, initially in the 5- to 20-kilobase size range, on a large scale and in a highly purified form, i.e., pharmaceutical grade.

A number of factors, including the choice of a host strain, such as *E. coli* K-12, that is safe and well characterized; growth conditions; defined medium; and a purification process that removes all of the undesired genomic DNA, RNA, proteins, lipids, and lipopolysaccharides, need to be considered for the large-scale production of plasmid DNA. Generally, to ensure that the plasmid DNA will be stable, a host strain should produce only low levels of nucleases. In addition, the presence of antibiotics in the culture medium should be avoided, since it is difficult to remove all traces of an antibiotic from the final preparation. For DNA isolation, the current method of choice for cell breakage is the alkaline-lysis procedure—treatment with 0.2 M NaOH and 1% sodium dodecyl sulfate—which breaks cells without disrupting plasmid DNA. The lysis solution must be added so that mixing is sufficient to lyse all of the cells and local pH extremes are avoided, while the shear that is generated by the mixing does not damage the plasmid DNA. After cell lysis, the precipitates formed, which contain cell debris, denatured proteins, and nucleic acids, must be removed. This is most commonly done by centrifugation using fixed-angle rotors, a process that is difficult to scale up. After removal of the precipitated material, plasmid purification from the remaining solution is best achieved by anion-exchange chromatography, followed by size exclusion (gel filtration) chromatography.

In the industrial production of many recombinant proteins, high-pressure homogenizers are used to continuously disrupt cells. However, the use of such homogenizers typically results in the degradation of a large portion of the plasmid DNA as a consequence of the severe shear stress generated during cell breakage. To work out a process for the large-scale preparation of plasmid DNA, one group of researchers developed and optimized a unique continuous thermal-lysis protocol. In this procedure, cells were harvested after high-cell-density batch culture, filtered to remove most of the growth medium, and resuspended in lysis buffer containing lysozyme. After incubation at 37°C and gentle stirring for 20 minutes, the cells were lysed by heating to 70°C for 20 seconds and then filtered to remove the cell debris (Fig. 9.24). Using this procedure, researchers have reported obtaining 100 mg of high-quality plasmid DNA, free of contaminating chromosomal DNA, per liter of high-cell-density cell culture. The lysozyme is the most expensive component of this process. Nevertheless, as a consequence of the high plasmid yield, the avoidance of time-consuming and expensive centrifugation steps, the short processing time—17 liters of *E. coli* cells can be processed in ~45 minutes—and the ease with which the process can be scaled up, this approach is likely to provide a highly effective means of preparing large amounts of highly purified plasmid DNA.

To increase the level of plasmid production (and the plasmid stability), researchers have mutated one or more genes in host *E. coli* strains (Table 9.5). Some of the genes that have been altered are involved in central carbon metabolism. These genes were altered in an effort to increase the flux of carbon metabolism in the bacterial cells toward nucleotide and amino acid precursor synthesis and to reduce the flux of wasteful by-products such as acetate. Other genes that have been altered encode proteins that are directly involved in plasmid DNA production or DNA replication. Most of

Figure 9.24 Schematic representation of a process for the continuous thermal lysis of *E. coli* cells for the large-scale preparation of purified plasmid DNA.

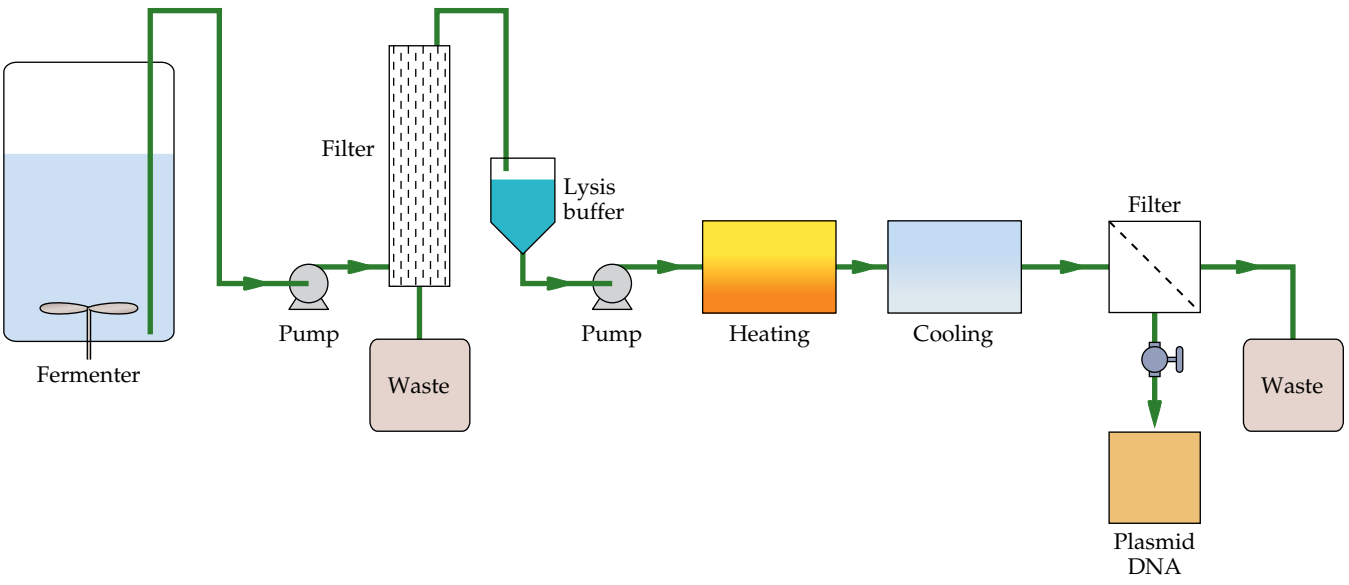


Table 9.5 *E. coli* genes that have been mutated to improve plasmid DNA production

Gene(s)	Activity	Mutation	Phenotype
recA	Protease and nuclease	Deletion	Decreased recombination of cloned DNA
endA	DNA-specific endonuclease	Deletion	Decreased plasmid digestion
rpiA	Ribose-5-phosphate isomerase	Overexpression	Increased synthesis of nucleotide precursors
pykA, *pykF*	Pyruvate kinase	Deletion	Increased pentose phosphate pathway and TCA cycle flux and reduced acetate synthesis
topA	Topoisomerase	Deletion	Increased RNA R-loop formation
polA	DNA polymerase	Overexpression	Increased enzyme that extends the RNA primer template
ligA	DNA ligase	Overexpression	Increased availability of enzyme that seals nicks during plasmid DNA synthesis
gyrAB	DNA gyrase	Overexpression	Increased availability of enzyme that negatively supercoils the covalently closed circular plasmid
priA, *priB*, *priC*	Primosome	Overexpression	Increased availability of primosomes
trxA, *grx1*	Thioredoxin, glutaredoxin	Overexpression	Increased level of hydrogen donors for ribonucleotide reductase
zwf	Glucose-6-phosphate-1-dehydrogenase	Overexpression	Increased pentose phosphate pathway flux
fruR	Transcriptional regulation	Deletion	Deletion of global regulator of carbon flow through the central metabolic pathway
relA, *spoT*	ppGpp synthetase	Deletion	Abolished stringent response to amino acid starvation
dcm	DNA cytosine methyltransferase	Deletion	Improved transgene expression and reduced immunogenicity

these strategies increase to some extent (depending on the bacterial growth conditions) the amount of plasmid DNA that is produced. Often, better results are obtained when several host genes are altered. In addition, other factors such as growth medium composition, carbon source, fermentation strategy (e.g., batch versus fed-batch), and the scale of the fermentation can influence plasmid yield and stability. Thus, it is necessary to elaborate the influence of these process variables in addition to the genetic changes in the host.

Recently, scientists engineered an *E. coli* K-12 host strain in which introduced plasmids were found to be extremely stable. This was achieved by reducing the size of the *E. coli* K-12 genome by approximately 15% by making multiple specific deletions, including nonessential genes, mobile DNA elements, and cryptic virulence genes. Strains carrying the variant genomes that exhibited good growth and protein production were selected. This genome reduction, which was originally undertaken in an effort to define the essential (core) genes in this bacterium, led to the unexpected result that some plasmids that were unstable in wild-type *E. coli* strains were readily and accurately propagated in this multiple deletion strain (Fig. 9.25). This increased plasmid stability in the cells with a reduced genome may reflect the existence of a decreased metabolic load in the altered *E. coli* host cells. Clearly, this strain is a prime candidate for the large-scale propagation of plasmid DNA.

Finally, some plasmid DNA vectors carry inverted-repeat sequences capable of forming secondary "stem-loop" structures that are difficult to replicate. These plasmids are generally unstable in most *E. coli* strains. This can be a serious problem in the large-scale production of plasmid DNA. Typically, these plasmids are produced at a much lower than expected level and/or some of the plasmids that are produced may have their stem-loop regions deleted, thereby changing the plasmid's properties. To remedy this problem, one company offers an *E. coli* host cell with a reduced genome,

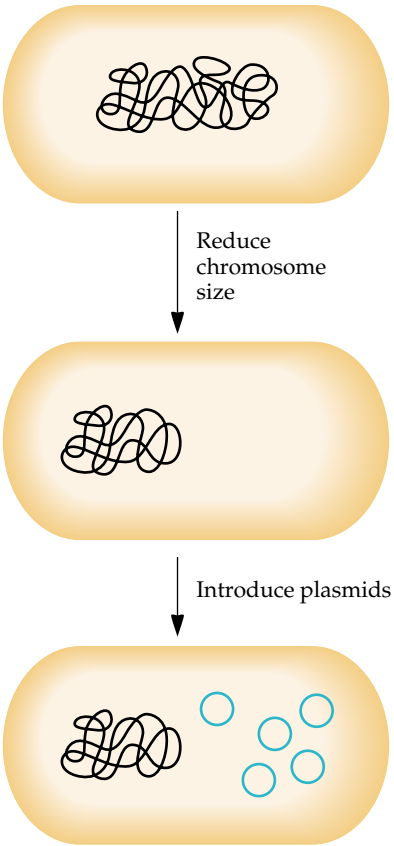


Figure 9.25 Schematic representation of the reduction of the chromosome size of a host *E. coli* cell by approximately 15% and the subsequent stable introduction and maintenance of normally unstable plasmid DNA into the modified *E. coli* strain.

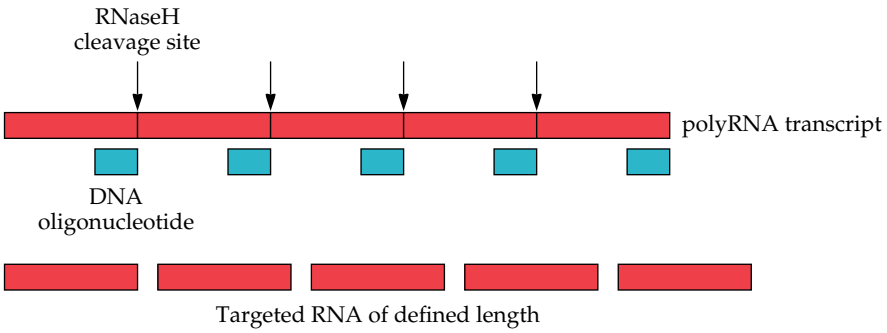
although the changes to the genome are not specified, that may be used in the large-scale production of plasmid DNA that contains stem-loop structures. According to the company's literature, use of their strain of *E. coli* as a host cell for the large-scale production of plasmid DNA results in the successful production of plasmids that contain otherwise unstable stem-loop structures.

mRNA

With the successful development of RNA vaccines directed against SARS-CoV-2, the virus that is the causative agent of COVID-19, and the probability that future vaccines may also be RNA-based, scientists have turned their attention to large-scale methods for the preparation of mRNA. Until recently, scientists produced RNA by transcribing linearized and highly purified DNA using T7 RNA polymerase. With an optimized version of this system, it was possible to produce milligram amounts of defined sequence RNA in as little as 30 minutes. Other scientists have produced large amounts of RNA through solid-phase chemical synthesis, a technique that is limited to relatively short (~100 nucleotides) RNA sequences and is relatively expensive.

While many of the commercial methods for the large-scale purification of RNA remain proprietary, very recently, one group reported a new and simple method to achieve this end (Fig. 9.26). On a plasmid, under the transcriptional control of the T7 promoter, scientists cloned multiple tandem repeats of the target sequence. The resulting transcript was then hybridized with short DNA oligomers. The mixture was then treated with the enzyme RNaseH, which cleaves RNA phosphodiester bonds in an RNA-DNA hybrid molecule at a site that is opposite to the DNA's 5′ end. This leaves the RNA with 3′-OH and 5′-PO_4 ends and enables the production of multiple identical copies of the RNA. These RNA copies may then be purified by ion-exchange chromatography. A United Kingdom patent application was filed for this method in mid-2019.

Figure 9.26 Schematic representation of the synthesis of multiple copies of a target RNA. Multiple copies of the targeted tandem RNA are encoded in plasmid DNA, then synthesized as a polyRNA transcript. Identical DNA oligonucleotides (shown in blue) that bind to the polyRNA are added. RNaseH cleaves the polyRNA (cleavage sites are indicated by arrows) opposite to the 5′ ends of the DNA oligomers, yielding identical RNA molecules, which are readily purified by ion exchange chromatography.



summary

The large-scale production of genetically engineered organisms in industrial-size bioreactors (>1,000 liters) is not achieved merely by extrapolating directly from laboratory growth conditions (0.1 to 1.0 liter). The temperature, pH, rate and nature of mixing, oxygen demand for aerobic organisms, and nutrient levels must be taken into consideration when large bioreactors are being designed.

Microbial fermentations can be performed in several different ways. In batch fermentation, an inoculum of cells is added to fresh medium, and the fermentation is allowed to proceed without supplementation until the maximum amount of the target product is synthesized. Under these conditions, the cell culture passes through six phases of growth: lag, acceleration, log, deceleration, stationary, and death. Protein production is optimal during the log phase, whereas peak production of many low-molecular-weight products occurs during the stationary phase. It is important to monitor batch fermentations closely to ensure that the cells are harvested at the appropriate time. In fed-batch fermentation, growth medium is added at various intervals, usually to prolong the log phase of the fermentation process. Continuous fermentation entails adding fresh growth medium throughout the course of the fermentation and simultaneously removing cells and spent medium.

Each of these fermentation systems has disadvantages and advantages for large-scale production of recombinant products. Although continuous fermentation is still a relatively untried industrial process, the approach has some inherent benefits, making it likely that its use will become more widespread in the future.

One way to increase the amount of a recombinant protein is to grow transformants to as high a cell density as possible. The best way to obtain high-density cell cultures is to use a fed-batch fermentation strategy. Fed-batch fermentation has been directly compared with batch fermentation for the production of several different proteins, and in all cases examined, fed-batch fermentation has resulted in a higher yield of the target protein.

To obtain the largest amount of product in a given volume, it is helpful to attain high cell densities, to avoid the loss of the recombinant plasmid, to utilize *E. coli* cells that can become quiescent during foreign-gene expression, to secrete the foreign protein into the growth medium, and to avoid the formation of acetate in the medium.

There are three basic bioreactor configurations: STRs, bubble columns, and airlift reactors. Currently, STRs are, by far, the most commonly used in industry, but interest in airlift bioreactors is increasing. In STRs, mixing is achieved by mechanical agitation. In airlift reactors, both aeration and mixing are performed by a gas, usually air, that is introduced through a sparger at the bottom of the vessel, with either internal baffles or external loops causing the fluid to circulate within the vessel. Bubble columns are similar to airlift reactors but lack the design features that cause the culture medium to circulate in the vessel. The problems of maintaining sterility, pH, temperature, and other fermentation parameters are overcome in different ways depending on the design of the bioreactor. Two-stage fermentation processes consisting of a growth phase followed by a protein induction phase with either tandem airlift reactors or a single STR have been successfully used for the production of recombinant proteins.

When the product is present within the cells, the cells can be harvested by either centrifugation or cross-flow filtration and lysed chemically, enzymatically, or mechanically. The preferred forms of mechanical cell lysis include wet milling, high-pressure homogenization, and impingement. The product is then separated from the cell lysate. Ultrafiltration has been found to be an effective method for large-scale dialysis, concentration, and initial fractionation of proteins produced by recombinant organisms.

Scientists have begun to establish procedures for the large-scale isolation of plasmid DNA. These procedures must take into account the host cell and its growth and metabolism, plasmid size, cell lysis methods, and the complete removal of a number of potentially contaminating cell components.

REFERENCES

Aristidou AA, San KY, Bennett GN. 1995. Metabolic engineering of *Escherichia coli* to enhance recombinant protein production through acetate reduction. *Biotechnol Prog* **11:**475–478.

Bailey JE, Olis DF. 1977. *Biochemical Engineering Fundamentals*. McGraw-Hill Book Co, New York, NY.

Chen H, Li N, Xie Y, Jiang H, Yang X, Cagliero C, Shi S, Zhu C, Luo H, Chen J, Zhang L, Zhao M, Feng L, Lu H, Zhu J. 2017. Purification of inclusion bodies using PEG precipitation under denaturing conditions to produce recombinant therapeutic proteins from *Escherichia coli*. *Appl Microbiol Biotechnol* **101:**5267–5278.

Choi J-I, Lee SY. 1999. High-level production of poly(3-hydroxybutyrate-co-3-hydroxyvalerate) by fed-batch culture of recombinant *Escherichia coli*. *Appl Environ Microbiol* **65:** 4363–4368.

Datar R. 1986. Economics of primary separation steps in relation to fermentation and genetic engineering. *Process Biochem* **21:**19–26.

Devi N, Adivitiya, Khasa YP. 2016. A combinatorial approach of N-terminus blocking and codon optimization strategies to enhance the soluble expression of recombinant hIL-7 in *E. coli* fed-batch culture. *Appl Microbiol Biotechnol* **100:**9979–9994.

Donovan RS, Robinson CW, Glick BR. 2000. Optimizing the expression of a monoclonal antibody fragment under the transcriptional control of the *Escherichia coli lac* promoter. *Can J Microbiol* **46:**532–541.

Eiteman MA, Altman E. 2006. Overcoming acetate in *Escherichia coli* recombinant protein fermentations. *Trends Biotechnol* **24:**530–536.

Engler CR. 1985. Disruption of microbial cells, p 305–324. *In* Cooney CL, Humphrey AE, Moo-Young M (ed), *Comprehensive Biotechnology*, vol **2**. Pergamon Press, Oxford, United Kingdom.

Feyrer H, Munteanu R, Baronti L, Petzold K. 2020. One-pot production of RNA in high yield and purity through cleaving tandem transcripts. *Molecules* **25:**1142.

Giorgio RJ, Wu JJ. 1986. Design of large scale containment facilities for recombinant DNA fermentations. *Trends Biotechnol* **4:**60–65 http://dx.doi.org/10.1016/0167-7799(86)90256-8.

Glick BR. 1995. Metabolic load and heterologous gene expression. *Biotechnol Adv* **13:**247–261.

Gonçalves GAL, Bower DM, Prazeres DMF, Monteiro GA, Prather KLJ. 2012. Rational engineering of *Escherichia coli* strains for plasmid biopharmaceutical manufacturing. *Biotechnol J* **7:**251–261.

Gosset G, de Anda R, Cruz N, Martínez A, Quintero R, Bolivar F. 1993. Recombinant protein production in cultures of an *Escherichia coli trp-* strain. *Appl Microbiol Biotechnol* **39:**541–546.

Grund G, Robinson CW, Glick BR. 1991. Cross-flow ultrafiltration of proteins, p 69–83. *In* White MD, Reuveny S, Shafferman A (ed), *Biologicals from Recombinant Microorganisms and Animal Cells: Production and Recovery*. Verlag Chemie, Weinheim, Germany.

Hägg P, de Pohl JW, Abdulkarim F, Isaksson LA. 2004. A host/plasmid system that is not dependent on antibiotics and antibiotic resistance genes for stable plasmid maintenance in *Escherichia coli*. *J Biotechnol* **111:**17–30.

Hart RA, Lester PM, Reifsnyder DH, Ogez JR, Builder SE. 1994. Large scale, *in situ* isolation of periplasmic IGF-I from *E. coli*. *Biotechnology (NY)* **12:**1113–1117.

Jungbauer A. 2013. Continuous downstream processing of biopharmaceuticals. *Trends Biotechnol* **31:**479–492.

Khosla C, Curtis JE, DeModena J, Rinas U, Bailey JE. 1990. Expression of intracellular hemoglobin improves protein synthesis in oxygen-limited *Escherichia coli*. *Biotechnology (NY)* **8:**849–853.

Kroner KH. 1986. Cross-flow filtration in the downstream processing of enzymes: current status. *Biotech-Forum* **3:**20–31.

Kroner KH, Nissinen H, Zeigler H. 1987. Improved dynamic filtration of microbial suspensions. *Bio/Technology* **5:**921–926.

Lee SY. 1996. High cell-density culture of *Escherichia coli*. *Trends Biotechnol* **14:**98–105.

Lénon M, Ke N, Szady C, Sakhtah H, Ren G, Manta B, Causey B, Berkmen M. 2020. Improved production of Humira antibody in the genetically engineered *Escherichia coli* SHuffle, by co-expression of human PDI-GPx7 fusions. *Appl Microbiol Biotechnol* **104:**9693–9706.

Levy MS, O'Kennedy RD, Ayazi-Shamlou P, Dunnill P. 2000. Biochemical engineering approaches to the challenges of producing pure plasmid DNA. *Trends Biotechnol* **18:**296–305.

Lim H-K, Jung K-H. 1998. Improvement of heterologous protein productivity by controlling postinduction specific growth rate in recombinant *Escherichia coli* under control of the P^L promoter. *Biotechnol Prog* **14:**548–553.

Lim H-K, Kim S-G, Jung K-H, Seo J-H. 2004. Production of the kringle fragments of human apolipoprotein(a) by continuous lactose induction strategy. *J Biotechnol* **108:**271–278.

Liu M, Feng X, Ding Y, Zhao G, Liu H, Xian M. 2015. Metabolic engineering of *Escherichia coli* to improve recombinant protein production. *Appl Microbiol Biotechnol* **99:**10367–10377.

March JC, Eiteman MA, Altman E. 2002. Expression of an anaplerotic enzyme, pyruvate carboxylase, improves recombinant protein production in *Escherichia coli*. *Appl Environ Microbiol* **68:**5620–5624.

McKillip ER, Giles AS, Levner MH, Hung PP, Hjorth RN. 1991. Bioreactors for large-scale t-PA production. *Bio/Technology* **9:**805–812.

Mendoza-Vega O, Hebert C, Brown SW. 1994. Production of recombinant hirudin by high cell density fed-batch cultivations of a *Saccharomyces cerevisiae* strain: physiological considerations during the bioprocess design. *J Biotechnol* **32:**249–259.

Merchuk JC. 1990. Why use airlift bioreactors? *Trends Biotechnol* **8:**66–71 http://dx.doi.org/10.1016/0167-7799(90)90138-N.

Mukherjee KJ, Rowe DCD, Watkins NA, Summers DK. 2004. Studies of single-chain antibody expression in quiescent *Escherichia coli*. *Appl Environ Microbiol* **70:**3005–3012.

Pósfai G, Plunkett G III, Fehér T, Frisch D, Keil GM, Umenhoffer K, Kolisnychenko V, Stahl B, Sharma SS, de Arruda M, Burland V, Harcum SW, Blattner FR. 2006. Emergent properties of reduced-genome *Escherichia coli*. *Science* **312:**1044–1046.

Packiam KAR, Ramanan RN, Ooi CW, Krishnaswamy L, Tey BT. 2020. Stepwise optimization of recombinant protein production in *Escherichia coli* utilizing computational and experimental approaches. *Appl Microbiol Biotechnol* **104:**3253–3266.

Park TH, Seo J-H, Lim HC. 1991. Two-stage fermentation with bacteriophage λ as an expression vector in *Escherachia coli*. *Biotechnol Bioeng* **37:**297–302.

Paulson DJ, Wilson RL, Spatz DD. 1984. Cross-flow membrane technology and its applications. *Food Technol* (December)**:**77–87.

Prazeres DMF, Ferreira GNM, Monteiro GA, Cooney CL, Cabral JMS. 1999. Large-scale production of pharmaceutical-grade plasmid DNA for gene therapy: problems and bottlenecks. *Trends Biotechnol* **17:**169–174.

Ramírez DM, Bentley WE. 1995. Fed-batch feeding and induction policies that improve foreign protein synthesis and stability by avoiding stress responses. *Biotechnol Bioeng* **47:**596–608.

Reuss M. 1995. Stirred tank bioreactors, p 207–255. *In* Asenjo JA, Merchuk J (ed), *Bioreactor System Design*. Marcel Dekker, Inc, New York, NY.

Riesenberg D, Guthke R. 1999. High-cell-density cultivation of microorganisms. *Appl Microbiol Biotechnol* **51:**422–430.

Robinson DK, Chan CP, Yu Lp C, Tsai PK, Tung J, Seamans TC, Lenny AB, Lee DK, Irwin J, Silberklang M. 1994. Characterization of a recombinant antibody produced in the course of a high yield fed-batch process. *Biotechnol Bioeng* **44:**727–735.

Rowe DCD, Summers DK. 1999. The quiescent-cell expression system for protein synthesis in *Escherichia coli*. *Appl Environ Microbiol* **65:**2710–2715.

Saraswat M, Musante L, Ravidá A, Shortt B, Byrne B, Holthofer H. 2013. Preparative purification of recombinant proteins: current status and future trends. *BioMed Res Int* **2013:**312709.

Sauer T, Robinson CW, Glick BR. 1989. Disruption of native and recombinant *Escherichia coli* in a high-pressure homogenizer. *Biotechnol Bioeng* **33:**1330–1342.

Sayadi S, Nasri M, Berry F, Barbotin JN, Thomas D. 1987. Effect of temperature on the stability of plasmid pTG201 and productivity of *xylE* gene product in recombinant *Escherichia coli*: development of a two-stage chemostat with free and immobilized cells. *J Gen Microbiol* **133:**1901–1908.

Schleef M.1999. Issues for large-scale DNA manufacturing, p 443–470. *In* Rehm HJ, Reed G (ed), *Biotechnology: a Multi-Volume Comprehensive Treatise,* vol. 5a. *Recombinant Proteins, Monoclonal Antibodies, and Therapeutic Genes*. Wiley-VCH, New York, NY.

Schügerl K, Lübbert A. 1995. Pneumatically agitated bioreactors, p 257–303. *In* Asenjo JA, Merchuk J (ed), *Bioreactor System Design*. Marcel Dekker, Inc, New York, NY.

Schütte H, Kula M-R. 1990. Pilot- and process-scale techniques for cell disruption. *Biotechnol Appl Biochem* **12:**599–620.

Siegel R, Ryu DDY. 1985. Kinetic study of instability of recombinant plasmid pPLc23trpAl in *E. coli* using two-stage continuous culture system. *Biotechnol Bioeng* **27:**28–33.

Siegel MH, Hallaille H, Merchuk JC. 1988. Air-lift reactors: design, operation, and applications. *Adv Biotechnol Processes* **7:**79–124.

Singh A, Upadhyay V, Upadhyay AK, Singh SM, Panda AK. 2015. Protein recovery from inclusion bodies of *Escherichia coli* using mild solubilization process. *Microb Cell Fact* **14:**41.

Sletta H, Tøndervik A, Hakvåg S, Aune TE, Nedal A, Aune R, Evensen G, Valla S, Ellingsen TE, Brautaset T. 2007. The presence of N-terminal secretion signal sequences leads to strong stimulation of the total expression levels of three tested medically important proteins during high-cell-density cultivations of *Escherichia coli*. *Appl Environ Microbiol* **73:**906–912.

Stark BC, Pagilla KR, Dikshit KL. 2015. Recent applications of *Vitreoscilla* hemoglobin technology in bioproduct synthesis and bioremediation. *Appl Microbiol Biotechnol* **99:**1627–1636.

Strandberg L, Andersson L, Enfors S-O. 1994. The use of fed batch cultivation for achieving high cell densities in the production of a recombinant protein in *Escherichia coli*. *FEMS Microbiol Rev* **14:**53–56.

Strandberg L, Köhler K, Enfors S-O. 1991. Large-scale fermentation and purification of a recombinant protein from *Escherichia coli*. *Process Biochem* **26:**225–234.

Strathman H. 1985. Membranes and membrane processes in biotechnology. *Trends Biotechnol* **3:**112–118.

Summers DK, Chant E. 2010. Chemical induction in quiescence in bacteria. US patent #7,790,435B2.

Tanny GB, Mirelman D, Pistole T. 1980. Improved filtration technique for concentrating and harvesting bacteria. *Appl Environ Microbiol* **40:**269–273.

Tutunjian RS. 1985. Scale-up considerations for membrane processes. *Bio/Technology* **3:**615–626.

Valdez-Cruz NA, Caspeta L, Pérez NO, Ramírez OT, Trujillo-Roldán MA. 2010. Production of recombinant proteins in *E. coli* by the heat inducible expression system based on the phage lambda pL and/or pR promoters. *Microb Cell Fact* **9:**18.

Velikodvorskaya GA, Tikhonova TV, Gurvits ID, Karyagina AS, Lavrova NV, Sergienko OV, Tashlitskii VN, Lunina NA, Lunin VG. 2010. Chimeric lactase capable of spontaneous and strong immobilization on cellulose and development of a continuous-flow system for lactose hydrolysis at high temperatures. *Appl Environ Microbiol* **76:**8071–8075.

Wang J, Liu Y, Chen Y, Wang A, Wei Q, Liu D, Zhang G. 2020. Large-scale manufacture of VP2 VLP vaccine against porcine parvovirus in *Escherichia coli* with high-density fermentation. *Appl Microbiol Biotechnol* **104:**3847–3857.

Wang ZW, Chen Y, Chao Y-P. 2006. Enhancement of recombinant protein production in *Escherichia coli* by coproduction of aspartase. *J Biotechnol* **124:**403–411.

White MD, Glick BR, Robinson CW. 1995. Bacterial, yeast and fungal cultures: the effect of microorganism type and culture characteristics on bioreactor design and operation, p 47–87. *In* Asenjo JA, Merchuk J (ed), *Bioreactor System Design*. Marcel Dekker, Inc, New York, NY.

Whitney GK, Glick BR, Robinson CW. 1989. Induction of T4 DNA ligase in a recombinant strain of *Escherichia coli*. *Biotechnol Bioeng* **33:**991–998.

Yamane T. 1995. Bioreactor operation modes, p 479–509. *In* Asenjo JA, Merchuk J (ed), *Bioreactor System Design*. Marcel Dekker, Inc, New York, NY.

Zhu K, Jin H, Ma Y, Ren Z, Xiao C, He Z, Zhang F, Zhu Q, Wang B. 2005. A continuous thermal lysis procedure for the large-scale preparation of plasmid DNA. *J Biotechnol* **118:**257–264.

review questions

1. What are the differences between batch, fed-batch, and continuous fermentations?
2. How can fed-batch fermentation be used to improve the production of the insulin B peptide, IFN-γ, Fab fragment, a viral protein antigen, or interleukin-7?
3. What parameters must be monitored and controlled in an optimized fermentation process?
4. What is the effect of a recombinant plasmid on the growth of microbial cells?
5. How does the mixing of a growing microbial culture affect the transfer of oxygen from the growth medium to the cells?
6. What are the advantages of a high cell density during a large-scale fermentation? What conditions lead to a high cell density during a large-scale fermentation?
7. What strategies can be employed to prevent acetate inhibition of the growth of recombinant *E. coli* strains?
8. What are the relative advantages and disadvantages of using an STR or an airlift fermenter?
9. Compare the growth and induction of a recombinant microbial culture using (i) two reactors in tandem and (ii) a single reactor.
10. What is downstream processing?
11. What strategy would you use to purify a recombinant protein that is secreted into the growth medium?
12. What are the advantages and disadvantages of large-scale mechanical lysis of cells compared to chemical lysis?
13. How are cells mechanically disrupted using (i) wet milling, (ii) high-pressure homogenization, and (iii) impingement?
14. How are microbial cells concentrated after the fermentation stage of a biotechnological process? What are the advantages and disadvantages of these procedures?
15. How are microbial cells harvested by continuous centrifugation?
16. How are large volumes of protein solutions partially purified using ultrafiltration?
17. How can the activity of an insoluble recombinant protein that is part of an inclusion body be recovered?
18. What factors should be considered for the large-scale isolation of plasmid DNA?
19. How can "quiescent *E. coli* cells" be engineered to produce large amounts of foreign protein?
20. What strategy would you use to ensure that a plasmid encoding a target protein is not lost during the large-scale growth of a recombinant bacterium?
21. What are some of the advantages of secreting a recombinant protein into the growth medium?
22. How can antibodies that are coupled to metallic particles be used to help purify recombinant target proteins?
23. How can a β-galactosidase and cellulose-binding domain fusion protein be used to develop a continuous process for the removal of lactose from milk?
24. How can a host *E. coli* bacterium be genetically engineered to increase the recoverable yield of plasmid DNA?

10 Genetic Engineering of Plants: Methodology

Plant Transformation with the Ti Plasmid of *A. tumefaciens*

Ti Plasmid-Derived Vector Systems

Increasing Transformation Efficiency

Microprojectile Bombardment

Chloroplast Engineering

Very-High-Level Protein Expression

Use of Reporter Genes in Transformed Plant Cells

Manipulation of Gene Expression in Plants

Transient Gene Expression

Plant Promoters

Manipulation of Genes in Plants

Facilitating Protein Purification

Protein Glycosylation

Gene Stacking

CRISPR-Based Directed Evolution

Polycistronic Gene Expression

Production of Marker-Free Transgenic Plants

Removing Marker Genes from Nuclear DNA

Removing Marker Genes from Chloroplast DNA

SUMMARY

REFERENCES

REVIEW QUESTIONS

There are three major reasons for developing transgenic plants. First, the addition of a gene(s) often improves the agricultural, horticultural, or ornamental value of a target plant. Second, transgenic plants can act as living bioreactors for the inexpensive production of economically important proteins or metabolites. Third, plant genetic transformation provides a powerful means for studying the actions of genes during development and other biological processes.

Some of the genetically determined traits that can be introduced into plants by a single gene or a small number of genes are insecticidal activity, protection against viral infection, resistance to herbicides, protection against pathogenic fungi and bacteria, delay of senescence, tolerance of environmental stresses, altered flower pigmentation, improved nutritional quality of seed proteins, and increased postharvest shelf life. In addition, transgenic plants can be made to produce a variety of useful compounds. For example, on a laboratory scale, plants have been used to produce therapeutic agents (Table 10.1), monoclonal antibodies and antibody fragments (Table 10.2), and vaccine antigens (Table 10.3). In 2015, a monoclonal antibody directed against human immunodeficiency virus (HIV) and produced using genetically modified tobacco plants successfully completed phase I clinical trials in the United Kingdom. The purified monoclonal antibody is produced in a greenhouse in Germany in the first facility in Europe to be granted a license to manufacture recombinant pharmaceutical products from transgenic plants. It has been estimated that this monoclonal antibody produced in plants is approximately

Molecular Biotechnology: Principles and Applications of Recombinant DNA, Sixth Edition. Bernard R. Glick and Cheryl L. Patten.

© 2022 American Society for Microbiology.

Table 10.1 Some pharmaceutical proteins that have been produced in plants

Pharmaceutical protein	Plant(s)	Application(s)
α-Tricosanthin	Tobacco	HIV therapy
Allergen-specific T-cell epitope	Rice	Pollinosis
Angiotensin-1-converting enzyme	Tobacco, tomato	Hypertension
Cyanovirin-N	Tobacco	HIV microbicide
Glucocerebrosidase	Tobacco	Gaucher disease
Human α1-antitrypsin	Rice, tomato	Cystic fibrosis, liver disease, hemorrhage
Human apolipoprotein	Safflower	Plaque reduction
Human aprotinin	Corn	Trypsin inhibitor for transplantation surgery
Human enkephalins	*Arabidopsis*, canola	Antihyperanalgesic by opiate activity
Human epidermal growth factor	Tobacco	Wound repair, control of cell proliferation
Human erythropoietin	Tobacco	Anemia
Human granulocyte-macrophage colony-stimulating factor	Tobacco	Neutropenia
Human growth hormone	Tobacco	Dwarfism, wound healing
Human hemoglobin	Tobacco	Blood substitute
Human hirudin	Canola, tobacco	Thrombin inhibitor, anticoagulant
Human homotrimeric collagen I	Tobacco	Collagen synthesis
Human insulin	Potato, *Arabidopsis*, safflower	Diabetes
Human α interferon	Rice, turnip	Hepatitis C and B
Human β interferon	Tobacco	Antiviral
Human interleukin-2 and interleukin-4	Tobacco	Immunotherapy
Human lactoferrin	Potato, rice	Antimicrobial, diarrhea
Human muscarinic cholinergic receptors	Tobacco	Central and peripheral nervous systems
Human placental alkaline phosphatase	Tobacco	Achondroplasia or cretinism in children
Human protein C	Tobacco	Anticoagulant
Human serum albumin	Tobacco	Liver cirrhosis, burns, surgery
Human somatotropin	Tobacco	Growth hormone
Lipase	Corn	Cystic fibrosis

100 times less expensive than when the same antibody is produced in animal cell tissue culture.

To date, over 150 different plant species have been genetically transformed, including many crop and forest species, in over 50 countries worldwide. Plant genetic engineering is having an enormous impact on plant-breeding programs because it significantly decreases the 10 to 15 years that it typically takes to develop a new variety using traditional plant-breeding techniques.

Despite all of the progress that has been made in the development of transgenic plants for a wide variety of purposes, a vocal minority of individuals in North America and a much larger number in Europe still oppose the use of this technology. Nevertheless, with each succeeding

Table 10.2 Examples of potentially therapeutic antibodies and antibody fragments that have been produced in plants

Host plant	Disease or antigen
Tobacco	38C13 mouse B-cell lymphoma
Tobacco	Anthrax
Tobacco	B-cell lymphoma
Tobacco	Breast and colon cancer
Tobacco	Broad-spectrum anticancer
Tobacco	Botulism
Tobacco	CD40 (cell surface protein)
Tobacco	Cell surface protein from mouse B-cell lymphoma
Tobacco	Hepatitis
Soybean	Herpes simplex virus
Tobacco	Human carcinoembryonic antigen
Pea	Human cancer cell surface antigen
Tobacco	Human CD40 cell surface protein
Tobacco	Human creatine kinase
Alfalfa	Human IgG
Tobacco	Rabies
Tobacco	*Salmonella* surface antigen
Tobacco	*Streptococcus mutans* cell surface antigen SA I/II
Tobacco	Substance P (neuropeptide)

Table 10.3 Some potential vaccine antigens that have been expressed in plants[a]

Disease or causative agent	Plant(s) or vector
Hepatitis B	Tobacco, potato, yellow lupin, lettuce
Malaria	Virus
Rabies	Tomato, spinach, virus
Human rhinovirus	Virus
HIV	Virus
E. coli	Tobacco, potato, corn
Norwalk virus	Tobacco, potato, corn
Diabetes	Tobacco, potato, carrot
Foot-and-mouth disease	*Arabidopsis*, alfalfa
Cholera	Potato, rice
Human cytomegalovirus	Tobacco
Dental caries	Tobacco
Respiratory syncytial virus	Tomato
Human papillomavirus	Potato, tobacco
Anthrax	Tobacco
SARS[b]	Tomato, tobacco
Staphylococcus aureus	Cowpea
Measles	Lettuce
Influenza virus	Tobacco
Tuberculosis	*Arabidopsis*
Rotavirus	Alfalfa

[a]Note that in some cases the antigen was cloned into a transient-expression system, such as a plant virus, often facilitating high levels of expression within a period of 1 to 2 weeks.

[b]SARS, severe acute respiratory syndrome.

year since the mid-1990s, the use of transgenic crops has continued to increase both in terms of the number of hectares of transgenic crops planted and in the number of countries using this technology (Box 10.1). Moreover, a 2014 review (by scientists in Italy, a country where there is a high level of public opposition to the use of transgenic crops) of more than 32,000 scientific studies that examined various aspects of the safety of transgenic crops concluded that there was no evidence of any significant hazards associated with the use of transgenic crops (see chapter 13). It is expected that in the not-too-distant future, the majority of agricultural crops worldwide will be transgenic. Finally, an economic

Is the Debate over Genetically Modified Crops Effectively Over?

box 10.1

Rice is the most important crop in the world. In 2004, China, the world's largest producer of rice, announced a ramping-up of efforts to commercialize genetically modified rice. India, the second-largest producer of rice, was quick to follow suit. With the two largest rice producers in the world (with ~40% of the world's population) switching to genetically modified rice, the rest of the world will eventually have little choice but to accept genetically modified foods.

From 1996 to 2018, the global area devoted to transgenic crops increased from 1.7 million to 191.7 million hectares (ha), an increase of more than 113-fold in 22 years. By 2018, more than 20 million farmers worldwide were using this technology, with more than 90% of those coming from developing countries. In 2018, transgenic crops were planted in 26 different countries by 18 million farmers, with the major producers of transgenic crop plants being in the United States, Brazil, Argentina, India, Canada, China, Paraguay, Pakistan, South Africa, Uruguay, and Bolivia. Major transgenic crops worldwide include (in order of the number of hectares planted) soybean, corn, cotton, canola, rice, squash, papaya, alfalfa, wheat, eggplant, potatoes, sugarcane, and apples. The major traits that have been introduced into plants include herbicide tolerance, insect resistance, and virus resistance. It has been estimated that between 1996 and 2016, as a consequence of using transgenic plants, worldwide pesticide spraying decreased by nearly 600 million kg (or about 20%).

In 2002, the Food and Agriculture Organization of the United Nations endorsed the development and use of genetically modified crops. According to researcher Dr. Florence Wambugu of Nairobi, Kenya, "The African continent urgently needs agricultural biotechnology, including transgenic crops, in order to improve food production. Famine provides critics with an opportunity to promote an antibiotech message that only results in millions of people, who urgently need food, starving to death." Wambugu urged the public to recognize the difference in needs between Europe and Africa. Europe, with a population that is under control, has surplus food and does not experience hunger, whereas Africa, in contrast, experiences mass starvation and death. In 2016, Dr. Margaret Karembu, Director, ISAAA AfriCenter, echoed those sentiments stating, "I dream of a world where those privileged with technology abundance would stop obscuring its access to those who need it most, especially in Africa."

Brinjal (eggplant) is grown on nearly 550,000 hectares in India, making it the world's second-largest producer, after China. Unfortunately, brinjal is prone to insect attack so that some farmers have experienced losses of 70 to 90% in some years. In 2009, Indian scientists began field-testing of transgenic brinjal that provided 98 to 100% protection against *Leucinodes orbonalis* (the major insect pest of brinjal in India). When this transgenic plant was approved for widespread use in 2014, it was determined that farmers had a net economic benefit of ~U.S. $400 per acre and at the same time used 77% less insecticide than with nontransgenic brinjal.

In 2019 the federal government of Nigeria approved the commercialization of a transgenic variety of cowpea (an annual herbaceous legume; black-eyed pea) that is resistant to pod borers (insect pest of legumes that can cause losses of 20 to 80% of cowpea yield). This makes Nigeria the first country in the world to commercialize transgenic cowpea. This variety was developed by scientists at Ahmadu Bello University in Nigeria. It is anticipated that this transgenic plant will greatly reduce the number of times that farmers in Nigeria have to spray their cowpea crops with chemical insecticides from the current six to eight times per season.

A recent report of the effect of transgenic cotton (engineered to be insect resistant) in India found that over a 6-year period, cotton yields had increased by 24% and that small farmers engaged in growing these plants realized a 50% gain in profit compared to growing nontransgenic cotton plants.

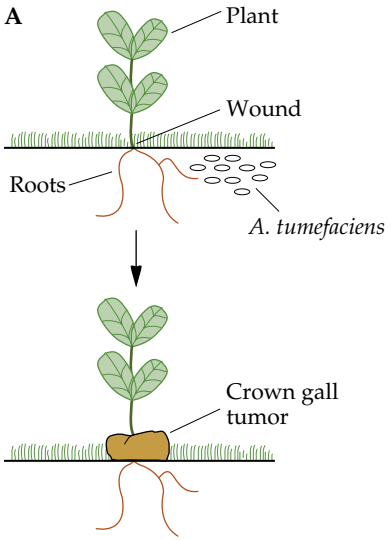
and environmental analysis of the potential impact of a worldwide ban on transgenic crops concludes that the total losses are estimated to be around U.S. $10 billion.

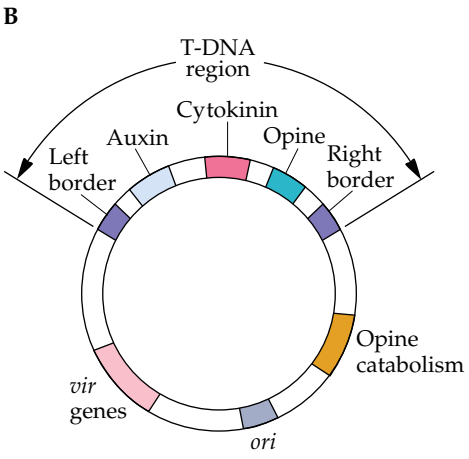
Plant Transformation with the Ti Plasmid of *A. tumefaciens*

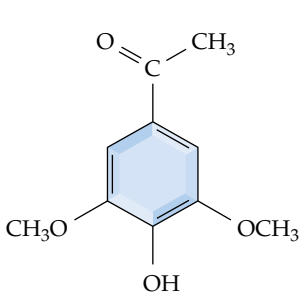
The Gram-negative soil bacterium *Agrobacterium tumefaciens* is a phytopathogen that, as a normal part of its life cycle, genetically transforms plant cells. This genetic transformation leads to the formation of crown gall tumors, which interfere with the normal growth of an infected plant (Fig. 10.1A).

Crown gall formation is the consequence of the transfer, integration, and expression of genes encoded within a specific segment of bacterial plasmid DNA—called the T-DNA (transferred DNA)—into the plant cell genome. The T-DNA is actually part of a much larger "tumor-inducing"

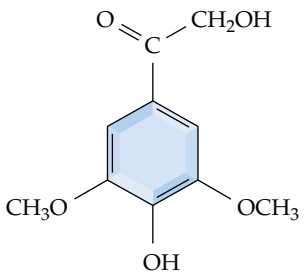
Figure 10.1 **(A)** Infection of a plant with *A. tumefaciens* and formation of a crown gall. **(B)** Schematic representation of a Ti plasmid. The T-DNA is defined by its left and right borders and includes genes for the biosynthesis of auxin, cytokinin, and an opine; these genes are transcribed and translated only in plant cells. Outside the T-DNA region, there is a cluster of *vir* genes, a gene(s) that encodes an enzyme(s) for opine catabolism, and an origin of DNA replication (*ori*) that permits the plasmid to be stably maintained in *A. tumefaciens*. None of these features is drawn to scale. **(C)** Structures of the plant molecules acetosyringone and hydroxyacetosyringone. These compounds are released in response to wounding and can induce the *vir* genes of the Ti plasmid. **(D)** Conserved bases on the right and left borders of the T-DNA of Ti plasmids. N indicates any one of the four nucleotides; i.e., there is no sequence conservation at these positions.







Acetosyringone



Hydroxyacetosyringone

D

Right	5′-TGNCAGGATATATNNNNNNGTNANN-3′
Left	5′-TGGCAGGATATATNNNNNTGTAAAN-3′

Regeneration of Normal and Fertile Plants That Express Octopine Synthase from Tobacco Crown Galls after Deletion of Tumour-Controlling Functions

H. De Greve, J. Leemans, J. P. Hernalsteens, L. Thia-Toong, M. De Beuckeleer, L. Willmitzer, L. Otten, M. Van Montagu, and J. Schell
Nature **300:**752–755, 1982

milestone

An efficient vector system is essential for the routine construction of transgenic plants. Most of the early efforts to develop such a vector focused on the Ti plasmid of the soil bacterium *A. tumefaciens* because, following infection of susceptible dicots, a portion of the Ti plasmid (the T-DNA) is inserted directly into the chromosomal DNA of the host plant cells. However, infection of plants with the native Ti plasmid results in the formation of a crown gall tumor that interferes with the normal growth of the plant. Therefore, before the Ti plasmid could be used as a vector to transform plants, crown gall tumor formation had to be prevented.

By studying the mRNAs that were transcribed from intact and modified T-DNAs, Schell and his colleagues determined that the tumor-inducing genes were a part of the T-DNA. Therefore, a modified T-DNA in which the tumor-inducing genes were deleted was constructed. This modified T-DNA was then introduced by homologous recombination into the Ti plasmid. Such a "disarmed" Ti plasmid could be introduced into plant cells and transfer its T-DNA into their chromosomal DNA. Under these conditions, the modified T-DNA was stably maintained in the plant genome, and importantly, no crown gall tumors were formed. The next logical step in the development of this system was the cloning of foreign marker and target genes into the T-DNA so that they could also be transferred to the chromosomal DNA of the host plant. This vector system utilizing the Ti plasmid has become the system of choice for creating transgenic plants and has been used successfully in thousands of laboratories around the world.

(Ti) plasmid (Fig. 10.1B) that is carried by most strains of *A. tumefaciens*. Depending on the Ti plasmid, the length of the T-DNA region can vary from approximately 10 to 30 kilobase pairs (kb). Strains of *A. tumefaciens* that do not possess a Ti plasmid cannot induce crown gall tumors.

After the initial attachment of the bacterium to a plant cell at the site of an open wound, *A. tumefaciens* responds to certain plant phenolic compounds, such as acetosyringone and hydroxyacetosyringone (Fig. 10.1C), which are excreted by susceptible wounded plants. These wound response compounds resemble some of the products of phenylpropanoid metabolism, which is the major plant pathway for the synthesis of plant secondary metabolites, such as lignins and flavonoids. These small molecules (i.e., acetosyringone and hydroxysyringone) act to induce the virulence (*vir*) genes that are carried on the Ti plasmid (Fig. 10.1B). Initially, the plant phenolic molecules are perceived by the VirA-VirG two-component regulatory system, resulting in the induction of the remainder of the *vir* genes.

The *vir* genes are located on a 35-kb region of the Ti plasmid that lies outside the T-DNA region. There are 25 *vir* genes arranged in 7 operons on the plasmid. The products of the *vir* genes are essential for the transfer and integration of the T-DNA region into the genome of a plant cell. Once the *vir* genes have been induced, the T-DNA is transferred by a process that is similar to plasmid transfer between bacterial cells during conjugation, from donor to recipient cell as a linear single-stranded molecule. Eventually the T-DNA from the Ti plasmid becomes integrated into the plant chromosomal DNA. The formation of a single-stranded form of T-DNA is initiated by strand-specific cutting by VirD1 and VirD2 at the inverted repeat regions in the borders of the T-DNA. Subsequently, VirD2 attaches to the 5′ end of one of the T-DNA strands. Protein VirE2 then coats the remainder of the T-DNA, and the complex exits the bacterium and enters the plant cell. Once inside the plant cell, the T-DNA strand, in a complex with VirD2

and VirE2, enters the nucleus through nuclear pores, where the VirD2 and VirE2 proteins are removed. Finally, the T-DNA strand integrates into the plant chromosomal DNA, where the chromosomal DNA is nicked.

The 5′ end of the single-stranded T-DNA carries the right-border sequence, and the left-border sequence is at the 3′ end. The integration of the T-DNA into the plant genome depends on specific sequences that are located at the right border of the T-DNA. This border contains a repeating unit that consists of 25 base pairs (bp) (Fig. 10.1D). Although the left border contains a similar 25-bp repeat (Fig. 10.1D), this region is not involved in the integration process. Notwithstanding the right and left border DNA sequences, the insertion of the T-DNA into the plant DNA occurs at random sites within the plant genome.

The genes that are located within the T-DNA region of the Ti plasmid are activated only after the T-DNA is inserted into the plant genome. This reflects the fact that these are essentially plant genes, which cannot be expressed in bacteria because of the differences in transcriptional and translational regulatory sequences between the two types of organisms. The products of these genes are responsible for crown gall formation. The T-DNA region includes the genes *iaaM* and *iaaH*, which encode enzymes that synthesize the plant hormone auxin (indoleacetic acid). Specifically, *iaaM* codes for the enzyme tryptophan 2-monooxygenase, which converts tryptophan to indole 3-acetamide, and *iaaH* encodes indole 3-acetamide hydrolase, which converts indole 3-acetamide to indoleacetic acid (Fig. 10.2A). In addition, the T-DNA region carries the *tmr* gene (also known as *ipt*), which encodes isopentenyltransferase. This enzyme adds an isopentenyl side chain to 5′-AMP to form isopentenyladenosine 5′-phosphate, the first committed step in the synthesis of the cytokinin isopentenyladenine (Fig. 10.2B). Hydroxylation of these two molecules by plant enzymes generates the cytokinins called transzeatin and transribosylzeatin, respectively. Both auxin and the cytokinins regulate plant cell growth and development; however, in excess, they can cause the plant to develop tumorous growths, such as crown galls (Fig. 10.1A).

In addition to auxin and cytokinin biosynthesis genes, the T-DNA region from each specific Ti plasmid carries a gene for the synthesis of a molecule called an opine. Opines are unique and unusual condensation products of either an amino acid and a keto acid or an amino acid and a sugar. For example, the condensation product of arginine and pyruvic acid is called octopine, arginine with α-ketoglutaraldehyde is nopaline, and agropine is a bicyclic sugar derivative of glutamic acid (Fig. 10.3). There are ~30 known opines that are synthesized by the transformed plant cells within the crown gall and then secreted. They can be used as a carbon source, and sometimes also as a nitrogen source, by any *A. tumefaciens* cell that carries a Ti plasmid-borne gene for the catabolism of that particular opine (Fig. 10.1B). The opine catabolism gene(s) is on the Ti plasmid but is not part of the T-DNA region. Thus, a unique set of mechanisms has evolved whereby each strain of *A. tumefaciens* carrying a Ti plasmid genetically engineers plant cells to be biological factories for the production of a carbon compound that it alone is able to use.

Ti Plasmid-Derived Vector Systems

The simplest way to exploit the ability of the Ti plasmid to genetically transform plants would be to insert a desired DNA sequence into the T-DNA region and then use the Ti plasmid and *A. tumefaciens* to deliver and insert this gene(s) into the genome of a susceptible plant cell. However,

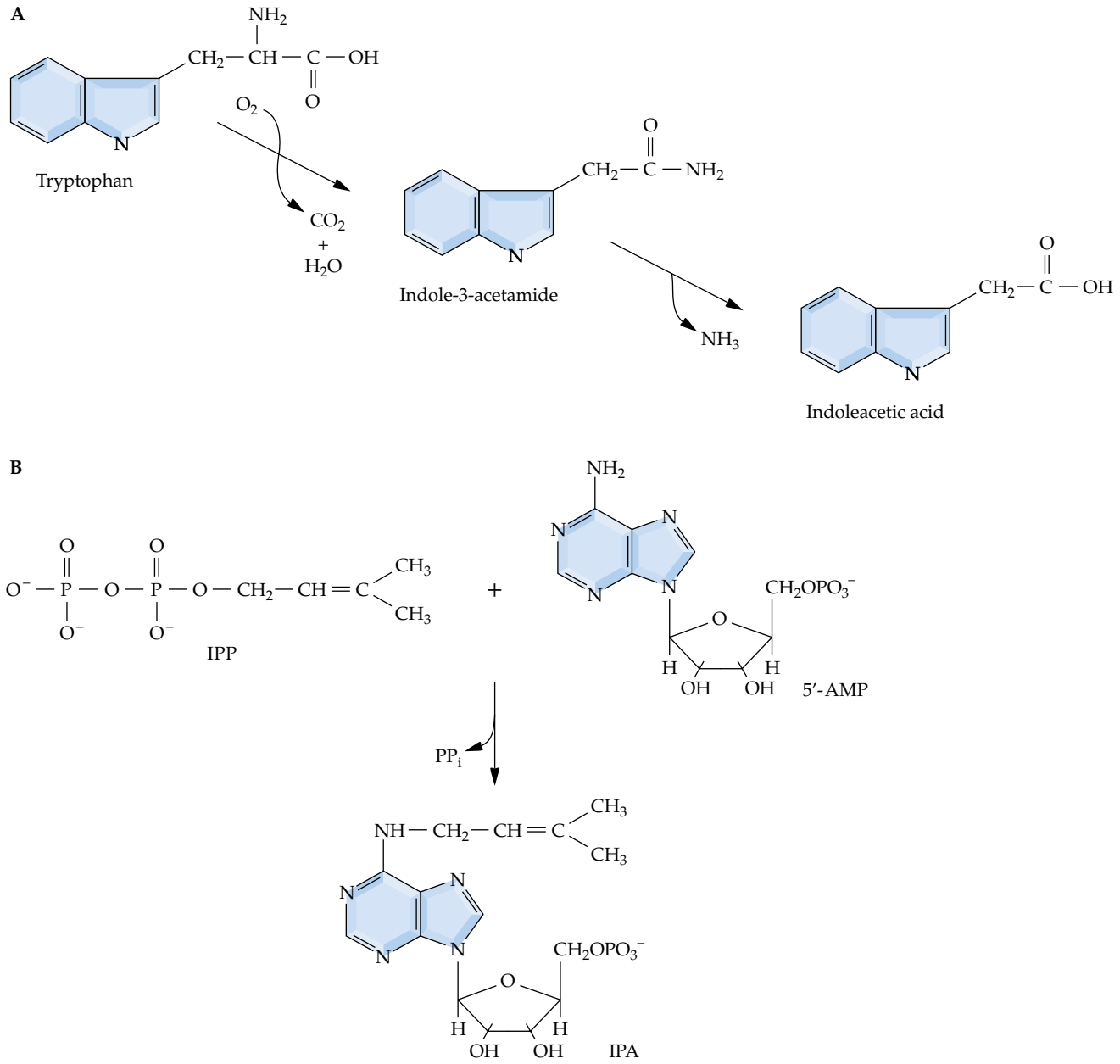


Figure 10.2 Biosynthesis of auxin and cytokinin by the enzymes encoded by the T-DNA genes of the Ti plasmid of *A. tumefaciens*. **(A)** The auxin pathway involves the conversion of tryptophan to indole 3-acetamide by tryptophan monooxygenase and then indole 3-acetamide to indoleacetic acid by indole-3-acetamide hydrolase. **(B)** The cytokinin synthesis reaction entails the attachment of an isopentenyl moiety from isopentenyl diphosphate (IPP) to 5′-AMP by the enzyme isopentenyltransferase to form isopentenyl adenosine monophosphate (IPA).

although the Ti plasmids are effective as natural vectors, they have several serious limitations as routine cloning vectors.

- In addition to being the cause of crown gall tumors, the production of phytohormones by transformed plant cells growing in culture prevents them from being regenerated into mature plants. Therefore, the

auxin and cytokinin biosynthesis genes must be removed from any Ti plasmid-derived cloning vector.

- A gene encoding opine synthesis is not useful to a transgenic plant and may lower the final plant yield by diverting plant resources into opine production. Therefore, the opine synthesis gene needs to be removed.
- Ti plasmids are large (typically ~180 to 210 kb) and do not contain any unique restriction enzyme sites, making it difficult to insert foreign DNA directly into this plasmid.
- Because the Ti plasmid does not replicate in *Escherichia coli*, the convenience of perpetuating and manipulating Ti plasmids carrying inserted DNA sequences in that bacterium is not available.

To overcome these constraints, recombinant DNA technology was used to create a number of Ti plasmid-based vectors. These vectors are similarly organized and contain the following components.

- A selectable marker gene, such as neomycin phosphotransferase, that confers kanamycin resistance on transformed plant cells. Because the neomycin phosphotransferase gene, as well as many of the other marker genes used in plant transformation, is prokaryotic in origin, it is necessary to put it under the control of plant (eukaryotic) transcriptional regulation signals, including both a promoter and a termination–polyadenylation sequence, to ensure that it is efficiently expressed in transformed plant cells.
- An origin of DNA replication that allows the plasmid to replicate in *E. coli*. In some vectors, an origin of replication that functions in *A. tumefaciens* has also been added.
- The right border sequence of the T-DNA region. This region is absolutely required for T-DNA integration into plant cell DNA, although most cloning vectors include both a right and a left border sequence.
- A polylinker (multiple cloning site) to facilitate insertion of the cloned gene into the region between T-DNA border sequences.

Because these cloning vectors lack *vir* genes, they cannot by themselves effect the transfer and integration of the T-DNA region into recipient plant cells. Two different approaches have been used to achieve this end. In one approach, a binary vector system is used (Fig. 10.4A). The binary cloning vector contains either *E. coli* and *A. tumefaciens* origins of DNA replication, i.e., an *E. coli–A. tumefaciens* shuttle vector, or a single broad-host-range origin of DNA replication. In either case, no *vir* genes are present on a binary cloning vector. All the cloning steps to insert the target gene into the T-DNA are carried out in *E. coli* before the vector is introduced into *A. tumefaciens* (often by conjugation between the donor *E. coli* and a recipient *A. tumefaciens*). The *A. tumefaciens* strain carries a modified (disarmed, helper) Ti plasmid that contains a complete set of *vir* genes but lacks portions, or all, of the T-DNA region, so this T-DNA cannot be transferred. With this system, the disarmed Ti plasmid synthesizes the *vir* gene products that mobilize the T-DNA region of the binary cloning vector. By providing the proteins encoded by the *vir* genes, the disarmed Ti plasmid acts as a helper plasmid, enabling the T-DNA from the binary cloning vector to be inserted into the plant chromosomal DNA. Since transfer of the T-DNA is initiated from the right border, the selectable marker, which will eventually be used to detect the presence of the T-DNA inserted into the plant chromosomal

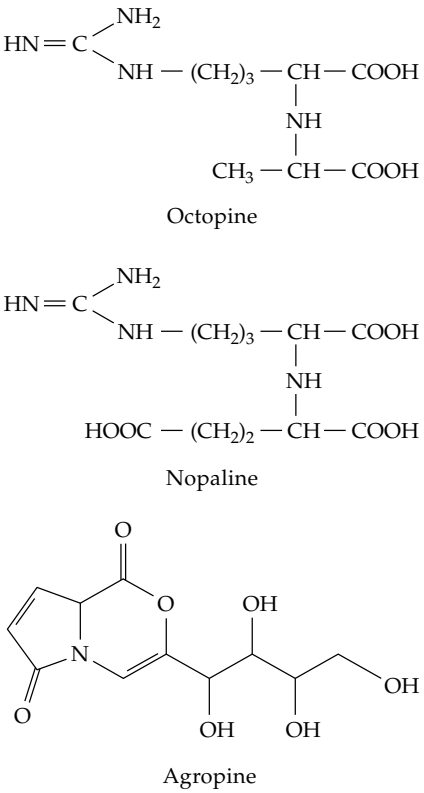


Figure 10.3 Chemical structures of three opines: octopine, nopaline, and agropine.

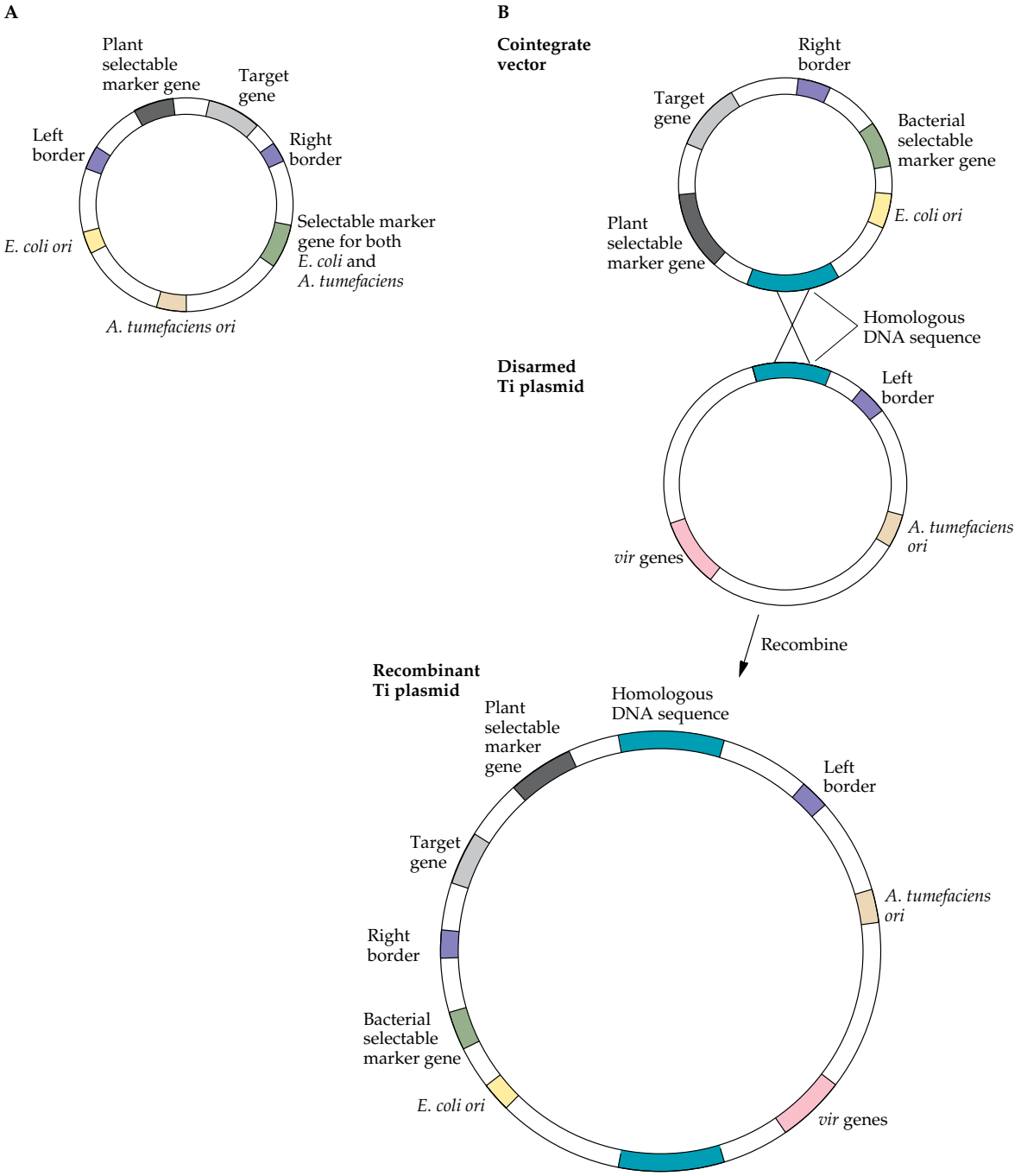


Figure 10.4 Two Ti plasmid-derived cloning vector systems. **(A)** The binary cloning vector has origins of DNA replication (*ori*) for both *E. coli* and *A. tumefaciens* (or a broad-host-range origin), a selectable marker gene that can be used in either *E. coli* or *A. tumefaciens*, and both a target gene and a plant selectable marker gene inserted between the T-DNA left and right borders. **(B)** The cointegrate cloning vector (top) carries only an *E. coli* origin of replication and cannot exist autonomously within *A. tumefaciens*. It also contains a selectable marker that can be used in either *E. coli* or *A. tumefaciens*, a T-DNA right border, a plant selectable marker (reporter) gene, a target gene, and a sequence of Ti plasmid DNA that is homologous to a segment on the disarmed Ti plasmid. The disarmed Ti plasmid (middle) contains the T-DNA left border, the *vir* gene cluster, and an *A. tumefaciens ori*. Following recombination between the cointegrate cloning vector and the disarmed Ti plasmid, the final recombinant plasmid (bottom) has the T-DNA

DNA, is usually placed next to the left border. This ensures that the entire T-DNA has been inserted into plant chromosomal DNA.

In the second approach, called the cointegrate vector system, the cloning (cointegrate) vector has a plant selectable marker gene, the target gene, the right border, an *E. coli* origin of DNA replication, and a bacterial selectable marker gene. Following its transfer by conjugation from *E. coli* to *A. tumefaciens*, the cointegrate vector recombines with a modified (disarmed) Ti plasmid that lacks both the tumor-producing genes and the right border of the T-DNA within *A. tumefaciens*, and the entire cloning vector becomes integrated into the disarmed Ti plasmid to form a recombinant Ti plasmid (Fig. 10.4B). The cointegrate cloning vector and the disarmed helper Ti plasmid both carry identical DNA sequences that provide a shared site for *in vivo* single cross-over homologous recombination; normally, these sequences lie inside the T-DNA region. Following recombination, the cloning vector becomes part of the disarmed Ti plasmid, which provides the *vir* genes necessary for the transfer of the T-DNA to the host plant cells. Since it does not have an origin of replication that is functional in *A. tumefaciens*, the only way that the cointegrate cloning vector can be maintained in *A. tumefaciens* is as part of a cointegrate structure. In this cointegrated configuration the genetically engineered T-DNA region can be transferred to plant cells.

In many instances it may be advantageous to transform plants with several foreign genes, for example, genes that encode an entire biochemical pathway. While this is not yet commonly done, it is nevertheless possible to introduce a large amount of foreign DNA into plants. Although the transformation efficiency is low, plants have been successfully transformed with large DNA fragments ranging from 30 to 150 kb.

Increasing Transformation Efficiency

Although *A. tumefaciens*-mediated gene transfer systems are effective in several species, monocotyledonous plants (monocots), including the world's major cereal crops (rice, wheat, and corn), are not as readily transformed by *A. tumefaciens*. However, by refining and carefully controlling conditions, protocols have been devised for the transformation of corn and rice (and other monocots) by *A. tumefaciens* carrying Ti plasmid vectors. Many of the early plant transformation experiments were conducted with limited-host-range strains of *Agrobacterium*. However, more recently, broad-host-range strains that infect most plants have been tested and found to be effective, so many of the plant species that previously appeared to be refractory to transformation by *A. tumefaciens* can now be transformed. Thus, when setting out to transform a new plant species, it is necessary to determine which *Agrobacterium* strain and Ti plasmid are best suited to that particular plant. In addition, modification of the tissue culture conditions by the inclusion of antioxidants during transformation of several different plants, including grape, rice, corn, and soybean, has been found to increase the transformation frequencies of those plant cells.

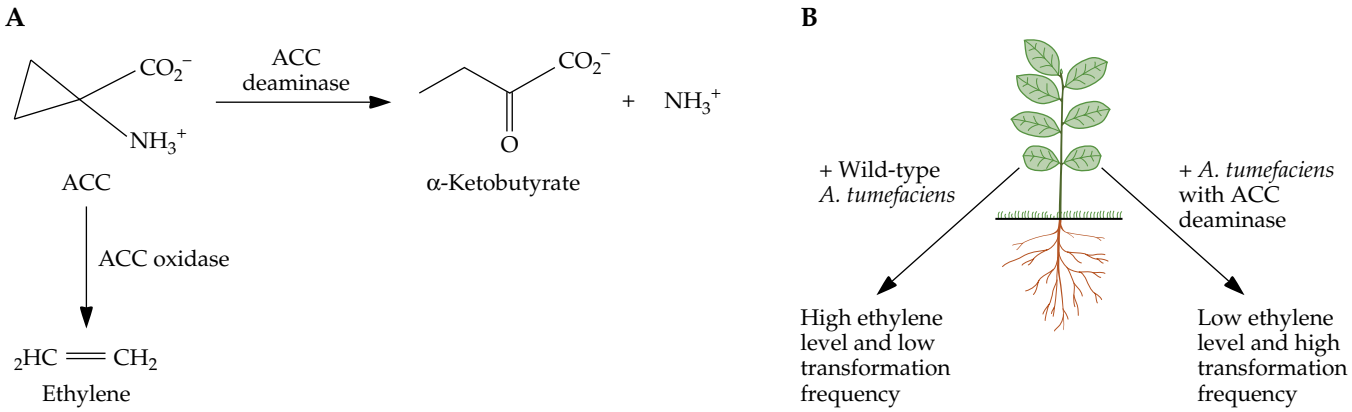
Regeneration of genetically modified plants from cell culture is dependent upon the plant species, its genotype, and the particular tissue being targeted for transformation. In fact, it was observed that some plant genes that are involved in the regulation of plant developmental processes, when overexpressed, were able to facilitate plant regeneration (including in plants and tissues that had previously been deemed to be recalcitrant to regeneration). For example, it was recently observed that some plant genes encoding growth regulating factors (GRFs) positively enhanced

regeneration and transformation in both monocot and dicot plant species. GRFs are part of a small family of highly conserved plant transcription factors that have roles in stem and leaf development, flower and seed formation, root development, and growth under adverse environmental conditions. In sugar beet, expression of exogenous GRF5 in callus culture cells (where callus culture represents a culture of dedifferentiated plant cells) accelerates shoot formation and increases transformation efficiency, thereby enabling stable transformation of otherwise recalcitrant varieties. Similarly, the introduction of GRF5 into canola, soybean, and sunflower increases the proliferation of transformed cells promoting transgenic shoot formation. These results suggest that the overexpression of GRF5 genes makes cells and tissues from a wide variety of crops more competent to regeneration and should significantly improve the transformation of both monocot and dicot plant species.

In another recent experiment, researchers showed that the expression of a fusion protein combining GRF4 and GRF-interacting factor 1 (GIF1; a cofactor of GRF4) significantly increased the efficiency and rate of regeneration in wheat and rice plants, even in the absence of the plant hormone cytokinin (which promotes cell growth and division) added to the medium, thereby facilitating selection of transgenic plants without selectable markers. In addition, the GRF-GIF1 fusion protein improved the regeneration efficiency in citrus crops, indicating that this strategy worked well in dicots as well as in monocots.

A systematic examination of the conditions that are used in *Agrobacterium*-mediated plant transformation revealed that the gaseous plant hormone ethylene significantly decreased the transfer of genes to plant genomes. During plant transformation, ethylene is produced when *Agrobacterium* infects plants. To remedy this, a bacterial gene encoding the enzyme aminocyclopropane-1-carboxylate (ACC) deaminase was introduced into an *A. tumefaciens* strain and, when expressed, lowered plant ethylene levels by cleaving ACC, which is the immediate precursor of ethylene (Fig. 10.5A). This genetically modified *A. tumefaciens* strain was then utilized to introduce foreign DNA into plants. When melon cotyledon

Figure 10.5 **(A)** Cleavage of ACC by bacterial ACC deaminase decreases the amount of ACC that can be converted into ethylene by plant ACC oxidase. **(B)** Improving the transformation of plants by *A. tumefaciens* by engineering the bacterium to produce the enzyme ACC deaminase lowers the amount of ethylene that forms following the interaction of the plant and bacterium.



segments were genetically transformed using an *A. tumefaciens* strain expressing ACC deaminase, the transformation frequency of the plants (as judged by the level of introduced marker enzyme activity) increased significantly (Fig. 10.5B). In addition, another group of researchers found that an *A. tumefaciens* strain that had been genetically modified to express an ACC deaminase gene was more efficient at transforming commercial canola cultivars than a comparable nonengineered strain. Thus, it is expected that the introduction of this ethylene-lowering gene will increase the transformation frequencies for a wide range of plants.

Microprojectile Bombardment

When the difficulties in transforming some plant species first became apparent, a number of procedures that are alternatives to transformation by *A. tumefaciens* were developed. However, with both the improvement of Ti plasmid-based transformation protocols and the success of microprojectile bombardment, the procedures currently used by nearly all researchers, other methods for transforming plants are now only rarely used. At the present time, essentially any plant can be transformed using one of these two procedures.

Microprojectile bombardment, also called biolistics, is by far the most important alternative to Ti plasmid DNA delivery systems for plants (Table 10.4). Spherical gold or tungsten particles (approximately 0.4 to 1.2 μm in diameter, or about the size of some bacterial cells) are coated with DNA that has been precipitated with $CaCl_2$, spermidine, or polyethylene glycol (Fig. 10.6). The coated particles are accelerated to high speed (300 to 600 meters/second) with a special apparatus called a particle gun (or gene gun). The original version of the gene gun used a small amount of gunpowder to provide the propelling force. The device that is currently used employs high-pressure helium as the source of particle propulsion (Fig. 10.7). The DNA-coated metal projectiles can penetrate plant cell walls and membranes; however, at the density used, the particles do not significantly damage the cells.

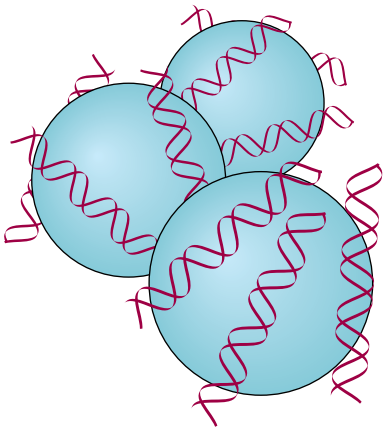
Once inside a cell, the DNA is slowly released from the particles and, in some cells, integrates into the plant genomic DNA. Microprojectile bombardment can be used to introduce foreign DNA into plant cell suspensions, callus cultures, meristematic tissues, immature embryos, protocorms, coleoptiles, and pollen in a wide range of plants, including monocots and conifers, plants that are less susceptible to *Agrobacterium*-mediated DNA transfer. Furthermore, this method has also been used to deliver genes into chloroplasts and mitochondria, thereby opening up the possibility of introducing exogenous (foreign) genes into these organelles.

Typically, plasmid DNA (here it is not necessary to utilize the Ti plasmid) dissolved in buffer is precipitated onto the surfaces of the microprojectiles. Using this procedure, it is possible to increase the transformation frequency by increasing the amount of plasmid DNA; however, too much plasmid DNA can be inhibitory. It is estimated that, on average, there are approximately 10,000 transformed cells formed per bombardment. With this technique, cells that appear to be transformed, based on the expression of a marker gene, often only transiently express the introduced DNA. Unless the DNA becomes incorporated into the genome of the plant, the foreign DNA will be degraded eventually.

Table 10.4 Some plants that have been stably transformed using microparticle bombardment

Plant species	Tissue bombarded
Arabidopsis	Roots
Barley	Embryos
Bean	Meristems
Cotton	Meristems
Cranberry	Stems
Tulip tree	Embryonic culture
Corn	Suspension culture, embryonic culture, embryos
Oat	Embryogenic callus
Orchid	Protocorm
Papaya	Somatic embryos
Poplar tree	Cultured nodules
Rice	Embryos, embryogenic callus
Sorghum	Suspension culture
Soybean	Embryos, apical meristem
Spruce tree	Embryogenic callus, somatic embryos
Sugarcane	Embryogenic callus
Sweet potato	Leaves
Tobacco	Suspension culture, leaves, leaf chloroplasts
Wheat	Embryogenic callus

Figure 10.6 DNA bound to a microprojectile particle.



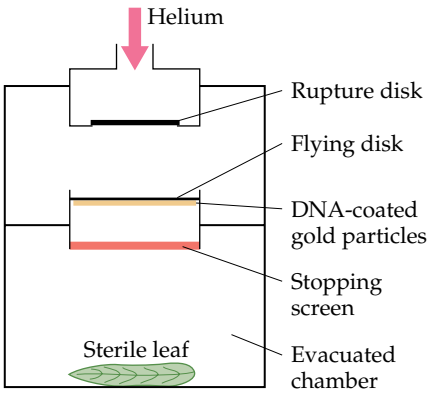


Figure 10.7 Schematic representation of a microprojectile bombardment apparatus. When the helium pressure builds up to a certain point, the plastic rupture disk bursts, and the released gas accelerates the flying disk with the DNA-coated gold particles on its lower side. The gold particles pass the stopping screen, which holds back the flying disk, and penetrate the cells of the sterile leaf.

The configuration of the vector that is used for biolistic delivery of foreign genes to plants influences both the integration and expression of those genes. For example, transformation using microprojectile bombardment is more efficient when linear rather than circular DNA is used. Moreover, large plasmids (>10 kb), in contrast to small ones, may become fragmented during microprojectile bombardment and therefore produce lower levels of foreign-gene expression. However, large segments of DNA may be introduced into plants using yeast artificial chromosomes (YACs). The YACs are engineered to contain plant selectable markers, as well as yeast selectable markers, which are already present on the YAC vector (chapter 3). YACs up to 150 kb in total size have a good chance of being transferred to plant cells, and the transferred DNA can be stably integrated into the plant cell. Thus, the production of transgenic plants that contain several foreign genes is feasible; eventually, entire biosynthetic pathways may be introduced into plant cells.

To predictably increase plant fruit yield, it is essential to alter a variety of plant processes. These processes include increasing the ability of the plant to assimilate carbon and nitrogen, transporting the newly synthesized sugars and amino acids to plant tissues, and converting the newly synthesized sugars and amino acids to storage compounds in the sink tissues. This is not an easy task. However, one group of researchers recently reported making multiple genetic modifications to tomato plants to increase the yield of tomato fruit. For a start, researchers needed to monitor the concentrations of a number of plant metabolites from plants grown under greenhouse conditions. This information was used to develop a computer model of the genes that were involved in fruit yield. Based on this model, these researchers introduced 20 separate genes into tomato plants by biolistic cotransformation. Of these genes, 6 were involved in metabolite assimilation, 3 in transport, and 11 in sink metabolism. Moreover, 16 genes were overexpressed and 4 were silenced. Although this procedure only increased the fruit yield by 23%, it was nevertheless quite impressive that so many foreign genes could be successfully introduced by microprojectile bombardment.

Although microprojectile bombardment enables scientists to transform monocots and numerous other plants, some plant varieties remain recalcitrant to this procedure. To overcome this hurdle, scientists overexpressed two plant growth-stimulating genes, *Bbm* (baby boom) and *Wus2* (Wuschel2). The products of both of these genes are considered to be regulators of plant embryogenesis, so their presence makes plants more likely to take up input foreign DNA. A number of different protocols have employed these genes to increase the frequency of transformation of several different monocots. One of those protocols is shown in Fig. 10.8. Some similar procedures employ only the *Wus2* gene, and some procedures simultaneously introduce the target gene and the *Wus2* (and *Bbm*) gene(s). With these changes, transformation frequencies of other recalcitrant plants are greatly improved. To date, this approach has been shown to be effective in corn, sorghum, sugarcane, and rice.

Chloroplast Engineering

Most higher plants have approximately 50 to 100 chloroplasts per leaf cell, and each chloroplast has about 10 to 100 copies of the chloroplast DNA genome (Fig. 10.9). Thus, while a single copy of a foreign gene inserted

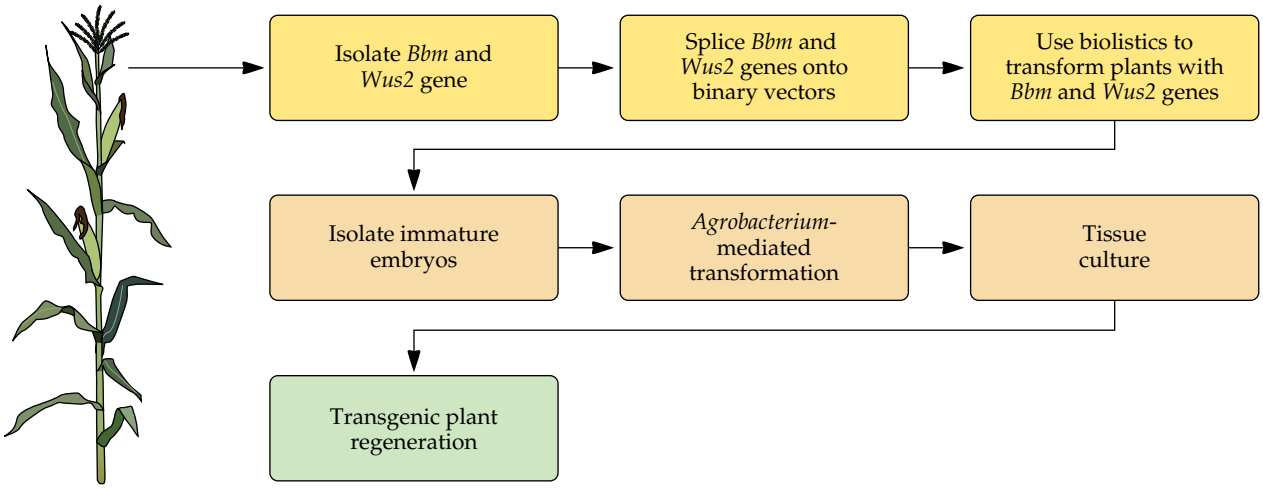


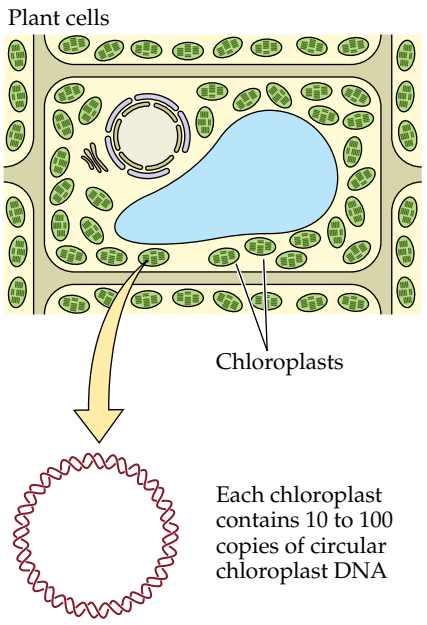
Figure 10.8 Schematic overview of a procedure developed to increase the transformation of monocots. Once the *Bbm* and/or *Wus2* genes are introduced into the target plant cells, the plant is more readily engineered by *Agrobacterium*-mediated transformation.

into plant chromosomal DNA is present in one copy per plant cell, a single copy of the same gene inserted into chloroplast DNA will be present in 500 to 10,000 copies per plant cell. The many copies of chloroplast DNA in each plant cell means that much higher levels of foreign gene expression should be attainable when the introduced gene is present as part of the chloroplast DNA. Researchers using this approach have reported that a recombinant protein may be expressed to a level of around 1 to 7% of the soluble plant protein. For example, active human somatotropin accumulated to 7% of the total soluble protein in tobacco chloroplasts. This is 10- to 100-fold higher than protein expression levels that are typically obtained when foreign DNA is inserted into the plant genomic DNA.

While the vast majority of plant genes are found as part of the nuclear DNA, both the chloroplast and mitochondrion contain genes that encode a number of important and unique functions. Moreover, not all of the proteins that are present in these organelles are encoded by organellar DNA. For chloroplasts, in particular, some proteins are encoded in the nuclear DNA, synthesized in the cell's cytoplasm, and then imported into the chloroplast by a specialized transport mechanism. Accordingly, there are two ways that a specific foreign protein can be introduced into the chloroplast. In one approach, a fusion gene encoding the foreign protein and additional amino acids that direct the transport of the protein to the organelle can be inserted into the nuclear chromosomal DNA using the *A. tumefaciens*–Ti plasmid system and a plant selectable marker (Fig. 10.10A). After the recombinant protein is synthesized, it can be transported into the chloroplast by the chloroplast targeting/transit peptide. This peptide is removed upon insertion of the target protein through the chloroplast membrane. This approach does not necessarily result in very high levels of expression of the target protein.

In the other approach, the gene for the foreign protein can be inserted directly into the chloroplast DNA (Fig. 10.10B). In this case, foreign DNA is typically introduced into the chloroplast genome by microprojectile bombardment of the genetic construct on a plasmid vector. The plasmid

Figure 10.9 Plant cells showing multiple chloroplasts and double-stranded circular chloroplast DNA.



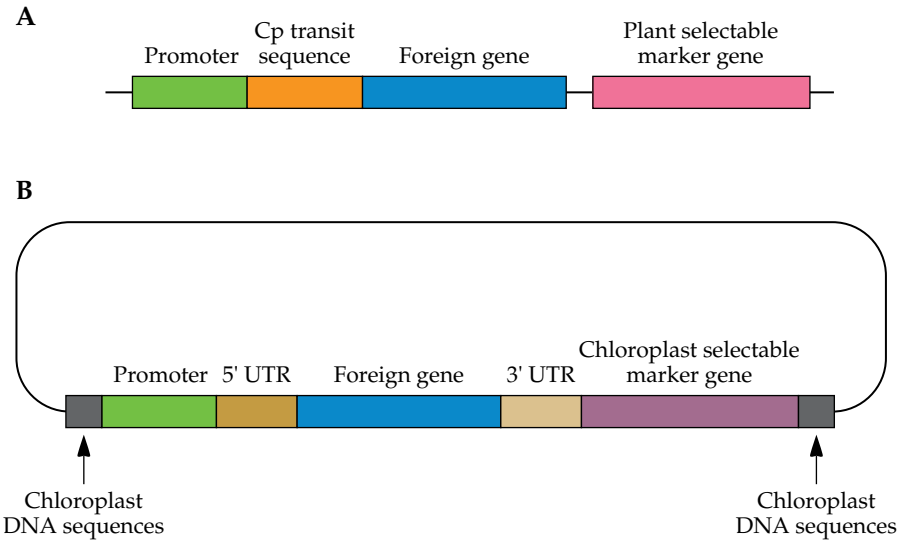


Figure 10.10 **(A)** A foreign gene that is introduced on a Ti plasmid and hence inserted into the nuclear/chromosomal DNA. The gene encodes a fusion protein that includes a chloroplast (Cp) targeting/transit sequence that facilitates the localization of the foreign protein in the chloroplast. The Cp targeting/transit sequence is cleaved during transport into the chloroplast. **(B)** Plasmid vector used for integrating a foreign gene and a marker gene directly into the chloroplast genome. Homologous recombination can occur between chloroplast DNA sequences on the vector and the chloroplast genome. The regulatory sequences controlling expression of the selectable marker genes are not shown. 5′ UTR and 3′ UTR, 5′ and 3′ untranslated regions, respectively.

generally contains both the foreign DNA and a chloroplast selectable marker flanked by specific chloroplast DNA sequences. Chloroplast selectable markers are designed to be expressed only within chloroplasts and not within the cytoplasm. Homologous recombination is the normal mode of DNA integration into the chloroplast genome; in this case, integration is targeted to a region of the DNA that does not encode proteins (an intergenic region).

The efficient expression of foreign genes in the chloroplast requires an appropriate (chloroplast-specific) promoter sequence and the presence of translational signals in the 5′ and 3′ untranslated regions of the mRNA. Translational control sequences that function only within the chloroplast and not within the cytoplasm are included on the cloned DNA together with the gene of interest and a 3′ chloroplast untranslated region containing a stem-and-loop structure, which is expected to act as a transcription termination signal (Fig. 10.10B). An additional advantage of inserting a target gene into the chloroplast genome is that since chloroplast transcription and translation are prokaryote-like, several genes may be introduced and expressed at the same time in a single operon.

Chloroplast DNA is inherited in a non-Mendelian fashion. In most angiosperms (flowering plants), it is maintained in egg cells but not sperm (pollen) cells and is therefore transmitted uniparentally by the female. Thus, pollen cannot transmit the contents of the chloroplast genome to the zygote. This has practical importance, since this trait can prevent the spread of foreign genes, localized in the chloroplast, through pollen to

neighboring plants, thereby addressing one of the concerns of some critics about the cultivation of genetically engineered plants.

Most chloroplast transformation protocols include selectable markers that confer resistance to antibiotics that inhibit protein synthesis on prokaryote-type ribosomes (Table 10.5). These antibiotics inhibit greening (chlorophyll production), the rate of plant proliferation, and shoot formation. Plant cells with transformed chloroplasts that express the antibiotic resistance genes are readily identified in the presence of (otherwise) inhibitory antibiotics by their greening, faster proliferation, and shoot formation. Since only one, or at most a few, of the chloroplasts actually incorporates the foreign DNA, repeated rounds of growth of cultured plant cells on agar plates containing selective antibiotics are often required so that the chloroplast population becomes enriched and eventually dominated by the transformants. Eventually, all chloroplast genomes that do not carry the transgenic DNA with their selectable marker gene are lost.

Chloroplasts belong to a group of plant organelles called plastids, which contain approximately 120 to 180 kb of circular double-stranded DNA bounded by a double membrane. Plastids include nonpigmented organelles called amyloplasts, which synthesize and store starch grains, chloroplasts, which contain chlorophyll, elaioplasts, which contain oil, and chromoplasts, which contain other pigments. While photosynthetic tissues, such as green leaves, contain chloroplasts, other tissues contain other types of plastids. For example, tomato fruit contains a large number of chromoplasts. Thus, in much the same way that foreign DNA may be expressed as a part of the chloroplast genome, it can also be targeted for expression in the chromoplast.

Very-High-Level Protein Expression

Facilitating the move toward greater use of plants to produce foreign proteins is the promise of very high expression levels following the expression of foreign genes in the chloroplast. With this is in mind, it is worth examining in some detail how one group of researchers was able to obtain an exceptionally high level of synthesis of a therapeutic protein in transgenic tobacco plants.

The target protein for experiments to optimize the expression of therapeutic proteins in plants was a bacteriophage lytic protein that acts as an antibiotic against pathogenic bacteria (group A and group B streptococci). The PlyGBS lysin gene was isolated from a bacteriophage that infects *Streptococcus agalactiae*, a group B streptococcus. Infections caused

Table 10.5 Some foreign genes that have been used as selectable markers and reporters of transgenic chloroplasts[a]

Gene	Gene product	Function
aadA	Aminoglycoside 3′-adenylyltransferase	Positive selection (spectinomycin and streptomycin resistance)
nptII	Neomycin phosphotransferase	Positive selection (kanamycin resistance)
uidA	β-Glucuronidase	Reporter gene
gfp	Green fluorescent protein	Reporter gene
coda	Cytosine deaminase	Negative selection (5-fluorocytosine sensitivity)

[a]Data from Hager and Bock, *Appl Microbiol Biotechnol* **54:**302–310, 2000.

by group A and B streptococci are the leading causes of neonatal sepsis (the presence in a newborn baby of a bacterial bloodstream infection) and meningitis (inflammation of the protective membranes covering the brain and spinal cord), and as persistent colonizers of the human genital and gastrointestinal tracts, pathogenic group B streptococci can sometimes be transmitted to a fetus during pregnancy. The PlyGBS lysin kills all known group B streptococcus strains and reduces colonization by these bacteria in both the vagina and the oropharynx (the part of the pharynx between the soft palate and the upper edge of the epiglottis). PlyGBS lysin is also active against group A streptococci such as *Streptococcus pyogenes*, a human pathogen that colonizes the skin and the mucous membranes of the upper respiratory tract, causing a variety of infections.

To maximize the production of the PlyGBS lysin in transgenic plants, several high-expression strategies were combined (Fig. 10.11). Following purification of the bacteriophage DNA, the *plyGBS* gene was isolated and its sequence was determined. To ensure that a high level of the PlyGBS lysin could be produced in chloroplasts, the *plyGBS* gene was completely resynthesized so that its codon usage corresponded to the codon usage found in the AT-rich tobacco chloroplast genome. The synthetic gene was then fused to the strongest known promoter in chloroplasts, i.e., the constitutive rRNA operon promoter. The promoter was followed by the gene 10 leader RNA from phage T7, which provides a strong ribosome-binding site. Next, the *plyGBS* gene expression cassette was inserted into a chloroplast transformation vector between chloroplast DNA sequences to target the transgene to the intergenic spacer between the *trnfM* and *trnG* genes by homologous recombination. Chloroplasts were transformed by microprojectile bombardment of tobacco leaves followed by selection on spectinomycin-containing plant regeneration medium. A spectinomycin resistance gene was used as the chloroplast selectable marker gene (Fig. 10.11).

Transgenic tobacco plants regenerated from the cells of the transformed tobacco leaves expressed an extraordinarily high level of the PlyGBS lysin. Not surprisingly, plants expressing the *plyGBS* gene cassette were somewhat debilitated in their growth compared to wild-type tobacco plants. This is a reflection of the enormous metabolic load imposed on these plants by the overproduction of the lysin protein. Notwithstanding the slower growth, the plants developed normally and produced viable seeds. When this technology is commercialized, these transgenic plants will likely be confined to a greenhouse environment to avoid release of lysin into the environment that might lead to the development of lysin-resistant bacteria.

It has been suggested that for bacteriophage lysins to be effective at killing their bacterial hosts, the lysins must be highly resistant to degradation by bacterial proteases. Thus, at least in part, the extraordinary stability of the lysin protein itself may facilitate its stable accumulation in chloroplasts. Since most other proteins are not as stable as the PlyGBS lysin, it may be difficult to express other foreign proteins in chloroplasts to the extent observed for the PlyGBS lysin. Nevertheless, the strategy of (i) optimizing the codon usage of the foreign gene, (ii) fusing the foreign gene to the rRNA operon promoter and the T7 gene 10 leader peptide, and (iii) expressing the construct in plant chloroplasts should provide an effective means of obtaining relatively high levels of many different foreign proteins.

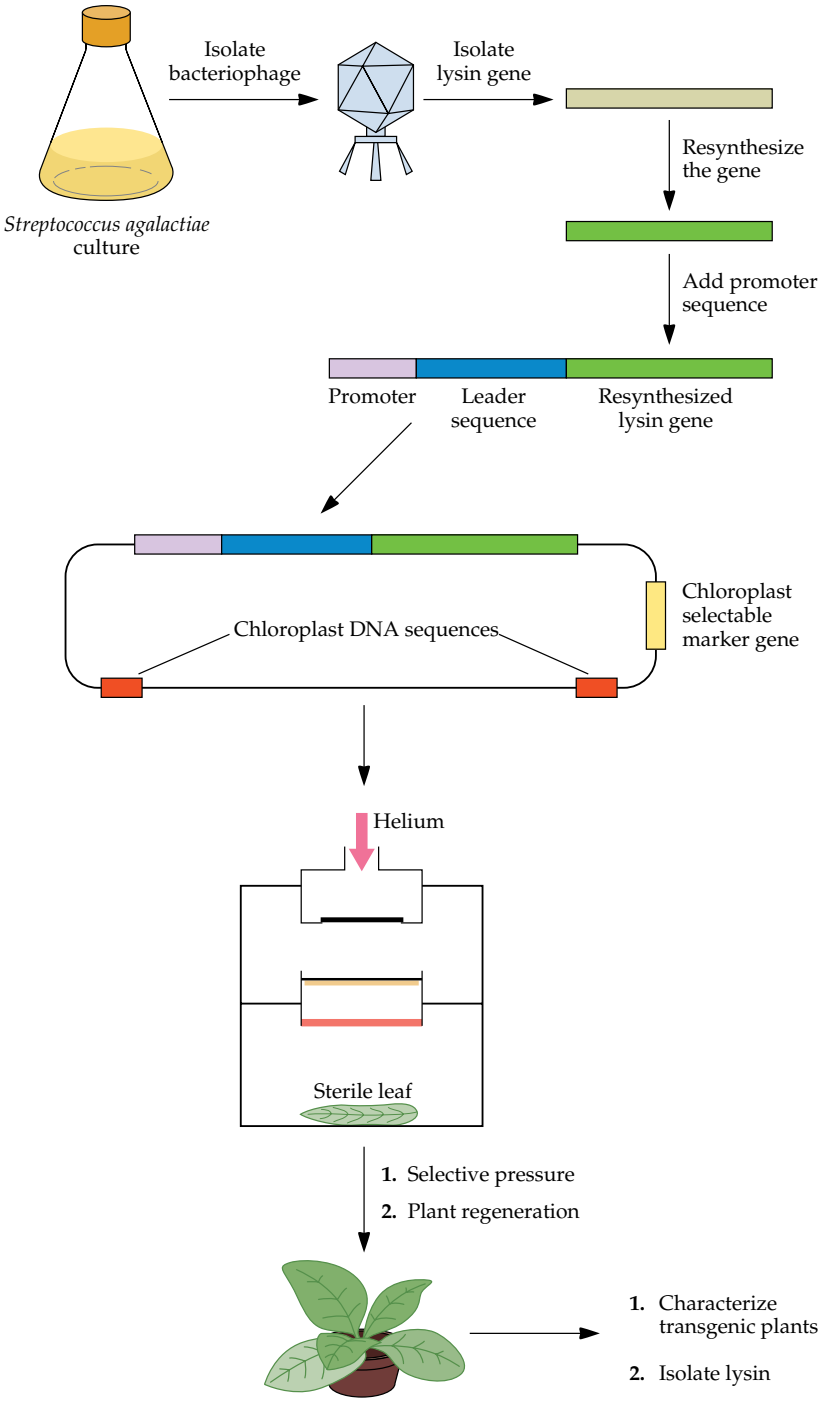


Figure 10.11 Strategy for the overproduction of a bacteriophage lysin protein in transgenic tobacco plants. Following its isolation, the native bacteriophage lysin gene was resynthesized to better reflect the codon usage in tobacco chloroplasts. For expression, the promoter was from chloroplast rRNA genes, while the leader sequence was a strong ribosomal binding site from bacteriophage T7. The chloroplast selectable marker gene encoded resistance to the antibiotic spectinomycin. The chloroplast DNA sequences on the plasmid vector ensured insertion by homologous recombination into a chloroplast intergenic region. The construct was used to transform tobacco using microprojectile bombardment.

Use of Reporter Genes in Transformed Plant Cells

It is essential to be able to detect the foreign DNA that has been integrated into plant genomic DNA so that those cells that have been transformed can be identified; this requires the use of a selectable marker. Dominant-marker selection provides a direct means of obtaining only transformed cells in culture. For example, in the presence of the antibiotic kanamycin, only plant cells that express a neomycin phosphotransferase gene can grow. Furthermore, in studies of plant transcriptional regulatory signals and the functioning of these signals in specific plant tissues (such as leaves, roots, and flowers), it is often important to be able to quantify the level of expression of a gene with a readily identified product. Quantification and other applications require the use of reporter genes that encode an activity that can be assayed.

A number of different genes have been tested as reporters for transformation, including genes that can be used as dominant selectable markers and genes whose proteins produce a detectable response to a specific assay (Table 10.6). Many of these reporter genes are from bacteria and have been equipped with plant-specific regulatory sequences for expression in plant cells. The desired outcome of a particular experiment often dictates which reporter gene will be used. Clearly, when the expression of a reporter gene interferes with normal plant functions, it cannot be used. Moreover, the presence of some reporter genes and their products may taint a commercial product. In this context, it is best to remove the reporter

Table 10.6 Plant cell reporter and selectable marker gene systems[a]

Enzyme activity	Selectable marker	Reporter gene
Neomycin phosphotransferase	Yes	Yes
Hygromycin phosphotransferase	Yes	Yes
Dihydrofolate reductase	Yes	Yes
Chloramphenicol acetyltransferase	Yes	Yes
Gentamicin acetyltransferase	Yes	Yes
Nopaline synthase	No	Yes
Octopine synthase	No	Yes
β-Glucuronidase	No	Yes
Streptomycin phosphotransferase	Yes	Yes
Bleomycin resistance	Yes	No
Firefly luciferase	No	Yes
Bacterial luciferase	No	Yes
Threonine dehydratase	Yes	Yes
Metallothionein II	Yes	Yes
enol-Pyruvylshikimate-3-phosphate synthase	Yes	No
Phosphinothricin acetyltransferase	Yes	Yes
β-Galactosidase	No	Yes
Blasticidin S deaminase	Yes	Yes
Acetolactate synthase	Yes	No
Bromoxynil nitrilase	Yes	No
Green fluorescent protein	No	Yes

[a]Data from Walden and Schell, *Eur J Biochem* **192:**563–576, 1990, and Gruber and Crosby, p. 89–119, in B. R. Glick and J. E. Thompson (ed.), *Methods in Plant Molecular Biology and Biotechnology* (CRC Press, Boca Raton, FL, 1993).

gene once transformants with the desired traits have been selected, especially in crop plants.

Some reporter gene products, e.g., β-D-glucuronidase (GUS), both firefly and bacterial luciferases, and green fluorescent protein (GFP), can be detected *in situ* in intact plant tissues. One of the most popular of these systems is the *E. coli uidA* gene, which encodes the protein GUS, a stable enzyme that is not normally present in plants and that catalyzes the cleavage of a range of β-D-glucuronides. The GUS activity in transformed plant tissues can be localized by the presence of a blue color that is formed after the hydrolysis of the uncolored synthetic substrate 5-bromo-4-chloro-3-indolyl β-D-glucuronic acid. Alternatively, GUS activity in plant extracts can be quantitatively and sensitively assayed by a fluorometric analysis that involves the hydrolysis of the substrate 4-methylumbelliferyl β-D-glucuronide to form a fluorescent product.

GFP is an ideal *in vivo* marker for monitoring transgenic plants because it fluoresces green when excited with either ultraviolet or blue light and does not require the addition of any substrates or cofactors. Normally, wild-type plants (which do not contain the gene for GFP) fluoresce reddish purple when they are excited by ultraviolet or blue light. Moreover, derivatives of the gene that encodes the GFP protein that have different spectral properties and increased levels of fluorescence have been constructed (see chapter 4). Importantly, the expression of GFP has no effect on the survival and growth of transformed plants under field conditions.

Manipulation of Gene Expression in Plants

Transient Gene Expression

The development of transgenic plants that can act as factories for the production of large amounts of therapeutic proteins is a time-consuming, expensive, and inefficient process. Nevertheless, it is possible to achieve high yields of recombinant proteins in plants. Unfortunately, all of the strategies currently available have some limitations. For example, high expression has been reported for nuclear transgenes in seeds; however, because the amount of seed harvested per hectare is much lower than the green biomass harvest (i.e., only a small percentage), the yield of the recombinant protein per hectare is relatively low. Similarly, recombinant proteins can be expressed in high yield in chloroplasts; however, proteins synthesized in chloroplasts are not glycosylated, a situation that is problematic for many therapeutic proteins.

To get around the above-mentioned constraints to the high-level expression of foreign proteins from genes stably integrated into the plant genome, it is possible to transiently express foreign genes in plants in large quantities. Foreign genes that are transiently expressed in plants do not integrate into the plant nuclear or chloroplast DNA. Rather, they exist stably for a limited period (often weeks to months) in the plant cell cytoplasm, typically under the control of strong expression signals such as those found in plant viruses. In this way, the protein may be produced rapidly (transgenic plants do not need to be produced, selected, or propagated) and in high yield. Since the foreign gene is not incorporated into the plant genome, it is not inherited.

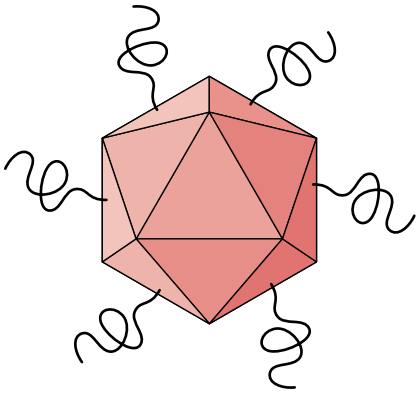
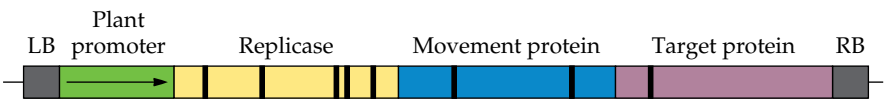


Figure 10.12 Schematic representation of a first-generation viral vector for transient expression of foreign peptides in plants. Here, the foreign peptide is fused to the viral coat protein.

With "first-generation viral vectors," the virus expresses all of its own genes in addition to the target gene, which is spliced into its genome, either under the control of a strong viral promoter or fused to a gene encoding the viral coat protein, which is highly expressed and is present on the surface of assembled viruses (Fig. 10.12). The foreign gene may be delivered to plant cells either using infectious (naked) nucleic acid versions of the vector that are not assembled into a whole virus or as mature viral particles. When mature viral particles are used, the entire process can be performed on a large scale by spraying plants in the field with a mixture of viral particles and an abrasive compound such as carborundum (silicon carbide). Depending on how efficiently the viral particles can move through the plant, it may take about 2 to 3 weeks for most of the tissues of the transformed plants to become infected. This approach has been somewhat successful in that there are numerous reports of plant expression of various therapeutic proteins as fusions to viral coat proteins (Fig. 10.12).

After purification of recombinant vaccine antigens from plants and injection into animals, some of these proteins have been found to be immunogenic and, in some instances, to demonstrate some level of protection against pathogenic agents in animal models, i.e., they have been used as vaccines. Unfortunately, only small epitopes (i.e., fewer than 25 amino acids) have been successfully expressed as coat protein fusions. This is probably because larger epitopes can prevent the viral particle from assembling. To overcome this limitation, scientists constructed "second-generation viral vectors." In this approach, the recombinant viral DNA is delivered to the plant by *A. tumefaciens* infection. This avoids the high level of metabolic resources that are devoted to the synthesis of large amounts of viral coat protein and instead facilitates a high level of expression of the target protein. The viral genome is modified so that only the viral elements required for efficient expression of the target gene are maintained (Fig. 10.13). Since the modified viral DNA is delivered to the plant by *A. tumefaciens*, viral genes responsible for infectivity, amplification/replication, cell-to-cell movement, assembly of viral particles, shutoff of the synthesis of plant cell components, and systemic spread of the virus are dispensable and therefore removed. In addition, by selectively modifying the vector, e.g., (i) introducing silent mutations (mutations that do not alter the encoded amino acids) to remove potential cryptic splice sites, (ii) changing the codon usage, and (iii) adding multiple plant introns, it has been possible to construct more efficient viral delivery systems. The plant introns are thought to make the viral RNA more plant mRNA-like and therefore more efficiently transcribed, processed, and translated. With the improved vectors, there is typically

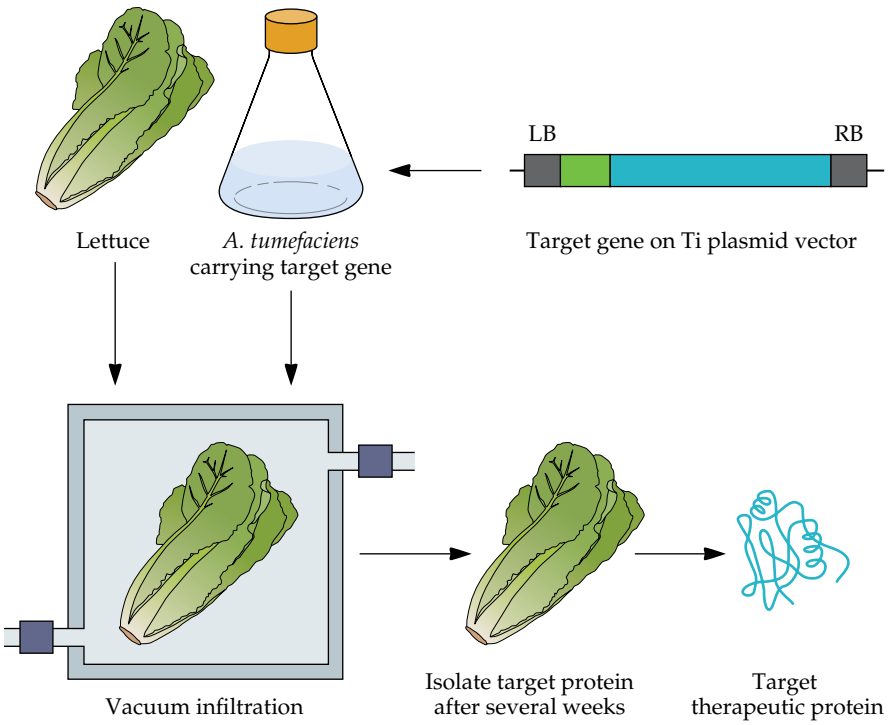
Figure 10.13 Second-generation transient-expression vector system. Shown is T-DNA carrying viral genes for replicase and movement protein as well as a target protein gene. The viral replicase is an RNA-dependent RNA polymerase. The movement protein facilitates movement from one plant cell to another. The arrow indicates the direction of transcription. The heavy vertical lines within genes represent introduced introns. In addition to the genes shown, between the left (LB) and the right (RB) borders there is a plant selectable marker gene.



one successful infection event per 10 to 20 copies of *A. tumefaciens* that are introduced by vacuum infiltration, a process in which the aerial part of the plant is immersed in an *A. tumefaciens* suspension and then a weak vacuum is then applied for 10 to 30 seconds. Vacuum infiltration replaces the viral functions of infection and systemic movement. Amplification of the viral genome within a plant cell and movement from one plant cell to another are performed by the viral proteins (i.e., replicase and movement protein) that remain encoded in the modified viral DNA. Starting with vacuum infiltration, the entire process takes 4 to 10 days before recombinant protein can be isolated from plant leaves. Moreover, the approach is effective with a range of plants. It is estimated that in a 1-hectare plant growth facility, this system may be able to produce up to 400 kilograms of recombinant protein per year. Since this system does not require a large amount of land, growth of transformed plants could be limited to a contained greenhouse facility.

One group of workers demonstrated that with a transient expression system, it was possible to dip a head of lettuce, obtained from a grocery store, into an *A. tumefaciens* cell suspension, vacuum infiltrate the bacterium (carrying the engineered viral DNA), and within about 1 week's time obtain around 20 to 100 milligrams of functional target protein per kilogram of fresh lettuce leaf tissue (Fig. 10.14). With this system, plant growth facilities are not required; the system is very efficient and inexpensive and is easily scaled up. This transient-expression system,

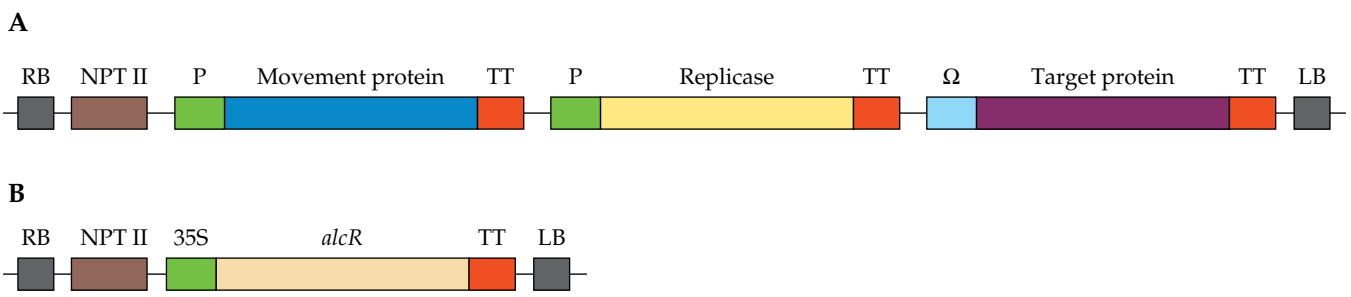
Figure 10.14 Transformation of a head of lettuce with T-DNA carrying viral DNA and a specific target gene using a strain of *A. tumefaciens*. Following incubation and vacuum infiltration, the target protein may be harvested from the transiently transformed lettuce within 1 to several weeks' time. LB and RB, left and right borders, respectively.



based on expression from viral RNA replicons (where an RNA replicon is RNA that replicates from a single origin of replication) delivered into plant cells by *A. tumefaciens*, allows production of recombinant proteins at yields of up to 5 grams per kilogram of fresh leaf biomass. Such high yields are possible because plant biomass accumulation takes place prior to infection.

Notwithstanding the high yields of foreign protein that can be obtained, this approach requires special equipment for vacuum infiltration and may be expensive to perform on a large scale. In addition, containment of the genetically engineered agrobacteria producing modified viral replicons that are used for transfection is required. To avoid these problems, researchers have developed an ethanol-inducible system that provides efficient release of viral RNA replicons from proreplicons (which lack a replication gene) contained in a stably transformed cassette (Fig. 10.15A). The cassette carrying the viral vector has two components, the viral replicon and a gene encoding the movement protein, and each of these components is placed separately under the transcriptional control of an ethanol-inducible promoter. In this system, the viral replicon and the movement protein are under the control of the alcohol dehydrogenase gene (*alc*) promoter from the fungus *Aspergillus nidulans*. The *alc* promoter is activated only when it is bound by the AlcR transcriptional activator, encoded on a separate construct, in the presence of ethanol (Fig. 10.15B). In practice, to activate the production of a target protein, transgenic plants carrying the two constructs shown in Fig. 10.15 are sprayed twice a day for around 5 days with a solution of 4% ethanol. Alternatively, expression of the target gene was strongly induced when the roots were soaked and then sprayed once with 4% ethanol. Ethanol was chosen as the inducer in this system because it is relatively inexpensive and is nontoxic and because the system exhibits a very low background level of expression in the absence of the inducer. This system should be readily amenable to scale-up. What remains is to determine whether this system is as effective and reliable in practice as it seems.

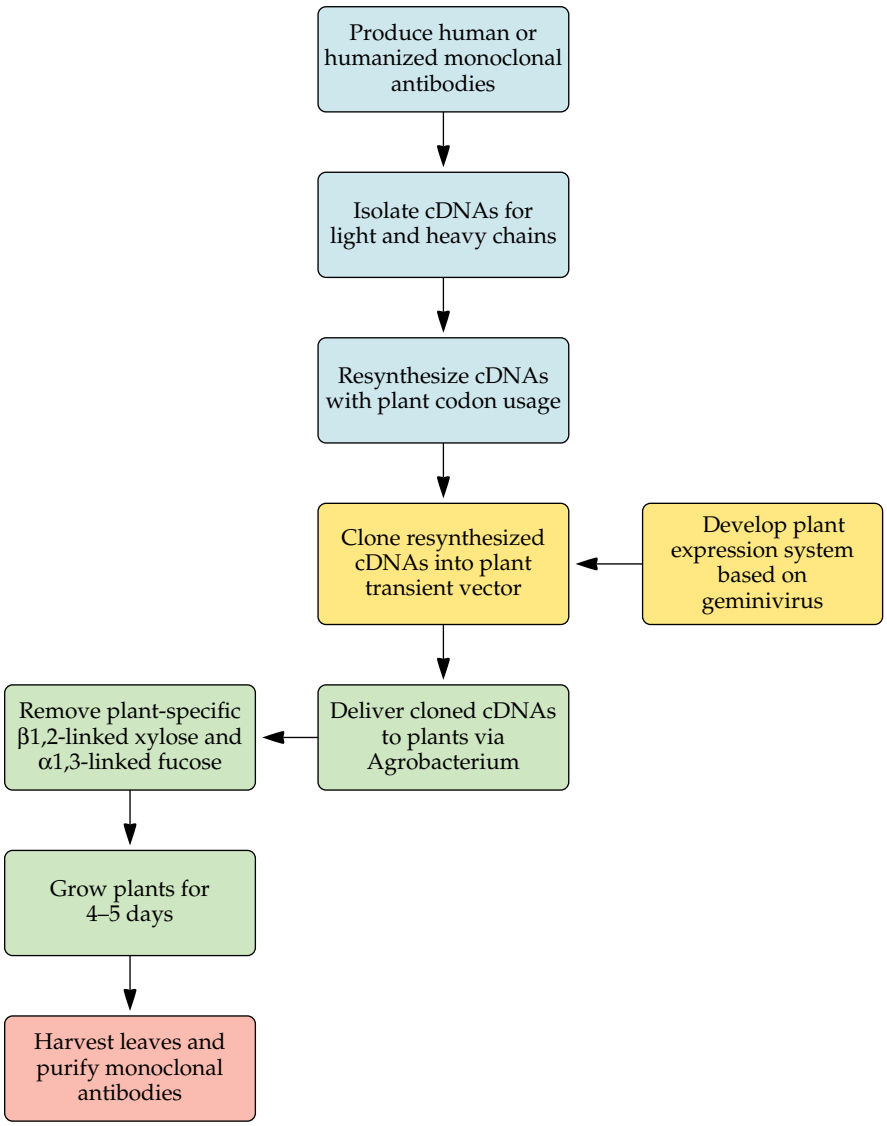
Figure 10.15 **(A)** Second-generation transient-expression vector system requiring the addition of ethanol for activation. The construct contains left (LB) and right (RB) T-DNA border sequences; a neomycin phosphotransferase gene (NPT II), whose promoter and transcription termination regions are not shown, that acts as a plant selectable marker; movement and replicase genes, each under the transcriptional control of an *alc* promoter (P) and transcription termination sequences (TT); and the target gene preceded by a sequence that enhances translation (Ω). **(B)** T-DNA encoding the *alcR* gene under the control of the constitutive 35S promoter.



Transient Expression of Antibodies

The production of human therapeutic agents, especially monoclonal antibodies, is very expensive, requiring large investments in specialized equipment with skilled personnel to run it and significant operating costs. However, with the enormous health benefits of these therapeutic agents, there is widespread demand for more inexpensive versions of these highly efficacious therapeutic molecules which can be achieved in relatively inexpensive plant-based systems. For example, to produce humanized monoclonal antibodies in the tobacco-like plant *Nicotiana benthamiana*, the following steps were undertaken (Fig. 10.16). Starting with isolated cDNA, the nucleotides encoding the amino acid sequence of a mouse monoclonal antibody directed against Zika virus were altered to (i) humanize the

Figure 10.16 Overview of the steps involved in producing humanized monoclonal antibodies in the tobacco-like plant *Nicotiana benthamiana*.



monoclonal antibody (removing all mouse DNA sequences) and (ii) optimize its codon expression in plants. The resynthesized cDNAs were spliced into geminivirus expression vectors (second-generation transient expression vectors), which were delivered to plant leaves using *Agrobacterium*. The strain of *N. benthamiana* that was used was previously engineered so that plant-specific β1,2-linked xylose and α1,3-linked fucose molecules were no longer added to the glycans on the surface of the foreign monoclonal antibody. This monoclonal antibody molecule was produced at a level of ~1.5 grams per kilogram of leaf tissue and was correctly assembled so that it recognized the Zika virus envelope protein and neutralized the Zika virus. Using affinity chromatography, this monoclonal antibody could be purified to near homogeneity in a single step (see chapter 3). This is an important demonstration of the ability of plant systems to be an effective and relatively inexpensive tool for the production of multiprotein human therapeutic proteins.

Plant Promoters

Many early experiments to express foreign proteins utilized promoters that were constitutively active in a wide range of plant cells. However, in order to minimize any deleterious effects on plant growth from the expression of foreign genes, it is necessary to regulate the expression of introduced genes, spatially and temporally, and the amount of the foreign protein that is produced. More recently, many additional plant promoters have been isolated and characterized and used to express foreign proteins in specific cells at certain times during the growth and development of the plant. For example, instead of the strong constitutive 35S promoter from cauliflower mosaic virus, which is expressed in all dicot plant tissues (and to a lesser extent in monocot plant tissues) and throughout the life of the plant, researchers have used the strong promoter for the gene encoding the small subunit of the photosynthetic enzyme ribulose bisphosphate carboxylase, which is active only in photosynthetic (green) tissues, such as leaves. Similarly, plant promoters active only in specific tissues, such as roots, seeds, or flowers, or only during periods of environmental stress—e.g., the pathogenesis-related promoters—have been used to control the expression of some foreign genes.

While a major factor in the initiation of transcription in plants is the binding of RNA polymerase II to the promoter sequence, other factors also affect this process. Much of the specificity of transcription in plants is controlled by sequence-specific transcription factors and/or enhancer-binding proteins (Fig. 10.17). Enhancer elements are regions on the DNA that bind to enhancer-binding proteins, which interact with transcription factors and RNA polymerase to increase the level of transcription from a particular promoter. Enhancer sequences are generally considered to be important determinants of the tissue and temporal specificity of gene expression. While the importance of promoter and enhancer sequences has been recognized, a thorough understanding of how these and other elements regulate plant gene expression is still not completely available. Nevertheless, using a variety of different promoters, scientists have been able to engineer both tissue-specific gene expression and high levels of gene expression.

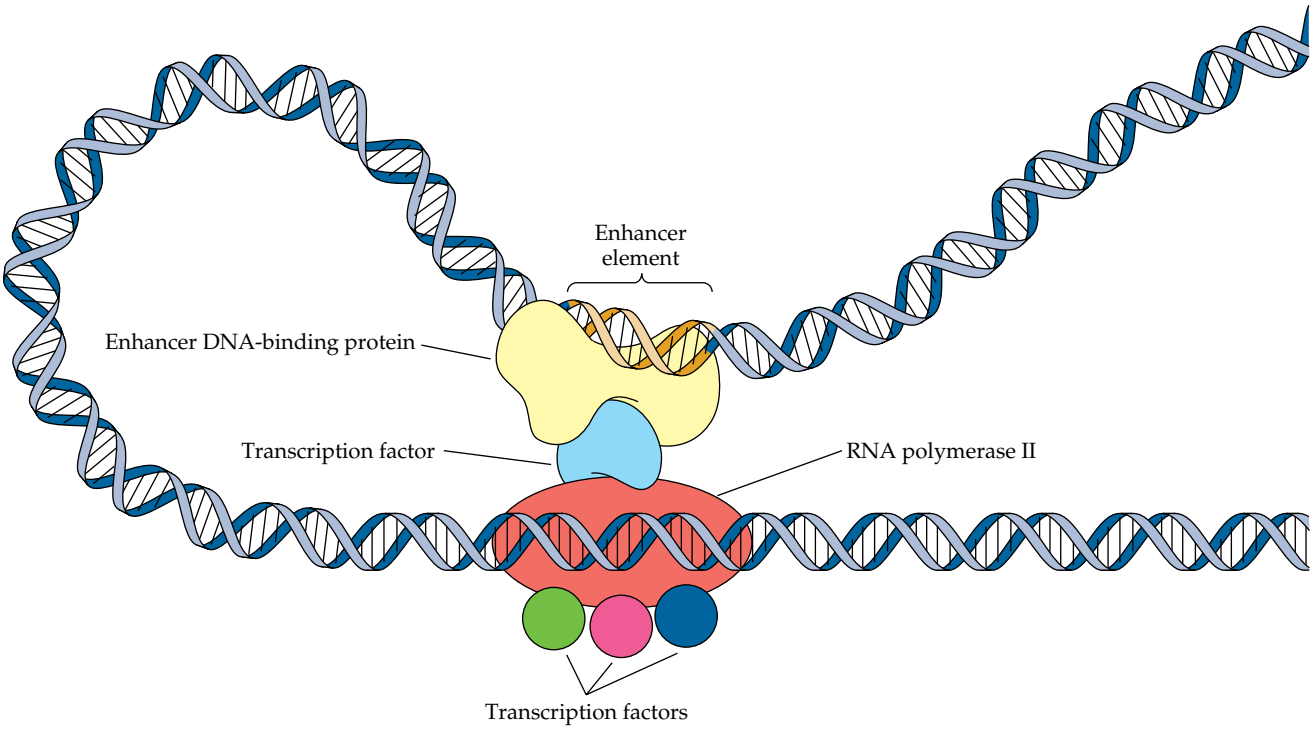


Figure 10.17 Schematic representation of a plant RNA transcription complex, including RNA polymerase II binding to the TATA box (an essential portion of a eukaryotic promoter), various transcription factors bound to the polymerase, and an enhancer-binding protein bound to an enhancer element on the DNA, which is located at some distance from the promoter sequence.

Manipulation of Genes in Plants

The targeted editing of a plant genome could offer tremendous advantages over traditional methods of plant breeding to create new cultivars with advantageous combinations of alleles, especially when several traits are needed for crop improvement. Traditional methods of plant breeding are designed to combine desirable alleles of different genes; they involve extremely tedious and time-consuming genetic crosses and selection for the sought-after phenotype, and often, as with closely linked loci, traditional methods cannot be used to produce the desired result. While the ability to create transgenic plants has solved some of these problems, consumers and regulators in a number of countries, especially in Europe, have expressed serious concerns and reservations about their commercialization and widespread use. On the other hand, *in situ* modification of an existing plant gene is currently not considered to be genetic engineering in many countries and therefore should allow some specifically altered plants to more rapidly reach the marketplace in those countries. With the targeted stable modification of plant genomes, no foreign DNA is introduced into the plant. Some recently developed and perfected techniques, including the use of zinc finger nucleases, transcription activator-like effector nucleases (TALENs), and the clustered regularly interspaced short palindromic repeat CRISPR/Cas9 system have dramatically increased the ease with which it is possible to target and modify specific plant genes.

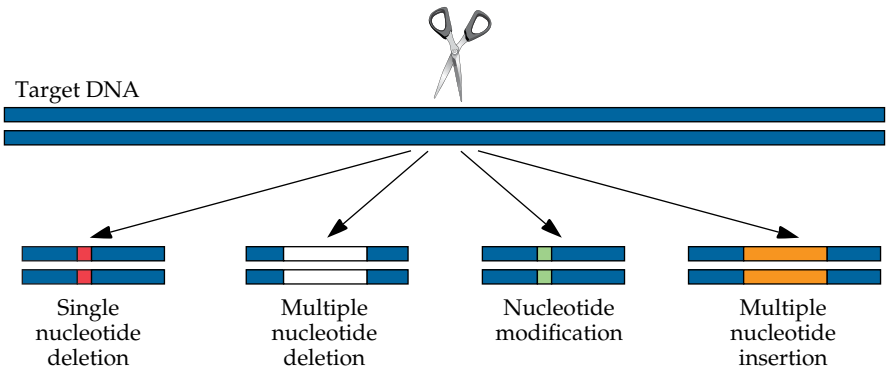
Table 10.7 Some of the plants whose genomes have been modified using the CRISPR/Cas system

Tomato	*Arabidopsis*
Strawberry	Corn
Banana	Poplar
Grape	Rice
Apple	Potato
Watermelon	Wheat
Kiwifruit	Soybean
Cherry	Cotton
Cucumber	Grapefruit
Sunflower	Sorghum
Tobacco	Camelina
Cassava	Orange
Cacao	Mushroom
Papaya	

The most promising of the three above-mentioned gene-targeting systems (and also the newest) exploits the bacterial CRISPR/Cas system (see chapter 2). The simplest form of this technology consists of a single guide RNA molecule (sgRNA) where part of the molecule matches the target DNA sequence and an endonuclease from the bacterium *Streptococcus pyogenes* (Cas9). The sgRNA and the Cas9 endonuclease form a complex in which the sgRNA guides the complex to its target DNA and the nuclease activity of Cas9 then cleaves the target DNA in both strands. With the CRISPR/Cas9 system, only the 20 nucleotides in the sgRNA need to be changed to confer different target specificities, and therefore it is possible to alter several genes at the same time, knocking out redundant or parallel pathways.

The first reports of the use of the CRISPR/Cas9 system to alter plant genomic DNA were published in mid-2013. Since that time, there have been hundreds of reports of the use of this technology to alter the genomes of a variety of plants (Table 10.7). The CRISPR/Cas9 system is used to create single-base-pair deletions and whole or partial gene knockouts and to introduce gene modifications, gene fragment deletions, and gene insertions (Fig. 10.18). Given the very rapid pace of the development of this technology, one can envision a wide range of potentially commercializable CRISPR/Cas9-modified plants being developed within the next 5 to 10 years. Initially, CRISPR/Cas9-modified plants had fewer regulatory hurdles to overcome in Europe, as they were viewed as not being genetically modified organisms. However, on 25 July 2018, the Court of Justice of the European Union ruled that crops developed using this technology must be treated the same way as other genetically modified organisms. In a different venue, in 2014, the New Zealand Environmental Protection Agency ruled that plants produced by gene-editing methods, where no foreign DNA was introduced into the edited plant, were not to be regulated as genetically modified organisms; however, a subsequent court challenge overturned this ruling.

Figure 10.18 Schematic representation of plants whose genome has been modified in a targeted manner by CRISPR/Cas9. The scissors represent the CRISPR/Cas9 complex. A single or multiple nucleotide deletion yields a knockout mutation that no longer encodes the synthesis of the original protein. A nucleotide modification mutation is designed to yield a protein with an altered activity. A multiple nucleotide insertion may, depending upon the specific sequences involved, encode a protein with an altered activity, or it may knock out an endogenous gene, replacing it with an exogenous gene.

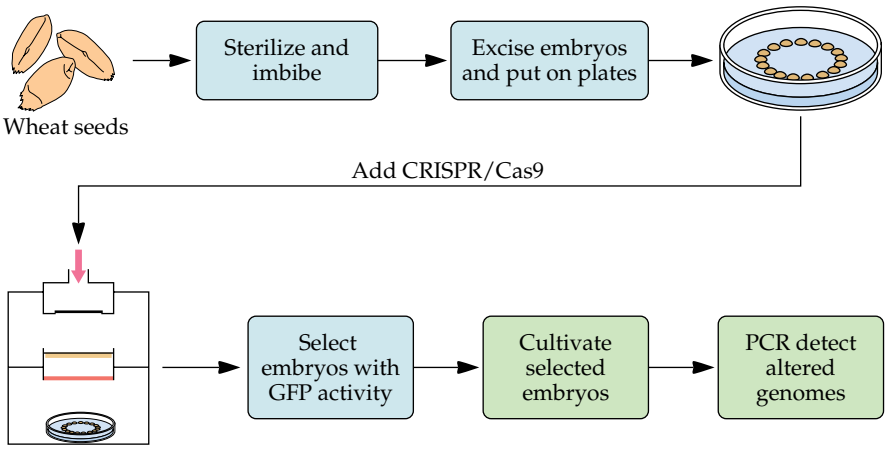


For the most part, sgRNA and Cas9 coding sequences have been introduced into plant genomes by either *Agrobacterium*-mediated transformation or by microprojectile bombardment of callus cultures followed by plant regeneration. This limits this technology to those cultivars that are amenable to tissue culture, significantly restricting the use of this approach. To circumvent this limitation, one group of researchers developed a microprojectile bombardment-based delivery system that enabled the components of the CRISPR/Cas9 system to be transiently expressed to edit genomes in wheat embryos (Fig. 10.19), thereby completely avoiding tissue culture and plant regeneration. In this case, gold particles coated with CRISPR/Cas9 components (and GFP as a marker gene) were delivered to wheat seed embryos. The bombarded embryos that exhibited transient GFP expression were subsequently grown into adult plants. Subsequent analysis of these plants by PCR for the target gene (i.e., targeted by the sgRNA) indicated that 5.2% of the selected embryos carried mutant (changed) alleles, with one-third of those plants inheriting the mutation in the next generation. Moreover, in no instances was the gene encoding the Cas9 enzyme integrated into the genomic DNA. It is suggested that this method should be useful in editing the genomes of plant species and cultivars that are otherwise recalcitrant to this procedure because they cannot be grown in culture and subsequently regenerated.

Clonal Seeds from Hybrid Rice

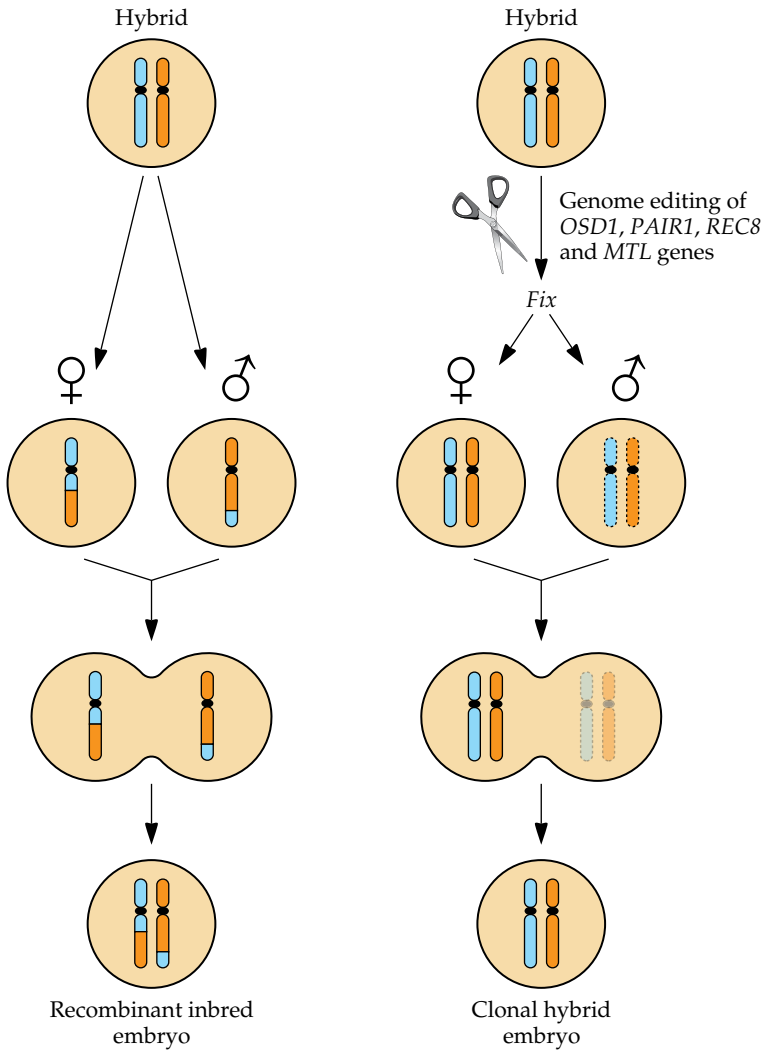
Hybrid plants are typically constructed from genetically different plant varieties in order to increase their vigor relative to their homozygous parents. Commercial hybrid plants are constructed to stably reproduce a valued attribute of each initial variety into the offspring. Typically, commercial hybrids are developed to improve traits such as disease resistance, plant size, flower color, fruit size, or taste. Unfortunately, hybrid seed production is a relatively slow process and is quite expensive for many crops. Generally, hybrids are produced by transferring the pollen from one plant to the pistil of another plant. To prevent plants from self-pollinating, all of the stamens (the pollen-producing organs) must be removed from the

Figure 10.19 Flow chart of the major steps involved in using microprojectile bombardment to introduce the CRISPR/Cas9 system into wheat plants, which are otherwise recalcitrant to genome editing.



plants that are going to be pollinated. Finding the desired hybrid may take years of constructing and testing genetic hybrids. In a recent effort to avoid the painstaking process of creating genetic hybrids with the desired traits, scientists, using the CRISPR/Cas9 system, simultaneously created mutants of several specific rice genes, including *REC8*, *PAIR1*, *OSD1*, and *MTL*. In this case, the genetically modified rice strains behaved identically to the initial wild-type strain. However, following the self-propagation of F_1 hybrids it was observed that the resultant plants could propagate clonally through seeds, a trait not usually associated with hybrid plants. Thus, the introduced genetic modifications had turned meiosis into a mitosis-like cell division (Fig. 10.20). Importantly, the genes that were targeted for

Figure 10.20 Fixation of rice heterozygosity by editing multiple genes. In normal plant sexual reproduction (shown on the left), recombinant inbred embryos are generated by the fusion of recombined haploid gametes. In the clonal reproduction strategy (shown on the right), meiosis is first turned into mitosis (by the combined action of the *OSD1*, *PAIR1*, and *REC8* genes = *FIX* genes) to produce clonal diploid gametes. Then, the genome of the male gamete is eliminated because the *MTL* gene has been knocked out. This results in the progeny of the self-pollinated plants being identical to the initial hybrid parent.



modification in rice are highly conserved in numerous other plants so that this method of producing stable hybrid plants should be applicable to a wide range of crops.

Conferring Male Sterility

Hybrid plant seed is typically produced from a genetic cross between two genetically different plant lines. The hybrid seeds usually result in larger, more vigorous plants (hence the term hybrid vigor). To facilitate the production of hybrid seeds, scientists often use plants that are cytoplasmically male sterile. Male sterile plants do not produce any functional pollen, although they do produce viable eggs. In addition, in jurisdictions where there is significant concern regarding the possible unintentional widespread dissemination of transgenic plants, the employment of male sterile transgenic plants that do not produce functional pollen (and hence do not spread their DNA) clearly obviates those concerns. Recently, a group of scientists reported using TALEN-mediated genome editing to knock out six (nearly identical) *CmDMC1* genes in chrysanthemum; this plant is a hexaploid species that contains six copies of the entire plant genome. The *CmDMC1* gene is involved in meiotic homologous recombination. Interestingly, the transgenic plants in which all six of these genes were knocked out exhibited both male and female sterility. Under the Cartagena Protocol on Biosafety, ornamentals must be unable to outcross with wild relatives prior to their approval for practical use. Thus, this manipulation obviates the difficulty of producing male sterile plants when the plants (such as chrysanthemum) are polyploid and may become a widely used approach to prevent the flow of transgenes from genetically engineered plants to their wild relatives.

Facilitating Protein Purification

Transgenic plants have some advantages over bacteria and animal cells as expression systems for foreign proteins. For example, transgenic bacteria and animal cells are typically grown in expensive bioreactors under precisely defined conditions and require highly skilled personnel to oversee the entire growth process. On the other hand, plants may be grown in the field or in greenhouses by somewhat less highly skilled individuals at a much lower cost and in more or less unlimited quantities. Thus, one of the main advantages of expressing a foreign protein in a transgenic plant is that it is relatively inexpensive to grow plants on a large scale.

For commercial production of a recombinant protein in any host organism, following growth, the organisms must be harvested and processed, and then the recombinant protein is purified. Protocols for the purification of proteins from transgenic bacteria, yeast, and animal cells are well established, and many proteins produced by bacteria and animal cells are currently in the marketplace. However, transgenic plants generally produce a much lower level of foreign proteins than bacteria and yeast; therefore, for commercial production of a range of protein products, the purification of target proteins produced by transgenic plants may require special strategies and approaches.

Oleosins

A novel way to facilitate recombinant protein purification from plants is to fuse the foreign protein to plant oleosins. Oleosins, or oil body proteins, are found in the seeds of a wide range of plants. These proteins are

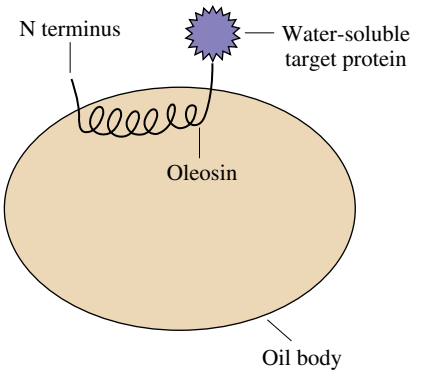


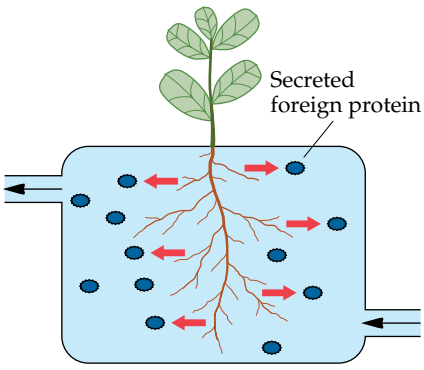
Figure 10.21 Schematic representation of a fusion protein including oleosin and a water-soluble target protein embedded in a plant seed oil body. The N- and C-terminal ends of the oleosin and the target protein are hydrophilic and are therefore found in the aqueous environment.

quite hydrophobic and are mostly found embedded within small oil droplets (0.6 to 2 mm in diameter) called plant oil bodies, thus stabilizing the oil bodies as discrete intracellular organelles for storing neutral lipids. However, the N- and C-terminal regions of oleosins are more hydrophilic than the rest of the protein and are exposed to the aqueous environment. It is therefore possible to engineer fusions between oleosins and water-soluble proteins at the DNA level (Fig. 10.21), with the expectation that the fusion protein will be targeted to plant oil bodies, making it relatively easy to purify. In this case, the water-soluble target protein will not be embedded in the oil body but, rather, will be exposed to the aqueous environment. Since it is relatively straightforward to purify oil bodies from plant seeds, the purification of the recombinant protein is simplified. A cleavable linker is included between the oleosin and the target protein so that the recombinant protein can be recovered by cleavage of the fusion protein. Expression in seeds is particularly attractive, as proteins accumulate stably and seeds can be stored without deterioration prior to being processed. This system has the potential to significantly lower the costs of purifying target proteins produced in plant seeds.

Rhizosecretion

Harvesting the variety of protein products that can be produced in transgenic plants can be difficult, since they are generally localized within plant cells that must be disrupted before their purification. Moreover, since the cost of purifying a recombinant protein can be as much as 90% of the total cost of producing that protein, if plants are going to be used as cost-effective bioreactors for protein production, it is essential that purification costs be kept to a minimum. One way around this problem is to engineer plants to secrete foreign proteins through the roots in a process that has been called "rhizosecretion." If a plant engineered for rhizosecretion is grown hydroponically, the protein will be secreted directly into the culture medium (Fig. 10.22).

Figure 10.22 Schematic representation of a plant in hydroponic culture secreting proteins and small molecules (red arrows) into the medium. The arrows at the inlet and outlet ports indicate the direction of flow of added nutrient solutions. The proteins secreted by the roots are concentrated and harvested from the hydroponic medium and then purified.



Normally, roots secrete large amounts of small molecules, such as sugars and amino acids; however, they secrete only low levels of relatively few proteins. The small organic molecules are first secreted to the root intercellular space (apoplast) before they are exuded/secreted by the roots. In one series of experiments, three different proteins—xylanase from the thermophilic anaerobic Gram-positive bacterium *Clostridium thermocellum*, GFP from the jellyfish *Aequorea victoria*, and human placental secreted alkaline phosphatase—were tested to determine whether they could be engineered for secretion through the roots of transgenic plants. The three proteins were directed to the root apoplast using three different secretion signals (Fig. 10.23). Each protein was efficiently exuded by the roots of transgenic tobacco plants, as long as the genetic construct contained a DNA sequence encoding a signal peptide, even one that was not of plant origin, placed upstream (at the 5′ end) of the gene whose protein was targeted for secretion. Both the 35S promoter, which is expressed in all plant cell types, and the *mas2′* promoter, which is preferentially expressed in roots, directed the synthesis of a significant amount of the target protein in root tissue. With the 35S promoter, the foreign protein could also be recovered from the guttation fluid (i.e., the leaf exudate). However, given the ease of collecting root as opposed to leaf exudate, this approach appears to have the most promise at the present time.

Protein Glycosylation

A large number of mammalian proteins, including many potentially therapeutic molecules, are glycosylated (i.e., they contain specific sugars attached to the hydroxyl group of either serine or threonine or the amide group of asparagine) (see chapter 3). While the addition of glycans (polysaccharides) containing high levels of mannose at specific asparagine residues on proteins is initially identical in mammalian and plant cells, subsequent trimming of the sugar residues generates complex N-glycans with very different structures and properties in the two different types of organisms. Although differences in glycosylation may not directly alter the activity of a protein, other properties, such as protein folding, stability, solubility, susceptibility to proteases, blood clearance rate, and antigenicity, can be affected profoundly. Thus, especially for monoclonal antibodies, glycosylation of the protein is considered to be a critical attribute of the final product.

To avoid some of the problems that result from the incorrect glycosylation of mammalian proteins that are produced in plant cells, it is possible to modify the plant so that it does not add "problematic" carbohydrate residues. For example, when the aquatic plant *Lemna minor* was engineered to produce light and heavy chains that are part of a human monoclonal antibody, the plant was simultaneously transformed with an RNA interference construct (see chapter 6) that specifically inhibited the expression of the plant glycan modification enzymes α-1,3-fucosyltransferase and β-1,2-xylosyltransferase. The antibodies that were produced by these transgenic plants contained a single major N-glycan species without any plant-specific N-glycan residues (Fig. 10.24A). While these truncated glycan side chains are not identical to the glycans that are produced when a human monoclonal antibody is synthesized in Chinese hamster ovary (CHO) cells in culture, they share the same core molecules. In fact, when human monoclonal antibodies produced in genetically modified plants were tested, they displayed better antibody-dependent cell-mediated cytotoxicity and effector cell receptor-binding activities than antibodies that were produced in CHO cells. Thus, with this system, it may be possible to produce a wide range of mammalian proteins in which the glycan side chains no longer limit the use of the final protein. In addition, it should be much less expensive to produce mammalian proteins in transgenic plants than in mammalian cell culture.

To further improve this system, the gene encoding the enzyme endo-*N*-acetyl-β-D-glucosaminidase (endo T) from the fungus *Hypocrea jecorina* was inserted into the genome of the plant *Arabidopsis thaliana* under the transcriptional control of a seed-specific promoter from *Phaseolus vulgaris* (common bean), and the enzyme was localized to the Golgi (of seed cells), where many proteins are glycosylated. The *A. thaliana* strain that was used was a mutant that lacked the enzyme *N*-acetylglucosaminyltransferase-I; this resulted in the accumulation of proteins with high mannose glycan side chains that are able to act as substrates for the enzyme endo T. Endo T hydrolyzed the bonds between adjacent *N*-acetylglucosamine residues in the glycans of some proteins, leaving only the innermost *N*-acetylglucosamine residue (Fig. 10.24B). This system, which has been called GlycoDelete, yields proteins with only a single *N*-acetylglucosamine residue attached to the amide group of asparagine. This reduces any difficulties in purifying the target protein that might otherwise arise as a consequence of heterogeneous glycosylation (with differently glycosylated proteins having somewhat different physical and biochemical properties)

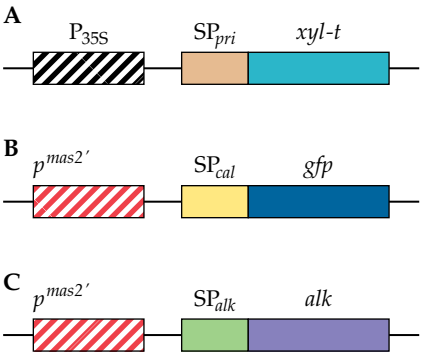


Figure 10.23 DNA constructs of proteins engineered for secretion into the apoplast in tobacco. **(A)** A truncated xylanase (*xyl-t*) gene from *C. thermocellum* is transcriptionally controlled by the cauliflower mosaic virus 35S promoter (P_{35S}) and targeted to the apoplast by a tobacco proteinase inhibitor signal peptide (SP_{pri}). **(B)** A GFP (*gfp*) gene from *A. victoria* controlled by the strong modified *Agrobacterium* mannopine synthase promoter ($p^{mas2'}$) that is preferentially active in plant roots and targeted to the apoplast by a *Nicotiana plumbaginifolia* (a type of tobacco) calreticulin proteinase inhibitor signal peptide (SP_{cal}). **(C)** A truncated human alkaline phosphatase (*alk*) gene controlled by the strong modified *Agrobacterium* mannopine synthase promoter ($p^{mas2'}$) that is preferentially active in plant roots and targeted to the apoplast by the human alkaline phosphatase signal peptide (SP_{alk}).

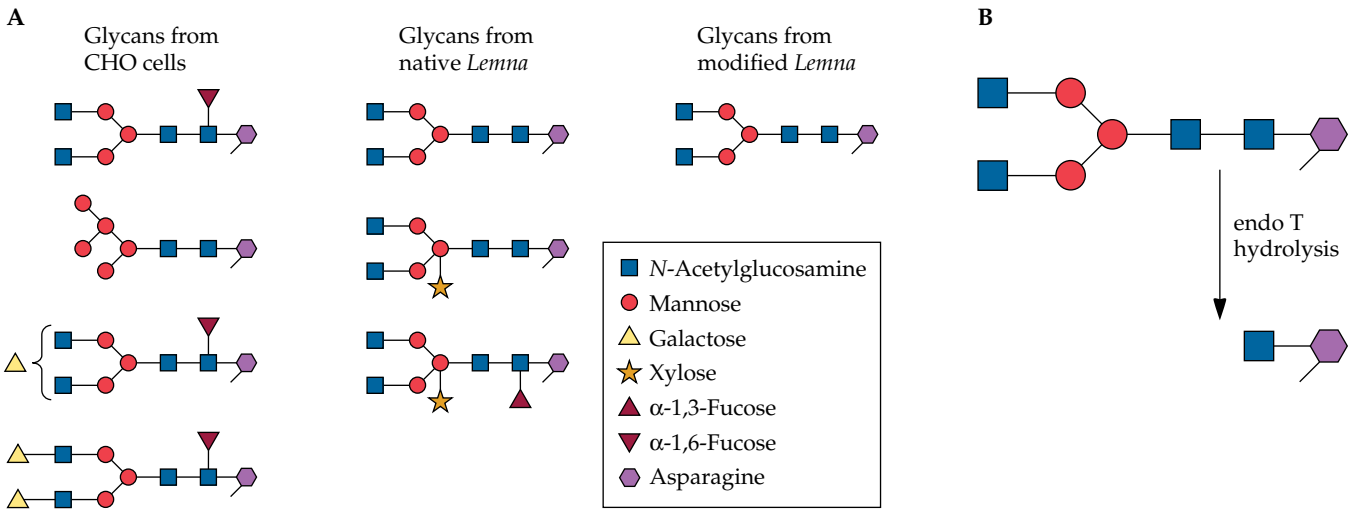


Figure 10.24 **(A)** Structures of the N-glycan side chains found on proteins synthesized by CHO cells, the native form of the aquatic plant *L. minor*, or a genetically modified form of *L. minor*. **(B)** Digestion of N-glycan protein side chains with endo T to yield protein side chains with only a single *N*-acetylglucosamine residue.

and does not interfere with the functioning of the target proteins. Moreover, the developers of this technology argue that plant-produced pharmaceuticals that are synthesized in seeds could be stored in a stable manner as part of the seed until the protein product is needed.

The technology for producing uniform Fc portions of monoclonal antibodies (which are all glycosylated) currently exists for a number of different plants. This enables enhanced Fc effector functions, greater consistency between batches, decreased risks related to the immunogenicity of the therapeutic monoclonal antibody, and the possibility for economically manufacturing monoclonal antibodies in plants. This approach should facilitate and significantly reduce the cost of monoclonal antibody production. In fact, glycoengineered monoclonal antibodies produced in plants have recently begun clinical trials.

Gene Stacking

In order to develop transgenic plants that have increased commercial value, it is often necessary to introduce more than a single foreign gene into plants. Currently, the two most commonly introduced traits in commercially available transgenic plants are insect resistance and herbicide tolerance, each of which may be encoded in a single gene (see chapter 11). To produce transgenic plants that are both insect resistant and herbicide tolerant requires plants to contain at least two transgenes. This stacked transgenic plant enables farmers to have a crop that is simultaneously insect resistant and herbicide tolerant. In fact, there are numerous commercial examples of transgenic plants that have stacked genes. In the future, crops that contain multiple transgenes are expected to become quite common. There are several methods that can be used to produce stacked transgenic plants (Table 10.8). These methods include hybrid stacking, cotransformation, multigene cassette transformation, and retransformation.

Table 10.8 Some common gene stacking methods used in the production of commercial transgenic plants

Gene stacking method	Details
Hybrid stacking	A transgenic plant containing one or more transgenes is cross-hybridized with another plant of the same genus and species containing other transgenes. Selection for a plant with all of the introduced traits follows multiple iterative hybridizations.
Cotransformation	A plant is simultaneously transformed with two or more independent transgenes (i.e., separate gene constructs).
Multigene cassette transformation	A plant is transformed with a single gene construct that contains two or more linked transgenes.
Retransformation	A plant that contains a particular transgene is genetically transformed with a DNA construct(s) containing other transgene(s)

With expanding populations worldwide, there has been a substantial loss of arable land necessary to feed this increasing population. One possible solution to this problem is to grow more food in urban environments. However, this approach is fraught with problems. Nevertheless, by judicious modification of the genomes of some plants, it may be possible to grow them to a smaller size than is normally the case on much smaller plots of land without decreasing the final crop yield. Recently, a group of scientists reported using a gene stacking strategy to generate and combine mutations in cherry tomatoes for condensed shoots, rapid flowering, and precocious growth termination. These mutations were all produced using CRISPR/Cas9 genome editing and produced tomatoes plants that were compact and early-yielding plants that were deemed suitable for urban agriculture. The engineered plants were then grown in hydroponic vertical farms and compared to wild-type plants grown in the same way. Although the yield of fruits per plant on the smaller genetically manipulated plants was slightly lower than with larger wild-type plants, this reduction in yield was readily compensated for by the fact that more, smaller genetically manipulated plants could be grown in the same space than the larger wild-type plants. Moreover, editing the same set of three genes in groundcherry led to a similar result, suggesting that this approach may be effective with a number of different crops.

CRISPR-Based Directed Evolution

The process of directed evolution involves introducing random mutations into a target gene in a host organism and then screening for proteins with potentially superior performance. In prokaryotes, directed evolution is often achieved by error-prone PCR of a target gene followed by extensive screening and testing (see chapter 3). However, this technique, which is highly effective in prokaryotes, is impractical in plants. Therefore, some researchers have developed the ability to produce random mutations in targeted genes in plants using CRISPR/Cas9. In this instance, the Cas9 endonuclease generates double-strand breaks within a region of the target gene defined by the sgRNA, and these breaks are then repaired by nonhomologous end joining, a procedure that generates small insertions, small deletions, or base substitutions. This procedure may also be performed using a nickase Cas9 enzyme that cuts only one of the two target DNA strands. In addition to generating random mutations in one portion of a gene, it is possible to simultaneously

target two or more regions of the same gene using two or more sgRNAs. CRISPR-based directed evolution is a technique that is still in its infancy; however, it is currently being used in attempts to alter a wide variety of traits in several different plants in a rapid and simple manner.

Polycistronic Gene Expression

Plants and other eukaryotes are monocistronic, where each gene is under the transcriptional control of a separate promoter. This makes it extremely difficult for scientists to genetically transform plants with several genes that are part of the same pathway and ensure that these genes are transcribed and translated to the same extent (e.g., as part of a metabolic pathway). For example, multiple rounds of genetic transformation to introduce several genes (even if those genes contain the same transcriptional promoter) generally result in those genes being inserted in different portions of the plant chromosomal DNA and therefore expressed at differing levels due to the influence of the epigenetic environment. However, scientists have recently developed systems that can get around this roadblock. Using synthetic biology, where the desired genes are synthesized and then logically assembled (see chapter 2), researchers were able to assemble multiple transcription units in a single DNA fragment, each with its own promoter, transcription terminator, and signal peptide-coding sequence for secretion of the produced protein (Fig. 10.25A). Those transcription units are separated by short linker sequences of nucleotides encoding 18 to 22 amino acids, known as 2A peptides, which are self-cleaving peptide sequences (Fig. 10.25B). These 2A sequences (peptides) were first discovered in the eukaryotic foot-and-mouth disease virus and function by inducing ribosomal skipping during protein translation between Pro and Gly in the C-terminal end of the peptide, where the amino acid sequence of the 2A peptide from foot-and mouth disease is Val-Lys-Gln-Thr-Leu-Asn-Phe-Asp-Leu-Leu-Lys-Leu-Ala-Gly-Asp-Val-Glu-Ser-Asp-Pro-Gly-Pro. Similar self-cleaving sequences have been found in a wide range of viral families, with different 2A peptides having different efficiencies of self-cleavage. With this system, comparable levels of transgene expression can be expected regardless of the gene positions in the polycistronic expression cassette. It now remains to be determined whether this system is a useful and effective means of introducing entire functional metabolic pathways into plants.

Production of Marker-Free Transgenic Plants

Usually, at the same time that a foreign gene is introduced into plants, a selectable marker gene is also introduced. Although none of these marker genes or their products have to date been shown to have an adverse effect on human, animal, or environmental safety, their inclusion in transgenic plants has raised some concerns. For example, it is possible that the products of some marker genes might be either toxic or allergenic. Also, the antibiotic resistance genes that are often used as selectable markers might be transferred to pathogenic soil microorganisms. While gene transfer from plants to environmental or medically related bacteria, or from plant products consumed as food to intestinal microorganisms or human cells, is generally considered to be highly unlikely (see chapter 13), there are nevertheless some inherent risks that have not been totally addressed, and therefore both regulatory and public concerns remain in many places in

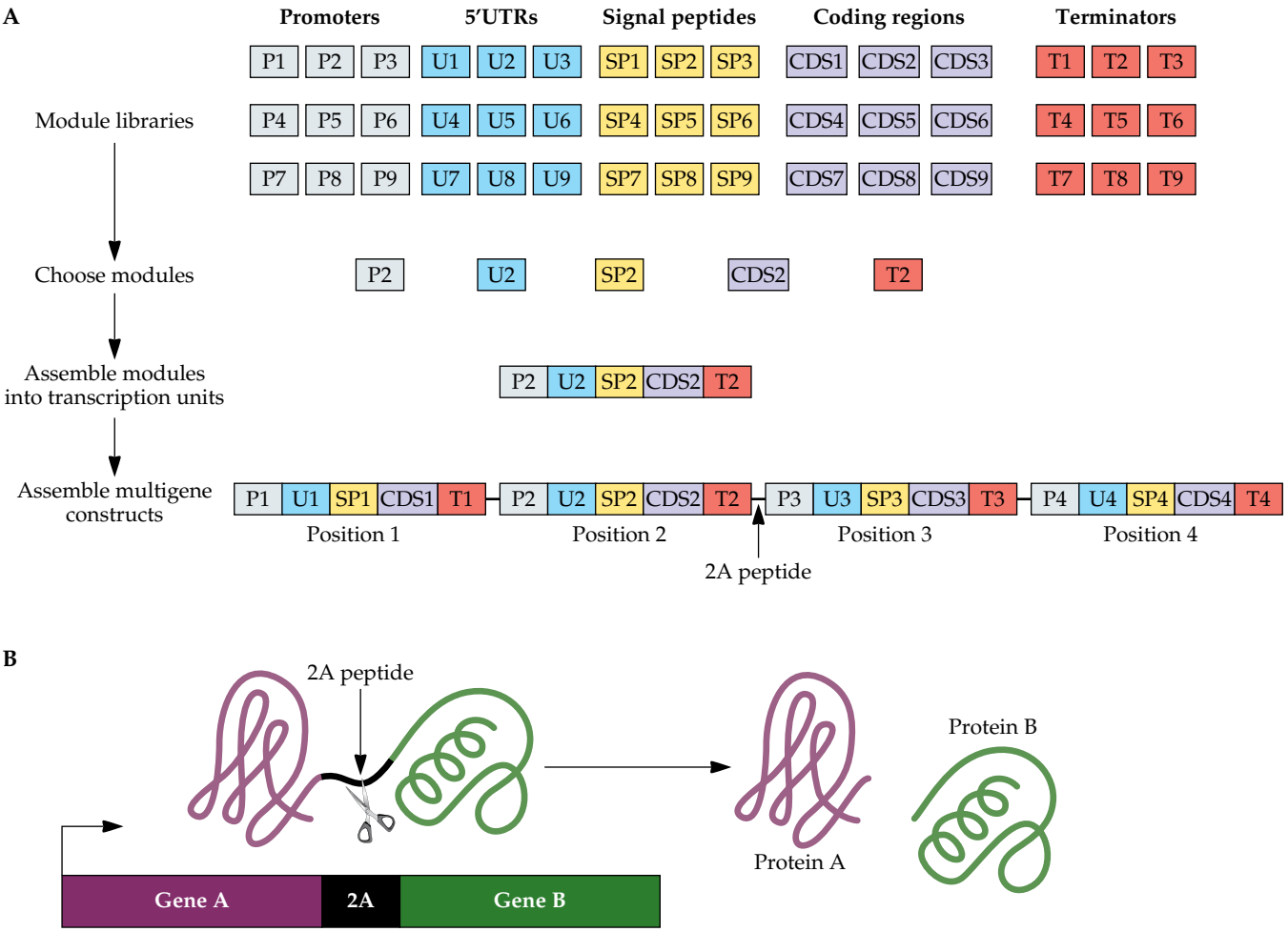


Figure 10.25 **(A)** Schematic overview of the system used to synthesize polycistronic genes in plants. To start, libraries of various gene components are assembled. These include DNA fragments that encode promoters, 5′ untranslated regions (5′ UTRs), signal peptides, protein-coding regions (CDS), and transcription terminators. From these module libraries, researchers can choose the modules that will make up each transcription unit. Secreted protein units will, while nonsecreted proteins will not, contain a signal peptide module. Assembled transcription units are each separated by a nucleotide sequence encoding the self-cleaving 2A peptide. **(B)** Schematic representation of the expression and self-cleavage of two proteins separated by a 2A peptide.

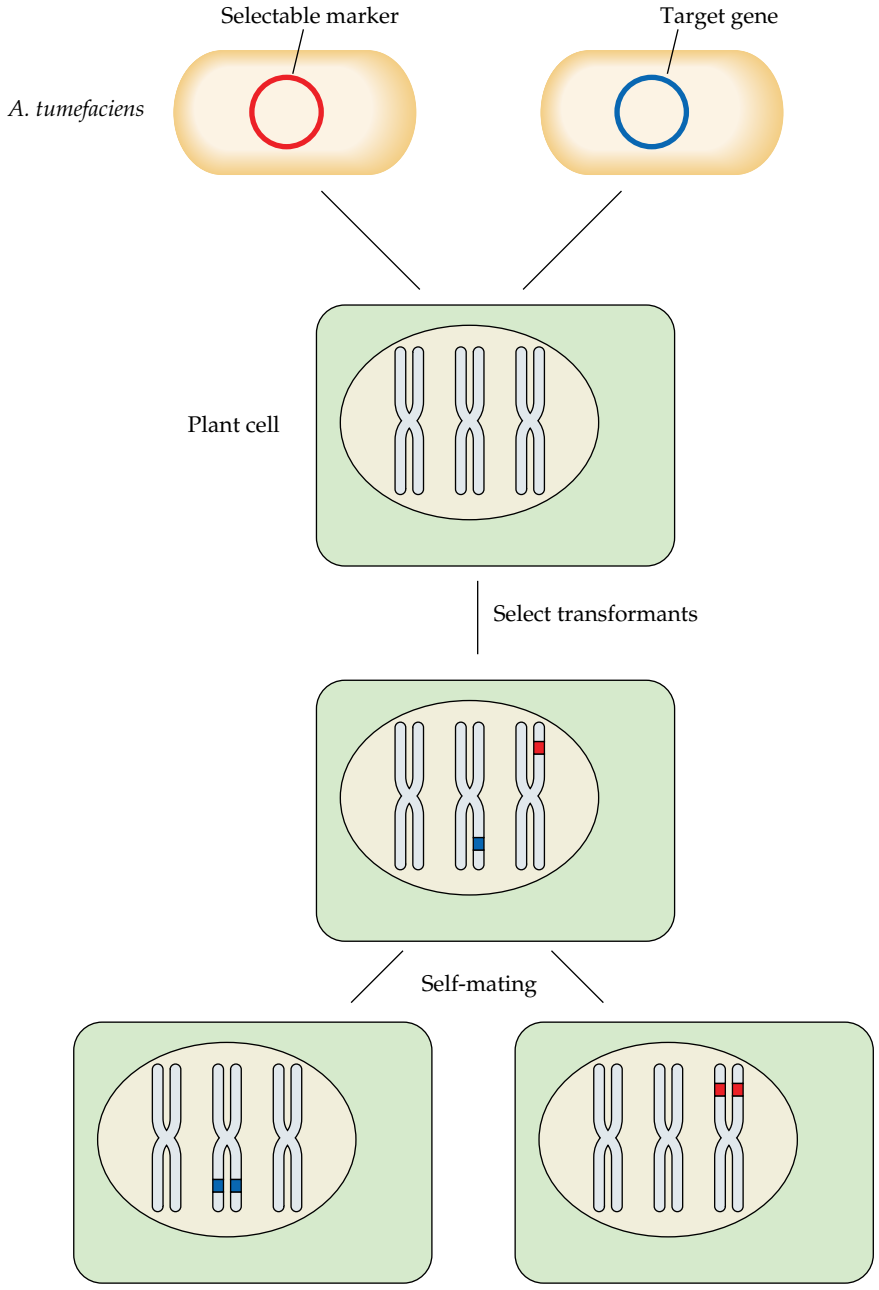
the world. Moreover, the presence of a selectable marker makes it technically more difficult to transform a transgenic plant with additional foreign genes, since the same selectable marker cannot be used more than once. To allay these concerns, strategies for the production of transgenic plants without any marker genes have been developed.

Removing Marker Genes from Nuclear DNA

One experimental approach that has been used to produce markerless transgenic plants includes cotransformation of plants with two separate DNAs, one carrying the marker gene and the other carrying the target foreign gene. With cotransformation, approximately 30 to 80% of the

transformed plants will contain both genes. However, since the two genes are integrated at different sites on the chromosomal DNA, provided that they are integrated at sites that are not too close to one another, traditional breeding techniques can be used to rid the transgenic plant of the selectable marker (Fig. 10.26). Thus, the two genes are separated by chromosome

Figure 10.26 Schematic representation of cotransformation of a plant cell with two separate DNAs. One DNA carries a selectable marker, while the other carries the target gene. On average, more than half of the transformed plants have both of these genes, but at separate sites on the chromosomes. Several rounds of self-mating allow the segregation of the marker and target genes into different plants.

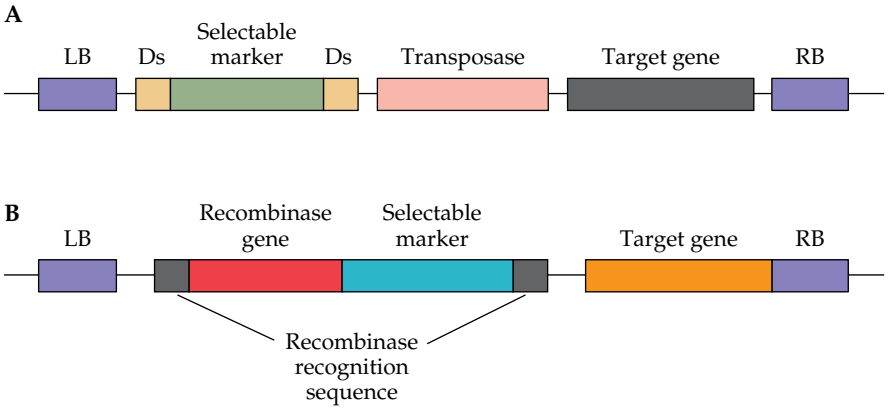


segregation during a few rounds of matings. In this way, the selectable marker is removed from the plant genome during the segregation and recombination that occur during sexual reproduction. At present, this strategy is the most widely used approach to produce marker-free transgenic plants.

As an alternative to cotransformation with two separate DNAs, a selectable marker gene may be cloned between two plant transposable elements (Ds elements) and inserted into the T-DNA along with the target gene and a transposase gene that, after plant transformation, excises the DNA between the Ds elements and integrates it at another chromosomal site (Fig. 10.27A). Any sequence that is between two Ds elements can be mobilized to a new location in the genome—excising the gene from the original location—provided that the appropriate transposase enzyme is available. Following insertion of the T-DNA into the host plant DNA, about 90% of the time, the selectable marker that is between the two Ds elements will be moved to another site on the chromosomal DNA. About half of the time, the new location for the selectable marker will be far away from the original location. Thus, a selectable marker gene can be used to identify transformed plant cells, and subsequently, it can be removed by breeding.

Any procedure that uses sexual crossing to segregate the selectable marker from the gene of interest cannot be applied to woody plants (because of their very long generation times), vegetatively propagated plants (where new plants grow from parts of the parent plant rather than from seeds), or sterile plants. One simple way around this problem is to flank the selectable marker with specific DNA sequences that mark it for eventual removal from the genome (Fig. 10.27B). In this way, the selectable

Figure 10.27 (A) Schematic representation of a T-DNA-based selectable marker gene excision system. Following integration of the T-DNA into the plant chromosomal DNA, the transposase can excise the selectable marker gene and insert it into a different chromosomal location. LB, the T-DNA left border; RB, the T-DNA right border. The promoter and transcription termination sequences of the genes are not shown. **(B)** Schematic representation of a DNA construct for generating marker-free transgenic plants. RB and LB are the right and left borders of the T-DNA, respectively. After transformation, selection of transformed cells, and growth in culture, recombinase cleaves the DNA at the recombination sites, and the selectable marker and recombinase genes are lost. Transgenic plants are formed from the cells without the selectable marker gene.



marker and flanking sequences are only transiently present as a component of the transgenic plant. For example, if the selectable marker gene is fused to a recombinase gene and flanked by sequences that are recognized by the recombinase, both from the yeast *Zygosaccharomyces rouxii*, the selectable marker gene will be initially expressed. However, the recombinase is also expressed and mediates excision of the region between its recognition sites, thus removing both the selectable marker and recombinase genes after growth in culture for several months. In this case, transformed plant cells are selected immediately following transformation, characterized to ascertain whether they contain and express the target gene, and then grown in culture until the selectable marker has been spontaneously excised. In addition to the system outlined here, several similar excision systems that result in the removal of the selectable marker have been developed. Importantly, regardless of the details, it is no longer necessary to produce transgenic plants that retain antibiotic resistance genes or other undesirable selectable markers.

In the selection of genetically modified plants following genome editing by the CRISPR/Cas9 methodology, it is necessary to distinguish modified from unmodified plants. This is a straightforward procedure if the modified plants carry a selectable marker. Procedures similar to those described above may be employed to remove the selectable markers from both transgenic plants and CRISPR/Cas-edited plants. Unfortunately, many of the procedures that have been developed to eliminate the selectable markers (typically antibiotic- or herbicide-resistant genes) in transgenic plants require growing the plants for two or more generations, where a generation is typically 4 to 12 months, depending on the plant. To shorten the time that is required to select marker-free plants following genome modification by CRISPR/Cas, one group of researchers used a gene encoding *Discosoma* (a genus of anemone) red fluorescent protein, i.e., DsRED, as a selectable marker (Fig. 10.28). Following CRISPR/Cas modification of rice, tomato, and *Arabidopsis* plants, DsRED fluorescence was readily and rapidly detected in plant seeds. With this fluorescent marker, transgene-free homozygous genome-edited crops could be detected within a single generation following *in vitro* transformation.

Any system used to remove the selectable marker generally requires growing the transformed plants for at least one and up to several generations and then testing to ensure that the selectable marker has been removed. To speed up this somewhat tedious process, one group of researchers did not include a selectable marker as part of their T-DNA construct but instead included a bacterial β-glucuronidase reporter gene, a gene that is not found in plants. The enzyme β-glucuronidase (often called GUS) can cleave the colorless substrate 5-bromo-4-chloro-3-indolyl glucuronide (X-Gluc; Fig. 10.29), a compound that turns blue at the site of GUS activity in the plant. Leaves or other tissues from putative transgenic plants may be easily tested to see whether they contain GUS activity, since upon cleavage, the colorless substrate turns an insoluble blue color that results from the synthesis of 5,5′-dibromo-4,4′-dichloro-indigo. Thus, putative transgenic cotton plants were identified by an *in situ* GUS histochemical assay of pollen grains, immature embryos, and tissues from putative first-generation transgenic plants, with 0.2 to 0.7% of the plants tested being transgenic.

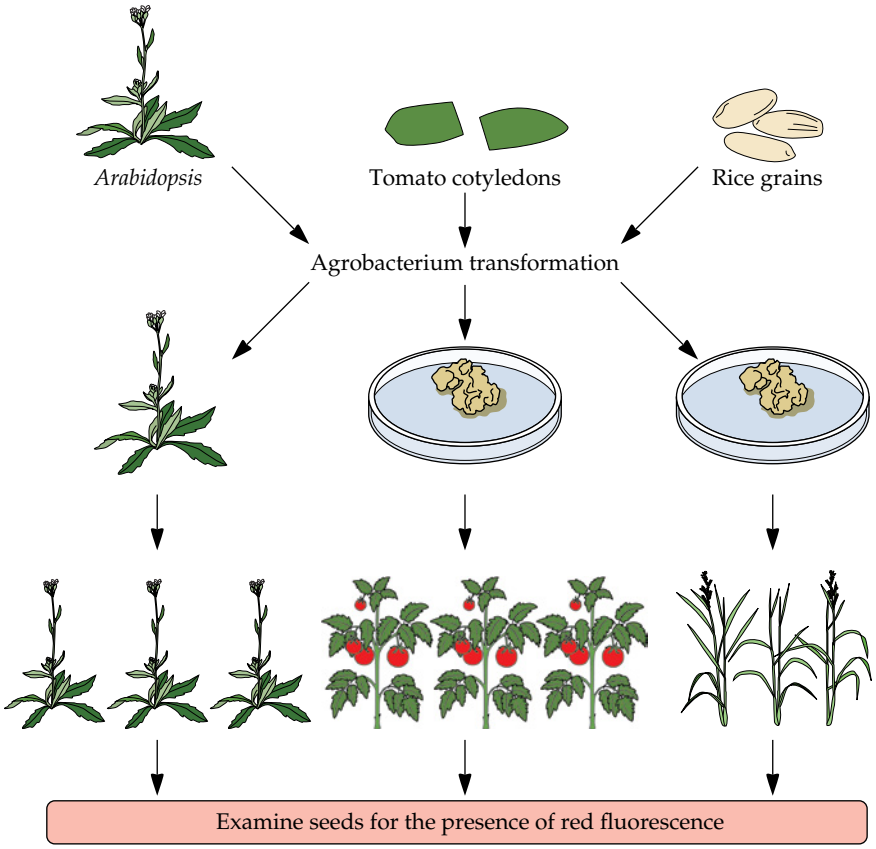
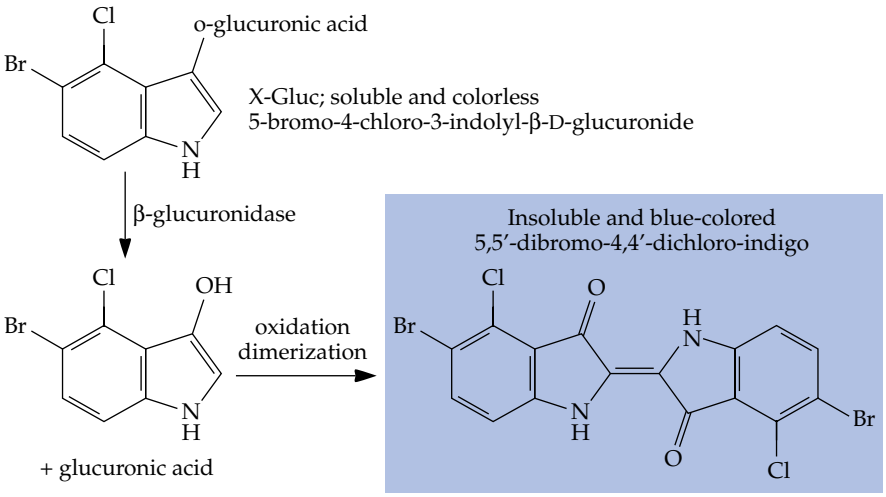


Figure 10.28 Schematic representation of plant transformation with the CRISPR/Cas system and selectable marker (DsRED gene). *Arabidopsis thaliana* transformation was done by dipping plant flowers while tomato and rice plants were transformed *in vitro* (in tissue culture).

Figure 10.29 The substrate 5-bromo-4-chloro-3-indolyl-β-D-glucuronide (X-Gluc) is cleaved by the enzyme β-glucuronidase to produce glucuronic acid and chloro-bromoindigo. The chloro-bromoindigo is subsequently air oxidized and then dimerizes to produce the insoluble blue precipitate 5,5′-dibromo-4,4′-dichloro-indigo.



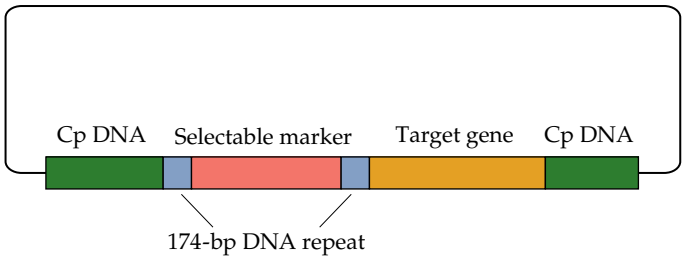


Figure 10.30 Schematic representation of a selectable marker gene excision system for transformed chloroplasts. Following introduction into chloroplasts, the chloroplast (Cp) DNA sequences on the plasmid direct the integration of the foreign gene into the chloroplast genome by homologous recombination. Transformants are selected by their resistance to spectinomycin. Then, transformants are grown without antibiotics, and the selectable marker gene is lost as a result of homologous recombination between the 174-bp repeat elements. The promoters, transcription termination regions, and ribosome-binding sites of both genes are not shown.

Removing Marker Genes from Chloroplast DNA

The possibility of high levels of foreign-gene expression, as well as expression of several related genes under the control of the same promoter (e.g., encoding a biochemical pathway), has resulted in an increasing number of foreign genes being introduced into plant chloroplast DNA. The foreign genes are typically introduced as part of a genetic construct that includes a selectable bacterial gene, e.g., *aadA*, which confers resistance to the antibiotics spectinomycin and streptomycin. For removal of the marker gene, it is flanked by directly repeating DNA sequences (174 bp). Following growth of the transformed cells in the absence of selective pressure, the selectable gene will be excised by homologous recombination between the direct repeat sequences (Fig. 10.30).

An alternative approach is to develop selectable markers that do not encode antibiotic resistance or any other undesirable trait. One possible selectable marker of this type is the spinach gene for the enzyme betaine aldehyde dehydrogenase. This enzyme, which is present in the chloroplasts of a limited number of plants, converts the toxic betaine aldehyde to the nontoxic glycine betaine. In addition, glycine betaine acts as an osmoprotectant, conferring some measure of salt or drought resistance. When resistance to betaine aldehyde was used to select transgenic plants, the transformation efficiency was approximately 25-fold higher than when transformants were selected with spectinomycin. The use of this approach should facilitate the chloroplast transformation of many important crops, including cereals that are naturally resistant to spectinomycin.

summary

Plants are genetically engineered by (i) introducing a gene into plant cells, (ii) selecting transformed cells, and then, if necessary, (iii) regenerating a fertile plant. Strains of the soil bacterium *A. tumefaciens* can genetically engineer plants naturally. In this system, after responding to chemical signals from a surface wound, *A. tumefaciens* makes contact with an exposed plant cell membrane. A series of steps then occurs that results in the transfer of a segment (T-DNA) of a plasmid (Ti plasmid) from the bacterium into the nucleus of the plant cell. The T-DNA region becomes integrated into the plant genome, and subsequently, the genes on the T-DNA region are expressed. The T-DNA region of natural Ti plasmids contains genes that encode enzymes for the production of phytohormones. These compounds cause the plant cells to enlarge and proliferate. Also, the plant cell becomes a factory for the production of an opine that is encoded by a T-DNA gene and can be catabolized only by *A. tumefaciens* with a specific Ti plasmid. Thus, *A. tumefaciens* has evolved a mechanism that converts a plant cell into a production center for a carbon and nitrogen source (opine) for its exclusive use.

The *A. tumefaciens*–Ti plasmid system has been modified for use as a delivery mechanism for cloned genes to some plant cells. In these vector systems, the phytohormone and opine metabolism genes have been removed from the T-DNA region, and the modified T-DNA sequence has been cloned into a plasmid that can exist stably in *E. coli*. A cloned gene that is inserted into this T-DNA region is part of the DNA that is transferred into the nucleus of a recipient plant cell. To achieve this transfer, *A. tumefaciens* is used as a delivery system. In one system, the shuttle vector with the T-DNA-cloned gene segment is introduced into an *A. tumefaciens* strain that carries a compatible plasmid with genes that are essential for transferring a T-DNA region into a plant cell (*vir* genes). In addition to this binary vector system, a cointegrate system has been designed so that, after the introduction of the shuttle vector carrying the target gene into *A. tumefaciens*, it recombines with the *vir* gene-containing, disarmed Ti plasmid to yield a single plasmid that has both *vir* gene functions and the T-DNA-cloned gene segment.

The *A. tumefaciens* T-DNA region has been used to produce a large number of transgenic plants. Unfortunately, this system is not effective with all plants. However, transformation of plants with Ti plasmids that lower plant ethylene levels can increase the transformation frequency of some plants. In addition, microprojectile bombardment (biolistics) has been an effective procedure for delivering DNA to a wider range of plant cells. Moreover, foreign DNA up to 150 kb in size that is cloned in YACs can be transferred to plant cells using a biolistic procedure. This transferred DNA can be stably integrated into the genome of the plant cells. Moreover, a two-step procedure was developed in which *Bbm* and/or *Wus2* genes are first introduced into plants by microprojectile bombardment, and then immature embryos from plants carrying these genes are readily transformed with *Agrobacteria* carrying a Ti plasmid vector.

Methods for the insertion of foreign genes directly into the chloroplast DNA have been developed. The large number of copies of chloroplast DNA in each plant cell means that very high levels of foreign gene expression are attainable when the introduced gene is present as part of the chloroplast DNA. Using the CRISPR/Cas9 system, protocols to make targeted changes to existing genes within the genomic DNA of plant cells have been developed. To obtain high-level expression of foreign proteins in plants, it is possible to transiently express foreign genes in plants in large quantities. To facilitate the purification of foreign proteins that are produced in plants, one can fuse a target gene to an oleosin gene so that the fusion protein that is produced is localized in seed oil bodies. A technique called rhizosecretion can be employed to secrete the foreign protein along with other root exudates. By judicious manipulation, several related proteins (e.g., as part of a metabolic pathway) may be synthesized from a polycistronic operon. To increase the usefulness of plants as production systems for mammalian proteins, it is possible to modify the glycosylation patterns of foreign proteins synthesized in plants to avoid problematic carbohydrate moieties. Finally, experimental protocols that remove marker genes from transgenic plants have been developed.

REFERENCES

Aliaga-Franco N, **Zhang C**, **Presa S**, **Srivastava AK**, **Granell A**, **Alabadí D**, **Sadanandom A**, **Blázquez MA**, **Minguet EG.** 2019. Identification and transgene-free CRISPR-edited plants of rice, tomato and *Arabidopsis* by monitoring DsRED fluorescence in dry seeds. *Front Plant Sci* **10:**1150.

Borisjuk NV, **Borisjuk LG**, **Logendra S**, **Petersen F**, **Gleba Y**, **Raskin I.** 1999. Production of recombinant proteins in plant root exudates. *Nat Biotechnol* **17:**466–469.

Bortesi L, **Fischer R.** 2015. The CRISPR/Cas9 system for plant genome editing and beyond. *Biotechnol Adv* **33:**41–52.

Butt H, **Zaidi SS**, **Hassan N**, **Mahfouz M.** 2020. CRISPR-based directed evolution for crop improvement. *Trends Biotechnol* **38:**236–240.

Callaway E. 2018. EU law deals blow to CRISPR crops. *Nature* **560:**16.

Capuano F, **Beaudoin F**, **Napier JA**, **Shewry PR.** 2007. Properties and exploitation of oleosins. *Biotechnol Adv* **25:**203–206.

Ceccon CC, **Caverzan A**, **Magris R**, **Salvadori JR**, **Grando MF.** 2020. Gene stacking as a strategy to confer characteristcs of agronomic importance in plants by genetic engineering. *Cienc Rural* **50:**e20190207.

Chen Y, **Lange A**, **Vaghchhipawala Z**, **Ye X**, **Saltarikos A.** 2020. Direct germline transformation of cotton meristem explants with no selection. *Front Plant Sci* **11:**575283.

Chiu W, **Niwa Y**, **Zeng W**, **Hirano T**, **Kobayashi H**, **Sheen J.** 1996. Engineered GFP as a vital reporter in plants. *Curr Biol* **6:**325–330.

Christou P. 1992. Genetic transformation of crop plants using microprojectile bombardment. *Plant J* **2:**275–281.

Cox KM, **Sterling JD**, **Regan JT**, **Gasdaska JR**, **Frantz KK**, **Peele CG**, **Black A**, **Passmore D**, **Moldovan-Loomis C**, **Srinivasan M**, **Cuison S**, **Cardarelli PM**, **Dickey LF.** 2006. Glycan optimization of a human monoclonal antibody in the aquatic plant *Lemna minor*. *Nat Biotechnol* **24:**1591–1597.

Dale EC, **Ow DW.** 1991. Gene transfer with subsequent removal of the selection gene from the host genome. *Proc Natl Acad Sci USA* **88:**10558–10562.

Daniell H, **Wiebe PO**, **Millan AF.** 2001. Antibiotic-free chloroplast genetic engineering: an environmentally friendly approach. *Trends Plant Sci* **6:**237–239.

Daniell H, **Lin C-S**, **Yu M**, **Chang W-J.** 2016. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biol* **17:**134.

Darbani B, **Eimanifar A**, **Stewart CN Jr**, **Camargo WN.** 2007. Methods to produce marker-free transgenic plants. *Biotechnol J* **2:**83–90.

Debernardi JM, **Tricoli DM**, **Ercoli MF**, **Hayta S**, **Ronald P**, **Palatnik JF**, **Dubcovsky J.** 2020. A GRF-GIF chimeric protein improves the regeneration efficiency of transgenic plants. *Nat Biotechnol* **38:**1274–1279.

De Greve H, **Leemans J**, **Hernalsteens JP**, **Thia-Toong L**, **De Beuckeleer M**, **Willmitzer L**, **Otten L**, **Van Montagu M**, **Schell J.** 1982. Regeneration of normal and fertile plants that express octopine synthase from tobacco crown galls after deletion of tumour-controlling functions. *Nature* **300:**752–755.

Diamos AG, **Hunter JGL**, **Pardhe MD**, **Rosenthal SH**, **Sun H**, **Foster BC**, **DiPalma MP**, **Chen Q**, **Mason HS.** 2020. High level production of monoclonal antibodies using an optimized plant expression system. *Front Bioeng Biotechnol* **7:**472.

Dubey KK, **Luke GA**, **Knox C**, **Kumar P**, **Pletschke BI**, **Singh PK**, **Shukla P.** 2018. Vaccine and antibody production in plants: developments and computational tools. *Brief Funct Genomics* **17:**295–307.

El-Mounadi K, **Morales-Floriano ML**, **Garcia-Ruiz H.** 2020. Principles, applications, and biosafety of plant genome editing using CRISPR-Cas9. *Front Plant Sci* **11:**56.

Gao Y, **Zhao Y.** 2014. Specific and heritable gene editing in *Arabidopsis*. *Proc Natl Acad Sci USA* **111:**4357–4358.

Gelvin SB. 2017. Integration of *Agrobacterium* T-DNA into the plant genome. *Annu Rev Genet* **51:**195–217.

Gleba Y, **Klimyuk V**, **Marillonnet S.** 2007. Viral vectors for the expression of proteins in plants. *Curr Opin Biotechnol* **18:**134–141.

Goldsbrough AP, **Lastrella CN**, **Yoder JI.** 1993. Transposition mediated re-positioning and subsequent elimination of marker genes from transgenic tomato. *Bio/Technology* **11:** 1286–1292.

Gould F, **Amasino RM**, **Brossard D**, **Buell CR**, **Dixon RA**, **Falck-Zepeda JB**, **Gallo MA**, **Giller K**, **Glenna L**, **Griffin TS**, **Hamaker BR**, **Kareiva PM**, **Magraw D**, **Mallory-Smith C**, **Pixley K**, **Ransom EP**, **Rodemeyer M**, **Stelly DM**, **Stewart CN**, **Whitaker RJ.** 2016. *Genetically Engineered Crops: Experiences and Prospects*. National Academies Press, Washington, DC.

Gruber MY, **Crosby WL.** 1993. Vectors for plant transformation, p 89–119. *In* Glick BR, Thompson JE (ed), *Methods in Plant Molecular Biology and Biotechnology*. CRC Press, Boca Raton, FL.

Hadi MZ, **McMullen MD**, **Finer JJ.** 1996. Transformation of 12 different plasmids into soybean via particle bombardment. *Plant Cell Rep* **15:**500–505.

Hager M, **Bock R.** 2000. Enslaved bacteria as new hope for plant biotechnologists. *Appl Microbiol Biotechnol* **54:**302–310.

Hamada H, **Liu Y**, **Nagira Y**, **Miki R**, **Taoka N**, **Imai R.** 2018. Biolistic-delivery-based transient CRISPR/Cas9 expression enables *in planta* genome editing in wheat. *Sci Rep* **8:**14422.

Hao Y, **Charles TC**, **Glick BR.** 2010. ACC deaminase increases the *Agrobacterium tumefaciens*-mediated transformation frequency of commercial canola cultivars. *FEMS Microbiol Lett* **307:**185–190.

Heifetz PB, **Tuttle AM.** 2001. Protein expression in plastids. *Curr Opin Plant Biol* **4:**157–161.

Heim R, **Tsien RY.** 1996. Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer. *Curr Biol* **6:**178–182.

Iamtham S, **Day A.** 2000. Removal of antibiotic resistance genes from transgenic tobacco plastids. *Nat Biotechnol* **18:**1172–1176.

Ishida Y, **Saito H**, **Ohta S**, **Hiei Y**, **Komari T**, **Kumashiro T.** 1996. High efficiency transformation of maize (*Zea mays* L.) mediated by *Agrobacterium tumefaciens*. *Nat Biotechnol* **14:**745–750.

Jefferson RA, **Kavanagh TA**, **Bevan MW.** 1987. GUS fusions: β-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J* **6:**3901–3907.

Jin S, **Daniell H.** 2015. The engineered chloroplast genome just got smarter. *Trends Plant Sci* **20:**622–640.

Kathage J, **Qaim M.** 2012. Economic impacts and impact dynamics of Bt (*Bacillus thuringiensis*) cotton in India. *Proc Natl Acad Sci USA* **109:**11652–11656.

Klein TM, **Wolf ED**, **Wu R**, **Sanford JC.** 1987. High-velocity microprojectiles for delivering nucleic acids into living cells. *Nature* **327:**70–73.

Komarnytsky S, **Borisjuk NV**, **Borisjuk LG**, **Alam MZ**, **Raskin I.** 2000. Production of recombinant proteins in tobacco guttation fluid. *Plant Physiol* **124:**927–934.

Kong J, **Martin-Ortigosa S**, **Finer J**, **Orchard N**, **Gunadi A**, **Batts LA**, **Thakare D**, **Rush B**, **Schmitz O**, **Stuiver M**, **Olhoft P**, **Pacheco-Villalobos D.** 2020. Overexpression of the transcription factor *GROWTH-REGULATING FACTOR5* improves transformation of dicot and monocot species. *Front Plant Sci* **11:**572319.

Kwon C-T, **Heo J**, **Lemmon ZH**, **Capua Y**, **Hutton SF**, **Van Eck J**, **Park SJ**, **Lippman ZB.** 2020. Rapid customization of Solanaceae fruit crops for urban agriculture. *Nat Biotechnol* **38:**182–188.

Lassoued R, **Macall DM**, **Smyth SJ**, **Phillips PWB**, **Hesseln H.** 2019. Risk and safety considerations of genome edited crops: expert opinion. *Curr Res Biotechnol* **1:**11–21.

Lee JH, **Won HJ**, **Oh E-S**, **Oh M-H**, **Jung JH.** 2020. Golden gate cloning-compatible DNA replicon/2A-mediated polycistronic vectors for plants. *Front Plant Sci* **11:**559365.

Lee WS, **Tzen JTC**, **Kridl JC**, **Radke SE**, **Huang AHC.** 1991. Maize oleosin is correctly targeted to seed oil bodies in *Brassica napus* transformed with the maize oleosin gene. *Proc Natl Acad Sci USA* **88:**6181–6185.

Lowe K, **Wu E**, **Wang N**, **Hoerster G**, **Hastings C**, **Cho M-J**, **Scelonge C**, **Lenderts B**, **Chamberlin M**, **Cushatt J**, **Wang L**, **Ryan L**, **Khan T**, **Chow-Yiu J**, **Hua W**, **Yu M**, **Banh J**, **Bao Z**, **Brink K**, **Igo E**, **Rudrappa B**, **Shamseer PM**, **Bruce W**, **Newman L**, **Shen B**, **Zheng P**, **Bidney D**, **Falco C**, **Register J**, **Zhao Z-Y**, **Xu D**, **Jones T**, **Gordon-Kamm W.** 2016. Morphogenic regulators *Baby boom* and *Wuschel* improve monocot transformation. *Plant Cell* **28:**1998–2015.

Mahaffey H, **Taheripour F**, **Tyner WE.** 2016. Evaluating the economic and environmental impacts of a global GMO ban. *J Environ Prot (Irvine Calif)* **7:**1522–1546.

Mahomed S, **Garrett N**, **Baxter C**, **Abdool Karim Q**, **Abdool Karim SS.** 2021. Clinical trial of broadly neutralizing monoclonal antibodies for human immunodeficiency virus prevention: a review. *J Infect Dis* **223:**370–380.

Maliga P. 2004. Plastid transformation in higher plants. *Annu Rev Plant Biol* **55:**289–313.

Mardanova ES, **Blokhina EA**, **Tsybalova LM**, **Peyret H**, **Lomonossoff GP**, **Ravin NV.** 2017. Efficient transient expression of recombinant proteins in plants by the novel pEff vector based on the genome of potato virus X. *Front Plant Sci* **8:**247.

Margolin EA, **Strasser R**, **Chapman R**, **Williamson A-L**, **Rybicki EP**, **Meyers AE.** 2020. Engineering the plant secretory pathway for the production of next-generation pharmaceuticals. *Trends Biotechnol* **38:**1034–1044.

Mastrangeli R, **Palinsky W**, **Bierau H.** 2019. Glycoengineered antibodies: towards the next-generation of immunotherapeutics. *Glycobiology* **29:**199–210.

Miki B, **McHugh S.** 2004. Selectable marker genes in transgenic plants: applications, alternatives and biosafety. *J Biotechnol* **107:**193–232.

Ming M, **Ren Q**, **Pan C**, **He Y**, **Zhang Y**, **Liu S**, **Zhong Z**, **Wang J**, **Malzahn AA**, **Wu J**, **Zheng X**, **Zhang Y**, **Qi Y.** 2020. CRISPR-Cas12b enables efficient plant genome engineering. *Nat Plants* **6:**202–208.

Mitsuhara I, **Ugaki M**, **Hirochika H**, **Ohshima M**, **Murakami T**, **Gotoh Y**, **Katayose Y**, **Nakamura S**, **Honkura R**, **Nishimiya S**, **Ueno K**, **Mochizuki A**, **Tanimoto H**, **Tsugawa H**, **Otsuki Y**, **Ohashi Y.** 1996. Efficient promoter cassettes for enhanced expression of foreign genes in dicotyledonous and monocotyledonous plants. *Plant Cell Physiol* **37:**49–59.

Nicolia A, **Manzo A**, **Veronesi F**, **Rosellini D.** 2014. An overview of the last 10 years of genetically engineered crop safety research. *Crit Rev Biotechnol* **34:**77–88.

Nonaka S, **Sugawara M**, **Minamisawa K**, **Yuhashi K**, **Ezura H.** 2008. 1-Aminocyclopropane-1-carboxylate deaminase enhances *Agrobacterium tumefaciens*-mediated gene transfer into plant cells. *Appl Environ Microbiol* **74:**2526–2528.

Oey M, **Lohse M**, **Kreikemeyer B**, **Bock R.** 2009. Exhaustion of the chloroplast protein synthesis capacity by massive expression of a highly stable protein antibiotic. *Plant J* **57:**436–445.

Ow DW, **DE Wet JR**, **Helinski DR**, **Howell SH**, **Wood KV**, **Deluca M.** 1986. Transient and stable expression of the firefly luciferase gene in plant cells and transgenic plants. *Science* **234:**856–859.

Pauls KP. 1995. Plant biotechnology for crop improvement. *Biotechnol Adv* **13:**673–693.

Piron R, **Santens F**, **De Paepe A**, **Depicker A**, **Callewaert N.** 2015. Using GlycoDelete to produce proteins lacking plant-specific N-glycan modification in seeds. *Nat Biotechnol* **33:**1135–1137.

Potenza C, **Aleman L**, **Sengupta-Gopalan C.** 2004. Targeting transgene expression in research, agricultural, and environmental applications: promoters used in plant transformation. *in vitro Cell Dev Biol Plant* **40:**1–22.

Potrykus I. 1990. Gene transfer to cereals: an assessment. *Bio/Technology* **8:**535–542.

Potrykus I. 1991. Gene transfer to plants: assessment of published approaches and results. *Annu Rev Plant Physiol* **42:**205–225.

Roychowdhury S, **Oh YJ**, **Kajiura H**, **Hamorsky KT**, **Fujiyama K**, **Matoba N.** 2018. Hydroponic treatment of *Nicotiana benthamiana* with kifunensine modifies the N-glycans of recombinant glycoprotein antigens to predominantly Man9 high-mannose type upon transient overexpression. *Front Plant Sci* **9:**62.

Ruf S, **Hermann M**, **Berger IJ**, **Carrer H**, **Bock R.** 2001. Stable genetic transformation of tomato plastids and expression of a foreign protein in fruit. *Nat Biotechnol* **19:**870–875.

Shinoyama H, Ichikawa H, Nishizawa-Yokoi A, Skaptsov M, Toki S. 2020. Simultaneous TALEN-mediated knockout of chrysanthemum *DMC1* genes confers male and female sterility. *Sci Rep* **10:**16165.

Someya T, Nonaka S, Nakamura K, Ezura H. 2013. Increased 1-aminocyclopropane-1-carboxylate deaminase activity enhances *Agrobacterium tumefaciens*-mediated gene delivery into plant cells. *MicrobiologyOpen* **2:**873–880.

Southgate EM, Davey MR, Power JB, Marchant R. 1995. Factors affecting the genetic engineering of plants by microprojectile bombardment. *Biotechnol Adv* **13:**631–651.

Sugita K, Matsunaga E, Ebinuma H. 1999. Effective selection system for generating marker-free transgenic plants independent of sexual crossing. *Plant Cell Rep* **18:**941–947.

Vain P, De Buyser J, Bui Trang V, Haicour R, Henry Y. 1995. Foreign gene delivery into monocotyledonous species. *Biotechnol Adv* **13:**653–671.

Vallarino JG, Kubiszewski-Jakubiak S, Ruf S, Rößner M, Timm S, Bauwe H, Carrari F, Rentsch D, Bock R, Sweetlove LJ, Fernie AR. 2020. Multi-gene metabolic engineering of tomato plants results in increased fruit yield up to 23%. *Sci Rep* **10:**17219.

Walden R, Schell J. 1990. Techniques in plant molecular biology: progress and problems. *Eur J Biochem* **192:**563–576.

Walden R, Wingender R. 1995. Gene-transfer and plant-regeneration techniques. *Trends Biotechnol* **13:**324–331.

Wang C, Liu Q, Shen Y, Hua Y, Wang J, Lin J, Wu M, Sun T, Cheng Z, Mercier R, Wang K. 2019. Clonal seeds from hybrid rice by simultaneous genome engineering of meiosis and fertilization genes. *Nat Biotechnol* **37:**283–286.

Wang T, Zhang H, Zhu H. 2019. CRISPR technology is revolutionizing the improvement of tomato and other fruit crops. *Hortic Res* **6:**77.

Weber E, Engler C, Gruetzner R, Werner S, Marillonnet S. 2011. A modular cloning system for standardized assembly of multigene constructs. *PLoS One* **6:**e16765.

Woo JW, Kim J, Kwon SI, Corvalán C, Cho SW, Kim H, Kim S-G, Kim S-T, Choe S, Kim J-S. 2015. DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins. *Nat Biotechnol* **33:**1162–1164.

Yamamoto T, Hoshikawa K, Ezura K, Okazawa R, Fujita S, Takaoka M, Mason HS, Ezura H, Miura K. 2018. Improvement of the transient expression system for production of recombinant proteins in plants. *Sci Rep* **8:**4755.

Yoder JI, Goldsbrough AP. 1994. Transformation systems for generating marker-free transgenic plants. *Bio/Technology* **12:**263–267.

Zambryski P. 1988. Basic processes underlying *Agrobacterium*-mediated DNA transfer to plant cells. *Annu Rev Genet* **22:**1–30.

Zambryski P, Tempe J, Schell J. 1989. Transfer and function of T-DNA genes from *Agrobacterium* Ti and Ri plasmids in plants. *Cell* **56:**193–201.

review questions

1. Why is the Ti plasmid from *A. tumefaciens* well suited for developing a vector to transfer foreign genes into plant chromosomal DNA?
2. How do (i) binary and (ii) cointegrate Ti plasmid-based vector systems for plant transformation differ from one another?
3. What are reporter genes, and how are they used when plant cells are transformed?
4. How are plants transformed by microprojectile bombardment?
5. How can *A. tumefaciens* transformation of plants be made more efficient?
6. How is foreign DNA targeted for integration into chloroplast DNA?
7. How would you produce a transgenic plant that does not contain a marker gene in its nuclear DNA? In its chloroplast DNA?
8. How would you use microprojectile bombardment to introduce the CRISPR/Cas9 system into wheat plants which are otherwise recalcitrant to genome editing?
9. How would you ensure that the progeny of self-pollinated plants is identical to the initial hybrid parent?
10. How would you ensure that a foreign gene that has been inserted into the chloroplast is expressed at a high level?
11. How can you create specific targeted alterations in a plant's nuclear DNA?
12. What is the advantage of introducing foreign genes into chloroplast rather than nuclear DNA?
13. How would you produce polycistronic genes in plants?
14. What is rhizosecretion? Why is it useful? How can it be engineered?
15. How can foreign proteins be expressed in plant seeds? How can these proteins be easily purified?
16. What steps need to be taken to efficiently produce functional monoclonal antibodies in plants?
17. How do enhancer sequences facilitate plant gene expression?
18. How would you modify the glycosylation pattern of a mammalian protein produced in plants?
19. How might you avoid using selectable markers when producing transgenic plants?
20. How can foreign genes be transiently expressed in plants? Why is this useful?
21. Why would you use different promoters in constructing transgenic plants?

Insect Resistance
Bacillus thuringiensis Insecticidal Toxin
Increasing Expression of the *B. thuringiensis* Protoxin
Other Strategies for Protecting Plants against Insects
Preventing the Development of *B. thuringiensis*-Resistant Insects
Targeting Aphids
Virus Resistance
Viral Coat Protein-Mediated Protection
Protection by Expression of Other Genes
Herbicide Resistance
Glyphosate
Dicamba
Other Herbicides
Fungus and Bacterium Resistance
Transgenic Plants
RNAi and CRISPR/Cas
Salt and Drought Stress
Increasing Trehalose Production
Sequestering Sodium Ions
Delaying Drought-Induced Senescence
Phytoremediation
Fruits and Flowers
Flavr Savr Tomato
Lowering Ethylene Levels
CRISPR Mutants
Modification of Plant Nutritional Content
Amino Acids
Lipids
Vitamins
Iron
Gluten
Modification of Food Plant Taste and Appearance
Preventing Discoloration
Starch
Plants as Bioreactors
Antibodies
Pharmaceuticals and Vaccines
Poly(3-Hydroxybutyric Acid)
Edible Vaccines
Edible Cholera Vaccines
Edible *E. coli* Vaccines
Plant Yield
Increasing Grain Yield
Increasing Harvest Index
Decreasing Lignin Content
Decreasing Pectin Content
Increasing Oxygen Content
SUMMARY
REFERENCES
REVIEW QUESTIONS

11 Transgenic Plants

The principal objective of plant biotechnology is to create new varieties of cultivated plants (cultivars). The majority of the initial studies of transgenic plants focused on developing crops that gave better yields. Genes that confer resistance to insects, viruses, herbicides, environmental stresses, and senescence (natural aging) were incorporated into various plants. Some of this work has been commercialized and has been the subject of considerable public scrutiny and discussion worldwide. Several examples are discussed below.

Transgenic crops developed more recently, most of which have not yet been commercialized, include plants with increased nutritional content, modified taste and appearance, and increased yield. In addition, researchers have begun to develop transgenic plants that can be used as edible vaccines or as bioreactors for the production of recombinant proteins, including a variety of therapeutic agents. Table 11.1 provides examples of some of the transgenic plants that were commercially available as of early 2021. The details of many of these transgenic plants are discussed in this chapter.

Insect Resistance

Bacillus thuringiensis Insecticidal Toxin

Insects negatively affect humans either by causing significant levels of crop damage or by acting as vectors of both human and animal diseases. During the 1940s, a number of chemical insecticides were developed as

Molecular Biotechnology: Principles and Applications of Recombinant DNA, Sixth Edition. Bernard R. Glick and Cheryl L. Patten.
© 2022 American Society for Microbiology.

Cloning and Expression of the *Bacillus thuringiensis* Crystal Protein Gene in *Escherichia coli*

milestone

H. E. Schnepf and H. R. Whiteley
Proc. Natl. Acad. Sci. USA **78:**2893–2897, 1981

Although it had been well known for a long time that the parasporal crystal that is produced by *B. thuringiensis* contained insecticidal activity, it took scientists many years before the conditions for solubilizing the crystal were discovered and the insecticidal protein toxin was isolated in a pure form. Moreover, although some *B. thuringiensis* mutants that were defective in the synthesis of the parasporal crystal were known, protocols for the genetic transformation of *B. thuringiensis* were not well developed. Therefore, when Schnepf and Whiteley decided to isolate *B. thuringiensis* insecticidal toxin protein genes, they were limited to screening *E. coli* transformants carrying *B. thuringiensis* DNA either immunologically, using antibodies directed against the whole crystal, or by the insecticidal activity of extracts of the transformants. Since evidence at the time suggested that the insecticidal toxin was probably plasmid encoded, clone banks were constructed from fractionated plasmid preparations with the idea of significantly enriching the clone banks for the presence of insecticidal toxin genes. Moreover, despite some concerns that antibodies against the whole (*B. thuringiensis*-produced and glycosylated) crystal protein might not interact with (*E. coli*-produced and nonglycosylated) crystal protein subunits in solution, this did not turn out to be a problem. Finally, extracts of the particulate fraction of *E. coli* transformants carrying the *B. thuringiensis* insecticidal toxin gene were found to be toxic to susceptible insects. This first cloning of a *B. thuringiensis* insecticidal toxin gene made it clear to workers in this field that these genes could be isolated in a straightforward manner and provided an impetus for increased activity both in the search for new strains of *B. thuringiensis* and in studies of the biochemistry of the insecticidal toxin.

Table 11.1 Some examples of commercially available transgenic plants as of early 2021

Plant	Transgenic trait
Alfalfa	Herbicide tolerance, altered lignin production
Apple	Nonbrowning
Bean	Virus resistance
Canola	Herbicide tolerance, modified oil, phytase production
Cotton	Herbicide tolerance, insect resistance, low gossypol
Eggplant	Insect resistance
Maize (corn)	Herbicide tolerance, insect resistance
Melon	Delayed ripening, virus resistance
Papaya	Virus resistance
Pineapple	Delayed ripening, modified fruit color
Plum	Virus resistance
Potato	Virus resistance, herbicide tolerance, insect resistance, fungal disease resistance, modified starch, low acrylamide
Rice	Herbicide tolerance, insect resistance, biofortified with provitamin A
Rose	Modified flower color
Safflower	Modified product quality
Soybean	Herbicide tolerance, insect resistance, abiotic stress tolerance, altered oil content
Squash	Virus resistance
Sugar beet	Herbicide tolerance
Sugar cane	Insect resistance, drought tolerance
Sweet pepper	Virus resistance
Tomato	Virus resistance, insect resistance, delayed ripening
Wheat	Herbicide tolerance

a means of controlling the proliferation of noxious insect populations. One of these, the chlorinated hydrocarbon DDT (dichlorodiphenyltrichloroethane) proved to be exceptionally effective in killing and controlling many species of insect pests. Chlorinated hydrocarbons such as DDT function by attacking the nervous system and muscle tissue of insects. Later, other chlorinated hydrocarbons were synthesized and applied on a massive scale worldwide against crop pests and insects that carry infectious agents. Organophosphates, another class of chemical insecticides, are used to control insect populations by inhibiting the enzyme acetylcholinesterase, which hydrolyzes the nerve transmitter acetylcholine, thereby disrupting the functioning of motor neurons and neurons in the brain of the insect.

By the early 1960s, researchers realized that chlorinated hydrocarbon insecticides, to a large extent, and organophosphate insecticides, to a lesser extent, had dramatic and immediate side effects and long-term and indirect effects on animals, ecosystems, and humans. Chlorinated hydrocarbons were found to persist in the environment for more than 20 years and to accumulate in increasing concentrations through food chains. This bioaccumulation had a significant biological impact on many organisms; e.g., many species of birds underwent severe population declines. In addition, as the insect pests became increasingly resistant to treatment with many chemical insecticides, higher concentrations of the insecticides were applied, resulting in the killing of beneficial insects along with those insects that were considered to be pests.

Given the drawbacks associated with the use of chemical insecticides, alternative means of controlling harmful insects have been sought. Using insecticides that are produced naturally by either microorganisms or plants was an obvious choice. These compounds are usually highly specific for a target insect species, biodegradable, and slow to select for resistance. However, their low potency and high cost of production initially limited their use for many applications.

The most studied, most effective, and most often utilized microbial insecticides are the toxins synthesized by *Bacillus thuringiensis* (Fig. 11.1). This Gram-positive bacterium comprises a large number of strains and subspecies, each of which produces a different toxin that can kill specific insects—there are more than 350 different known subspecies of *B. thuringiensis* (e.g., Table 11.2).

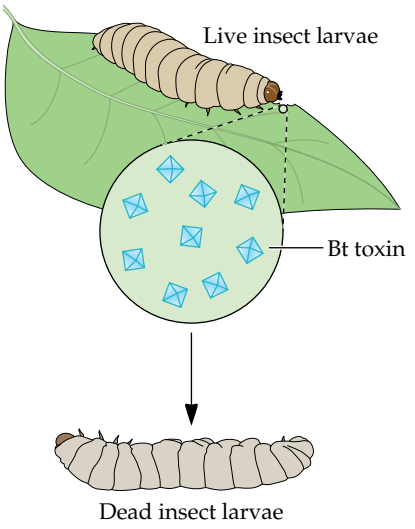


Figure 11.1 An insect larva feeding on a plant leaf. By consuming the *Bacillus thuringiensis* (Bt) parasporal crystal (insecticidal protoxin), the insect larva is eventually killed.

Table 11.2 Some properties of the insecticidal toxins from several common strains of *B. thuringiensis*

B. thuringiensis strain or subspecies	Protoxin size (kDa)	Target insects
berliner	130–140	Lepidoptera
kurstaki KTO, HD-1	130–140	Lepidoptera
entomocidus 6.01	130–140	Lepidoptera
aizawai 7.29	130–140	Lepidoptera
aizawai IC 1	135	Lepidoptera, Diptera
kurstaki HD-1	71	Lepidoptera, Diptera
tenebrionis (*san diego*)	66–73	Coleoptera
morrisoni PG14	125–145	Diptera
israelensis	68	Diptera

The insecticidal activity of *B. thuringiensis* subsp. *kurstaki* and other strains is contained within a very large structure called a parasporal crystal, which is synthesized during bacterial sporulation. By synthesizing the crystal, the bacterium is "providing for its future" in that a dead insect provides sufficient nutrients to allow germination of the dormant spore. The parasporal crystal usually consists mainly of protein (~95%) and a small amount of carbohydrate (~5%). The crystal is an aggregate of insecticidal Cry protein subunits that can generally be dissociated by mild alkali treatment into subunits. The subunits can be further dissociated *in vitro* by treatment with β-mercaptoethanol, which reduces disulfide linkages (Fig. 11.2). *B. thuringiensis* insecticidal Cry proteins are assigned designations based on their degree of evolutionary divergence.

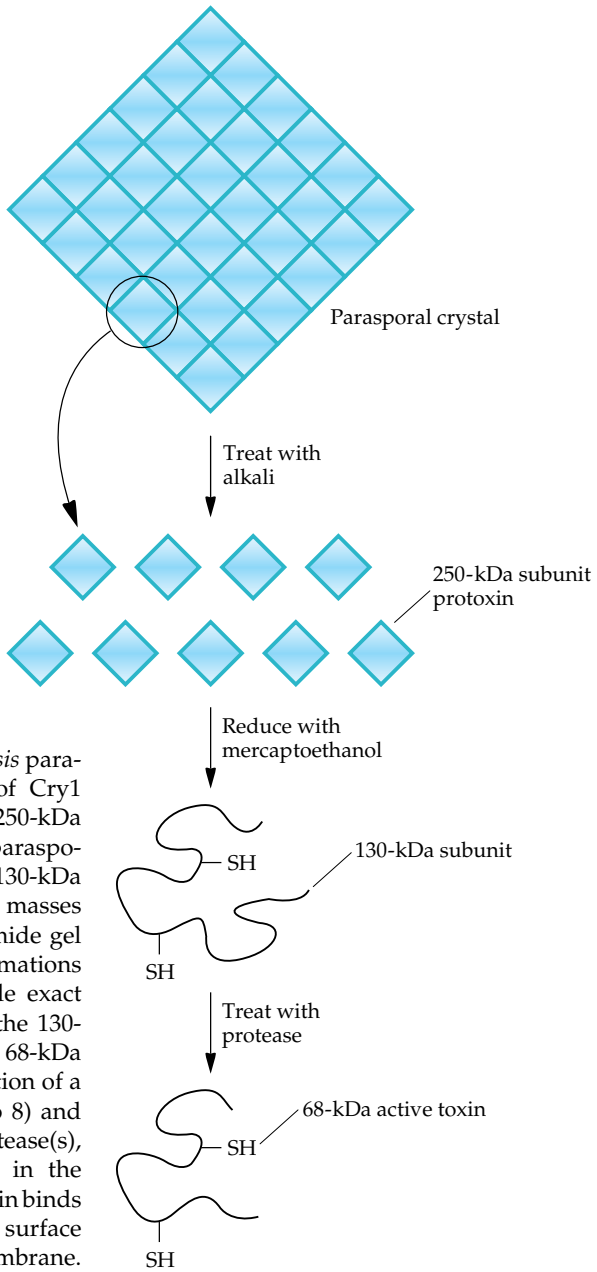


Figure 11.2 A *B. thuringiensis* parasporal crystal composed of Cry1 protoxin protein. Each 250-kDa protein subunit of the parasporal crystal contains two 130-kDa polypeptides. (Molecular masses determined by polyacrylamide gel electrophoresis are approximations and do not always provide exact multiples.) Conversion of the 130-kDa protoxin into an active 68-kDa toxin requires the combination of a slightly alkaline pH (7.5 to 8) and the action of a specific protease(s), both of which are found in the insect gut. The activated toxin binds to protein receptors on the surface of the gut epithelial cell membrane.

The parasporal crystal does not usually contain the active form of the insecticide. Rather, once the crystal has been solubilized, the protein that is released is generally a protoxin, a precursor of the active toxin. The protoxins of many of the Cry toxins that are directed against lepidoptera (order of insects comprising butterflies, moths, and skippers) have a molecular mass of approximately 130 kilodaltons (kDa) (Fig. 11.2). When a parasporal crystal is ingested by a target insect, the protoxin is activated within its gut by the combination of alkaline pH (7.5 to 8.0) and specific digestive proteases, which convert the protoxin into an active toxin with a molecular mass of approximately 68 kDa (Fig. 11.2). In its active form, the toxic protein inserts itself into the membranes of the gut epithelial cells of the target insect and creates an ion channel, which leads to an excessive loss of cellular ATP (Fig. 11.3). About 15 minutes after this ion channel forms, cellular metabolism ceases; the insect stops feeding within a few hours, becomes dehydrated, and eventually dies (in about 2 to 5 days). Because the conversion of the protoxin to the active toxin requires both alkaline pH and the presence of specific proteases, it is extremely unlikely that nontarget species, such as humans and farm animals, will be affected.

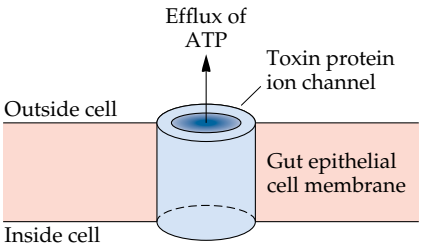


Figure 11.3 Insertion of the *B. thuringiensis* toxin into the membrane of an insect gut epithelial cell. The toxin forms an ion channel between the cell cytoplasm and the external environment.

To kill an insect pest, *B. thuringiensis* parasporal crystals must be ingested. This requirement further limits the susceptibility of nontarget insects and other animals to the insecticide. Insects that bore into plants or attack plant roots are less likely to ingest a *B. thuringiensis* toxin that has been sprayed on a host plant or expressed in leaves of genetically engineered plants.

Sunlight degrades over 60% of the tryptophan residues of the parasporal crystal within a 24-hour period, thereby rendering the protein inactive. Depending on the amount of sunlight present, parasporal crystals may persist in the environment for as little as a day or as long as a month. The lack of long-term persistence of the insecticidal protoxin in the natural environment means that natural selection of resistant insects is highly unlikely.

Because *B. thuringiensis* protoxin does not persist in the environment, and is not hazardous to mammals, it is a safe means of protecting plants. Biological insecticides are usually highly specific for a limited number of insect species and therefore less disruptive to the ecosystem. In addition, by reducing the damage to plants from insect predation, a corresponding decrease in the damage to plants from several fungal diseases should result, since many pathogenic fungi often invade a plant either together with or as a consequence of insect infection. Since the mid-1980s, *B. thuringiensis* has been used for aerial spraying programs for spruce budworm of more than 8 million hectares of insect-infested forests in Canada, completely replacing the use of chemical insecticides.

The genetic engineering of crop plants to produce functional insecticides makes it possible to develop crops that are intrinsically resistant to specific insect predators and do not need to be sprayed (often six to eight times during a growing season) with costly and potentially hazardous chemical pesticides. It has been estimated that in 2020 the amount spent on chemical insecticides worldwide was ~$18 billion, an amount that would be much higher if we weren't already growing a large number of *B. thuringiensis* transgenic crops. The cost of maintaining genetically engineered insect-resistant crops is significantly lower than that for nonresistant crops. In addition, it is both simpler and less costly to express the genes for *B. thuringiensis* toxins in plants than to spray *B. thuringiensis* preparations onto the surface of the plant. This mode of insecticidal toxin delivery limits

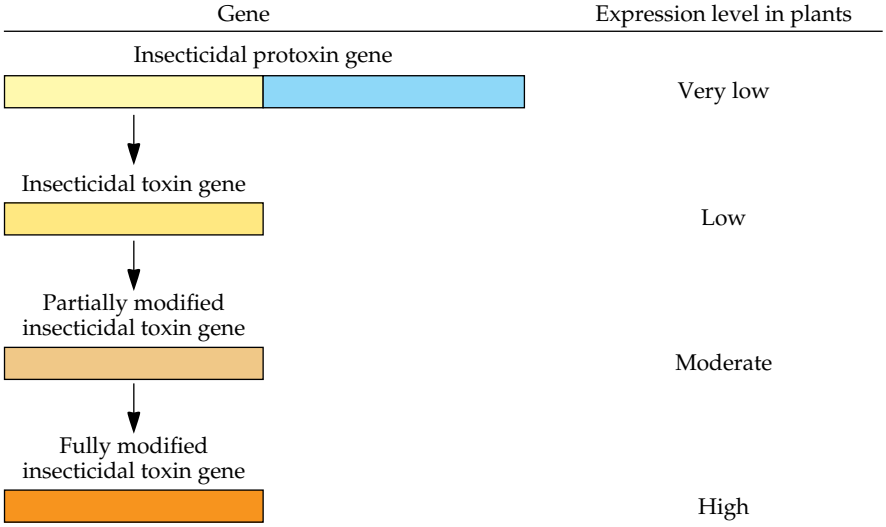
the environmental distribution of the toxin and avoids problems associated with spraying *B. thuringiensis* preparations, such as limited environmental stability and the timing of the toxin application.

Increasing Expression of the *B. thuringiensis* Protoxin

Once the isolation and sequencing of a toxin gene were accomplished, the complete amino acid sequence was determined. The amino acid sequences of protoxins from various strains of *B. thuringiensis* were compared to determine whether there is a common insecticidal (toxin) domain. This analysis showed that the N-terminal portion of the protoxin protein molecule is highly conserved (~98%) and the C-terminal region is more variable (~45% conserved). Further work showed that all of the insecticidal toxin activity resides within the first 646 amino acids from the N terminus of the 1,156-amino-acid protoxin.

In initial experiments, the *B. thuringiensis* subsp. *kurstaki* insecticidal-protein genes, *cry1Aa*, *cry1Ab*, and *cry1Ac*, were not particularly well expressed in plants. This is problematic, because high levels of expression of these insect control proteins are needed to produce commercially viable insect-resistant plants. To raise the level of the expressed protein, scientists truncated the gene so that only the N-terminal portion of the insecticidal protoxin was produced and inserted a strong plant promoter to direct gene expression. When the segment of the protoxin gene that encodes the highly conserved amino acid sequence was cloned and expressed in bacteria, the shortened protein was more active than the native (protoxin) form in protecting plants against lepidopteran insects in laboratory trials (Fig. 11.4). When the shortened version of the toxin gene was partially modified to

Figure 11.4 Relative expression levels of *B. thuringiensis* protoxin and toxin genes in transgenic plants. The *B. thuringiensis* protoxin gene is the native form of the gene, the insecticidal toxin gene results from the portion of the gene encoding the precursor portion of the protein, in the partially modified insecticidal toxin gene some of the codons have been changed to reflect codons that are more commonly used in plants, and in the fully modified insecticidal toxin gene the entire gene has been resynthesized to reflect codons that are commonly used in plants.



include more plant codons, expression was further increased, and when the entire toxin gene was synthesized to include only plant-preferred codons, the level of expression was relatively high. Thus, going from the native insecticidal protoxin gene to the fully modified insecticidal toxin gene, there was a significant increase (~100- to 1,000-fold) in the level of insecticidal toxin produced, affording transgenic plants significant protection against damage from insect predation. This higher level of insecticidal toxin synthesis was directly correlated with increased insecticidal activity.

To construct transgenic tomato plants expressing a truncated and fully modified *B. thuringiensis* toxin gene, the following construct was used. The shortened version of the toxin gene was put under the transcriptional control of the strong constitutive 35S promoter from *Cauliflower mosaic virus* and the nopaline synthase transcription termination–polyadenylation site and was then cloned into the T-DNA (transfer DNA) region of a cointegrate-type Ti plasmid vector (Fig. 11.5). The vector contained a spectinomycin resistance (Spc^r) gene that allowed it to be selected in either *Escherichia coli* or *Agrobacterium tumefaciens*, an *E. coli* origin of DNA replication, and a neomycin phosphotransferase gene that was under the control of the nopaline synthase promoter and transcription termination–polyadenylation sites and enabled the selection of transformed plant cells in the presence of kanamycin. In addition, the cointegrate cloning vector had the right border of the T-DNA from a nopaline Ti plasmid and a segment of the octopine Ti plasmid that provides a region of homology for cointegrate formation by homologous recombination with a disarmed Ti plasmid (see chapter 10). The plasmid was constructed and manipulated in *E. coli* before it was transferred by conjugation to a strain of *A. tumefaciens* that contained a disarmed Ti plasmid. After recombination in *A. tumefaciens*, the shortened toxin gene construct was transferred to the chromosomal DNA of tomato plants.

In both greenhouse and field trials, transgenic tomato plants that expressed the short form of the toxin were protected to varying degrees against damage caused by tobacco hornworms (*Manduca sexta*), tomato fruitworms (*Heliothis zea*), and tomato pinworms (*Keiferia lycopersicella*). The extent of the protection was not the same for each of the insects, nor was it complete.

In another approach to increasing the expression of the protoxin, one group of researchers expressed the fully modified protoxin gene under the control of the promoter for the gene that codes for the small subunit of the plant enzyme ribulose bisphosphate carboxylase and downstream from the chloroplast transit peptide sequence of this enzyme, so that the overproduced protoxin became localized within the chloroplast. This strategy led to a very high level of expression (nearly 1% of the total leaf protein) of the insecticidal protoxin. Other researchers have introduced an insecticidal protoxin gene directly into the chloroplast DNA of the host plant. The *B. thuringiensis* protoxin gene was integrated into a specific site on the chloroplast DNA by constructing a vector that contained the protoxin gene flanked by two single-copy chloroplast genes (Fig. 11.6). Integration of the introduced genes occurs by homologous recombination. Once integrated into the chloroplast DNA, a protoxin gene under the transcriptional control of a strong chloroplast promoter may be expressed at high levels (see chapter 10), so the protoxin may account for as much as 2 to 3% of the total soluble protein in the leaf, yielding a very high level of insecticidal activity.

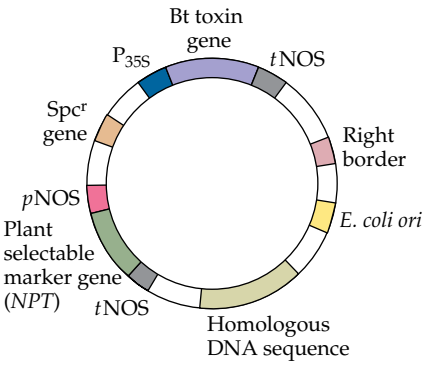


Figure 11.5 Cointegrate cloning vector carrying a *B. thuringiensis* (Bt) insecticidal toxin gene. The toxin gene is under the control of the strong, constitutive 35S promoter (P_{35S}) from *Cauliflower mosaic virus* and the nopaline synthase transcription terminator–polyadenylation site (*t*NOS). The vector has an *E. coli* origin of DNA replication (*ori*) and an Spc^r gene, which allow the vector to be maintained and selected in *E. coli* cells, a T-DNA right border, a plant selectable marker gene, and a region of DNA that is homologous to DNA in the disarmed Ti plasmid, for integrating the two plasmids. The neomycin phosphotransferase gene (*NPT*), which acts as a plant reporter gene, is under the transcriptional control of nopaline synthase gene sequences (*p*NOS and *t*NOS) and is used to select for kanamycin-resistant transformed plant cells.

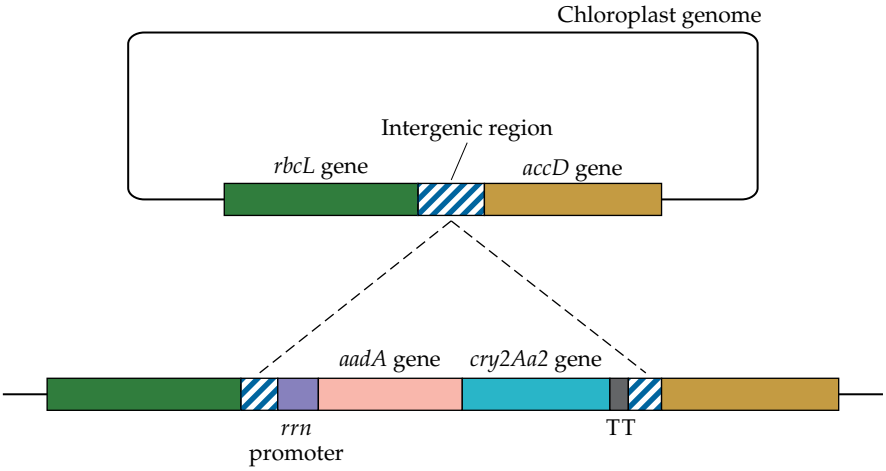


Figure 11.6 Site on the chloroplast genome where a foreign gene encoding the *B. thuringiensis* Cry2Aa2 protoxin is integrated by homologous recombination. The genes *rbcL* and *accD* are both present in a single copy per chloroplast genome. The intergenic region between these two genes, which is the site of insertion of the foreign genes, is smaller than it appears in this representation. The *aadA* gene (encoding spectinomycin and streptomycin resistances) and the *cry2Aa2* gene are both under the transcriptional control of the constitutive chloroplast *rrn* promoter and transcription terminator (TT), and each contains its own ribosomal binding site. Integration of foreign DNA into the intergenic spacer region prevents insertion of a foreign gene from interfering with the expression of any endogenous chloroplast genes. Data from Kota et al., *Proc. Natl. Acad. Sci. USA* **96:**1840–1845, 1999.

Integration of the *B. thuringiensis* protoxin gene into chloroplast DNA has several potential advantages over inserting it into the chromosomal DNA. First, the protoxin gene does not have to be modified, because the chloroplast transcriptional and translational apparatuses are typically prokaryotic. Second, because there are many chloroplasts per cell and many copies of chloroplast DNA per chloroplast, the protoxin gene is present in multiple copies and therefore is more likely to be expressed at a high level. Third, in most plants, chloroplasts are transmitted only through the egg (ovule) and not through pollen, which means that plants receive all their chloroplast DNA from their female parent. Consequently, there is no risk of unwanted transfer of the protoxin gene to other plants in the environment by pollen. The disadvantage of expressing the *B. thuringiensis* protoxin in chloroplasts is that insects that attack stems or fruit will not encounter the protoxin, since these tissues do not have any chloroplasts.

To date, some form of the gene for the protoxin has been introduced and expressed in a wide variety of plant species, including alfalfa, apple, broccoli, cabbage, canola, corn (maize), cotton, cranberry, eggplant, grape, hawthorn, juneberry, peanut, pear, poplar, potato, rice, rutabaga, soybean, spruce, sugar cane, tobacco, tomato, walnut, white clover, and white spruce. Following several seasons of successful field trials, these transgenic plants were approved for commercial release in the United States, Canada, and Argentina, and large-scale growth of the plants in the field began in 1996. It is estimated that worldwide, in 2020, farmers planted ~70 million hectares of transgenic *B. thuringiensis* insecticidal protein-containing crops. Although insect populations still must be monitored to keep track of the frequency of resistant organisms, the use of crops expressing

B. thuringiensis insecticidal proteins has already exceeded the length of time (often 2 to 4 growing seasons) that it typically takes for resistance to arise in insects to conventional insecticides. This technology has more than lived up to the hopes and expectations of scientists. Notwithstanding the initial concerns about the technology, especially in Europe, this approach to crop protection has gained widespread commercial acceptance throughout much of the world.

Other Strategies for Protecting Plants against Insects

No single *B. thuringiensis* protoxin is effective against a broad range of insect species. This may limit the overall usefulness of these protoxins. However, plants have evolved general insect defense mechanisms that are sufficient for plant survival but not always effective enough to keep the damage to a level that would be acceptable for crop plants. Therefore, some of the genes encoding plant defense proteins could be engineered to enhance the activity of *B. thuringiensis* protoxins.

Protease Inhibitors

Some plants produce protease inhibitors that, when ingested, prevent the feeding insect from hydrolyzing plant proteins, thereby effectively starving the predator insect. Consequently, it seemed reasonable to isolate a plant gene for a protease inhibitor, add a strong promoter, and create transgenic crop plants that produce sufficiently high levels of the protease inhibitor to reduce damage from insect predation.

In one early study, researchers isolated a DNA sequence that encodes cowpea trypsin inhibitor from a complementary DNA (cDNA) library (see chapter 2) by using a chemically synthesized DNA probe based on the amino acid sequence of the cowpea trypsin inhibitor protein. The full-length cDNA was subcloned onto a Ti plasmid binary cloning vector (Fig. 11.7) and introduced into a strain of *A. tumefaciens* carrying a disarmed Ti plasmid that contained active *vir* genes. Following *A. tumefaciens* infection of tobacco leaf disks with this vector, cells that incorporated the cloned DNA were selected for growth on kanamycin, and transgenic plants were regenerated. The damage caused by *Heliothis virescens* (tobacco budworm) larvae to transgenic plants that expressed more than 2 μg of cowpea trypsin inhibitor per mg of protein was significantly reduced compared to the damage inflicted on nontransformed plants.

Cowpea (*Vigna unguiculata*) seeds that contain relatively high levels of cowpea trypsin inhibitor are not toxic to either animals or humans. However, if the amount of protease inhibitor produced by a transgenic plant is determined to be a potential hazard, then it is possible to limit the expression of the protease inhibitor to the plant tissues that the major insect pests prefer but that are not used as food by humans or animals. In other words, a cloned protease inhibitor gene could be active in the leaves and roots of a plant but not in the commercially valuable fruit.

α-Amylase Inhibitor

Another way of imparting insect resistance to susceptible plants entails using a gene that encodes an α-amylase inhibitor. The cowpea weevil (*Callosobruchus maculatus*) and the azuki bean weevil (*Callosobruchus chinensis*) are seed-feeding beetles that cause considerable economic loss of these legume crops, especially in developing countries. When larvae of

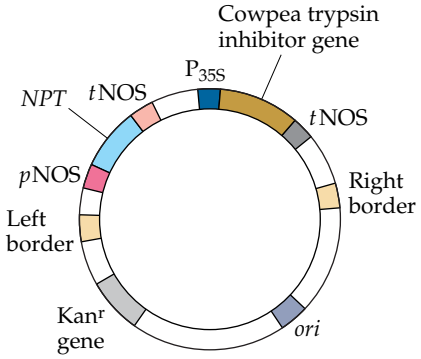


Figure 11.7 Binary cloning vector carrying a cowpea trypsin inhibitor gene. The vector contains a broad-host-range origin of DNA replication (*ori*) and a kanamycin resistance (Kan^r) gene, which function in both *E. coli* and *A. tumefaciens*. Between the T-DNA left and right borders, there are (i) a neomycin phosphotransferase gene (*NPT*) under the transcriptional control of nopaline synthase signals (*p*NOS and *t*NOS), which enables kanamycin-resistant transformed plant cells to be selected, and (ii) the cowpea trypsin inhibitor gene, which is under the control of the 35S promoter (P_{35S}) from *Cauliflower mosaic virus* and the transcription terminator–polyadenylation region from the nopaline synthase gene (*t*NOS).

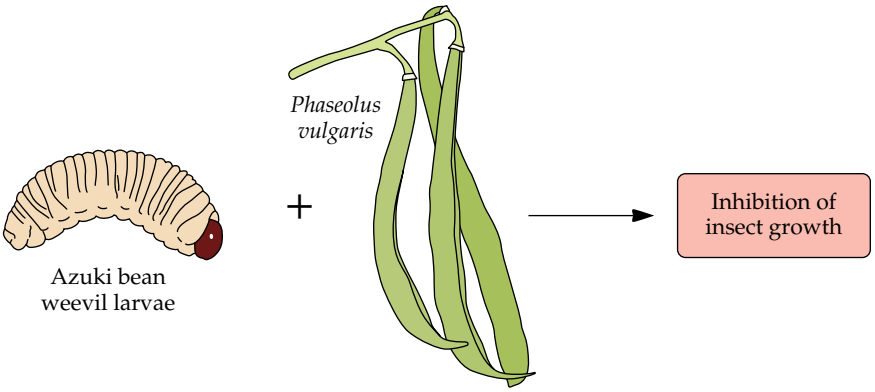


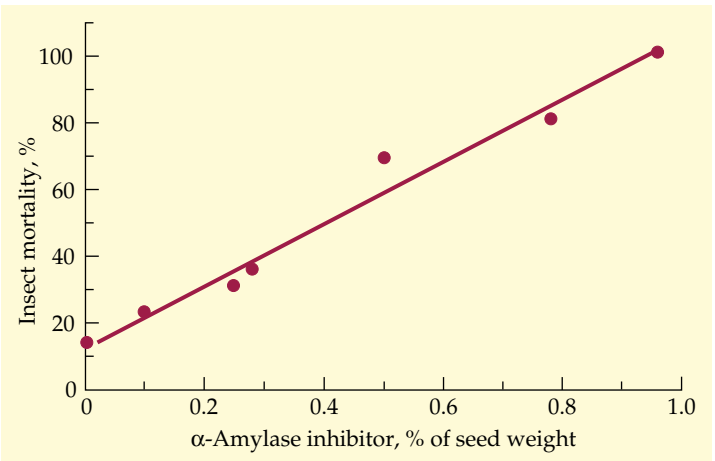
Figure 11.8 Inhibition of Azuki bean weevil larvae by *Phaseolus vulgaris* (common bean plants).

these insects are fed a diet that includes the common bean (*Phaseolus vulgaris*), insect growth is inhibited (Fig. 11.8). This growth inhibition is attributable to the presence of an α-amylase inhibitor in the seed proteins of the common bean. Accordingly, the cDNA for the α-amylase inhibitor from the common bean was isolated, placed under the transcriptional control of the strong seed-specific promoter for the bean phytohemagglutinin gene, and used to transform pea plants (*Pisum sativum*). Peas are usually quite susceptible to damage by both cowpea weevils and azuki bean weevils. However, transgenic pea plants that expressed the α-amylase inhibitor were resistant to both of these insects. The level of resistance to cowpea weevils was found to be proportional to the amount of α-amylase inhibitor that the transgenic plant produced (Fig. 11.9).

Cholesterol Oxidase

Another approach to developing insect-resistant transgenic plants makes use of a bacterial cholesterol oxidase gene. Cholesterol oxidase, which is present in a range of different bacterial genera, catalyzes the oxidation of

Figure 11.9 Mortality of cowpea weevil larvae reared on transgenic pea plants that produce different amounts of α-amylase inhibitor.



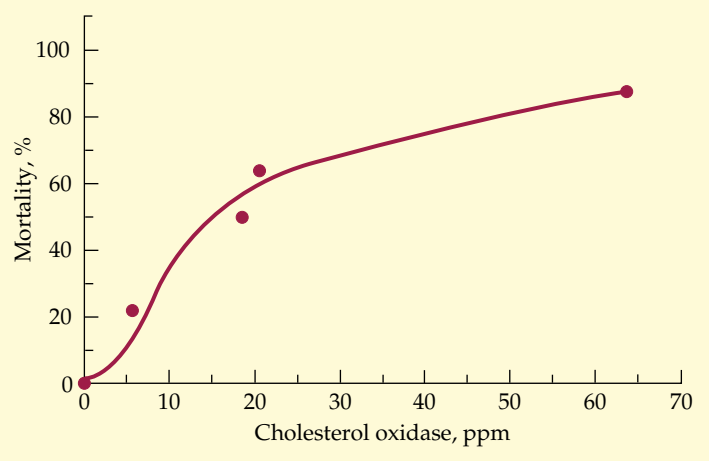


Figure 11.10 Effect of increasing amounts of cholesterol oxidase on the mortality of boll weevil larvae. ppm, parts per million. Data from Corbin et al., *Appl. Environ. Microbiol*. **60:**4239–4244, 1994.

3-hydroxysteroids to ketosteroids and hydrogen peroxide. Low levels of the enzyme have a high level of insecticidal activity against larvae of the boll weevil (*Anthonomus grandis grandis*) (Fig. 11.10), a common and economically important insect (Coleoptera) pest of cotton, and have lower levels of activity against some lepidopteran pests. Cholesterol oxidase is thought to act by disrupting the insect's midgut epithelial membrane, thus killing the insect. A cholesterol oxidase gene encoding a protein with a molecular mass of approximately 55,000 daltons with a length of 504 amino acids plus a leader peptide of approximately 5,000 daltons (43 amino acids) was isolated from a strain of *Streptomyces* and cloned into a vector under the control of a plant virus (*Figwort mosaic virus*) promoter and a termination sequence from the 3′ region of the *A. tumefaciens* nopaline synthase gene. When this construct was introduced into tobacco cell protoplasts, the transformed cells actively expressed the cholesterol oxidase. Following the targeting of this gene to the plant cytoplasm, cholesterol oxidase expression resulted in severe abnormalities in plant development. However, expression of this gene as a fusion with a chloroplast-targeting peptide did not cause any observable deleterious phenotypic effects. When the gene is introduced into cotton plants for commercial production, either by itself or in combination with genes for other biological insecticides, it should be an effective means of protecting plants against damage from insect predation.

Vegetative Insecticidal Toxins

In addition to the large number of well-characterized Cry insecticidal toxins, many strains of *B. thuringiensis* produce a secreted insecticidal protein during their vegetative growth stage. To date, two major groups of vegetative insecticidal proteins (Vip) have been identified. One group consists of the proteins Vip1 and Vip2, which are not toxic to lepidoptera, and Vip3, which targets several major lepidopteran pests. The less-well-characterized Vip proteins may act synergistically with Cry proteins to kill their target insects, providing a double-barreled approach to insect toxicity, so that it is extremely difficult for susceptible insects to develop resistance. It would

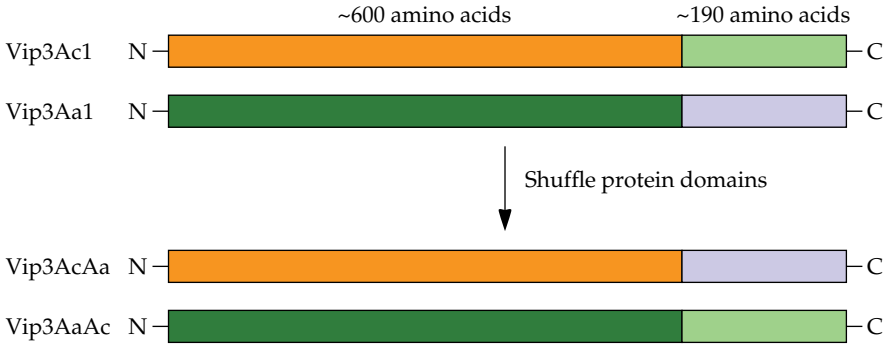


Figure 11.11 Generation of two hybrid insecticidal proteins constructed by domain shuffling. The approximate sizes of the amino- and carboxy-terminal fragments are indicated. The hybrid protein Vip3AcAa was found to have the greatest level of insecticidal activity with the insect larvae tested.

therefore be advantageous if transgenic plants expressing both Cry and Vip proteins were created. As a first step, researchers shuffled the two major domains (see chapter 3) of two Vip3 proteins, Vip3Ac1 and Vip3Aa1 (Fig. 11.11). One of the hybrid Vip3 proteins (i.e., Vip3AcAa) displayed the highest activity of the four proteins against fall armyworms, cotton bollworms, and silkworms. Moreover, only the Vip3AcAa construct was toxic to a strain of cabbage looper that was resistant to the well-characterized *B. thuringiensis* insecticidal protein Cry1Ac. The chimeric toxin Vip3AcAa enriches the diversity of Vip toxins that can be used together with conventional Cry proteins to generate transgenic plants that are highly unlikely to select for resistant insects.

Other Proteins

The activities of several other proteins have been utilized in an effort to protect plants from insect predation. For example, some lectins, i.e., carbohydrate-binding proteins found in the seeds and storage tissues of a variety of plant species, are toxic to certain species of insects. While many plant lectins are toxic to mammals as well as insects, the lectin from the snowdrop plant (*Galanthus nivalis*) is toxic only to insects. With this in mind, the snowdrop lectin gene has been introduced into approximately a dozen different plants, with the result that plants that expressed this protein were damaged by aphids (small sap-sucking insects) to a lesser extent than nontransformed plants. However, although the snowdrop lectin significantly lowered the amount of leaf material eaten, insect mortality was only slightly increased.

The whitefly (*Bemisia tabaci*) is a sap-sucking insect pest that damages crops and transmits a wide range of plant viral diseases, and this insect pest is not killed or inhibited by any known strains of *B. thuringiensis*. Recently, a group of researchers in India reported identifying a protein (Tma12) from the edible fern, *Tectaria macrodonta* (Fee) C. Chr., that is insecticidal to whitefly with a median lethal concentration of ~1.5 μg/ml in feeding experiments. The gene encoding this protein was isolated and used to construct transgenic cotton plants. Transgenic cotton plants that expressed Tma12 at ~0.01% of total leaf protein were completely protected from damage caused by whitefly as well as from viral diseases carried by

whitefly. In toxicity tests with rats that were fed relatively high levels of Tma12, there was no indication of any toxicity caused by this insecticide. Scientists are currently testing these transgenic cotton plants in the field.

When the gene encoding the enzyme tryptophan decarboxylase from Madagascar periwinkle plants (*Catharanthus roseus*) is expressed in tobacco, the plants are protected from damage by the whitefly (*B. tabaci*). While the precise mechanism of this protection is unknown, it has been suggested that the tryptamine that is produced by this enzyme, following the decarboxylation of tryptophan, is used in the production of insect-inhibiting plant alkaloids.

The Gram-negative bacterium *Photorhabdus luminescens* produces a 283-kilodalton protein, toxin A, that is highly toxic to a variety of insects. This protein was expressed in transgenic *Arabidopsis thaliana* (Box 11.1) plants in amounts of ≥700 ng/mg of extractable plant protein and was highly toxic to the tobacco hornworm, as well as the southern corn rootworm, with insect mortality typically at 100%.

Researchers have recently found a protein, IPDO72Aa, from *Pseudomonas chlororaphis* that can limit insect damage from the coleopteran western corn rootworm (*Diabrotica virgifera virgifera* LeConte) in transgenic corn plants expressing this protein. These insects had become resistant to several different *B. thuringiensis* insecticidal proteins previously expressed in these plants. While this is a preliminary report, it is consistent with the notion that several soil bacteria produce insecticidal proteins that might complement the activity of the more commonly utilized Cry proteins from *B. thuringiensis*.

Finally, transgenic corn plants that express avidin—a glycoprotein isolated from chicken eggs that binds the coenzyme biotin with extremely high affinity—caused biotin deficiency that led to stunted growth and death in several different insect species. Importantly, the levels of avidin that are toxic to insects are not toxic to mice, suggesting that protecting plants with an avidin transgene is not a problem for humans.

Arabidopsis thaliana

box 11.1

A. thaliana (thale cress) is a small weed in the same family (*Brassicaceae*) as canola, mustard, and broccoli. It is native to Europe, Asia, and northwestern Africa. *A. thaliana* is popular with scientists as a model organism in plant biology and genetic studies. It has one of the smallest genomes (at 7×10^7 bp, it is similar to the size of the yeast genome, which is approximately 1.5×10^7 bp) of any flowering plant, which makes it readily amenable to molecular genetic studies. The small size of its genome has made *A. thaliana* useful for the generation and selection of mutants, and it was the first plant genome to be sequenced, in 2000.

The plant's small size and short life cycle are also advantageous for research. Laboratory strains of *A. thaliana* take about 6 weeks from germination to mature seed. The small size of the plant is convenient for cultivation in limited space, and it produces many seeds—an individual plant can produce several thousand seeds.

Plant transformation in *Arabidopsis* is straightforward and has become a routine procedure in many laboratories, using *A. tumefaciens* to transfer DNA to the plant genome. The current *Arabidopsis* transformation protocol, termed "floral dip," involves dipping a flower into a solution containing *Agrobacterium*, the DNA of interest, and a detergent. This method avoids the need for tissue culture or plant regeneration. The idea is that some of the *Agrobacterium* cells will transfer their T-DNA containing the target DNA into the reproductive tissue of the plant. As a consequence of the above-mentioned traits and the relative complexity of most other plants, *Arabidopsis* has become the *E. coli* of the plant world.

RNA Interference

The ingestion or microinjection of double-stranded RNA into some worms and insects has been used to silence genes in these organisms. This gene silencing works through the generation of RNA interference (RNAi) (see chapter 6). In one study, 290 different double-stranded RNAs (dsRNAs) derived from sequences thought to encode essential or important insect functions were fed (one at a time), at a low level, to larvae of the western corn rootworm (*D. virgifera* LeConte). Growth of this coleoptera insect was significantly inhibited by 67 of the RNAs; 14 of the dsRNAs had 50% lethal concentrations of ≤5.2 ng/cm^2. Based on these results, transgenic corn expressing one of the most effective dsRNAs (see chapter 6), targeting the transcript for vacuolar ATPase, was constructed (Fig. 11.12). The transgenic plants were protected against the western corn rootworm to an extent comparable to the protection afforded by a *B. thuringiensis* transgene. The demonstration that it is possible to produce RNAi in coleoptera following oral delivery of dsRNA is an important first step in the development of a completely novel approach to developing a wide range of insect-resistant plants, including resistance to some insects that are refractory to the *B. thuringiensis* insecticidal toxin.

In a variation on the above-mentioned strategy, another group of researchers first identified the mechanism that the cotton bollworm uses to protect itself against the compound gossypol, which is produced by cotton plants to prevent insect predation. Gossypol (Fig. 11.13A) is a yellow polyphenolic aldehyde derived from cotton plants that permeates cells and acts as an inhibitor of several of the insect's dehydrogenase enzymes. It has been used as a male oral contraceptive in China, possesses antimalarial properties, and may have anticancer properties. The cotton bollworm protects

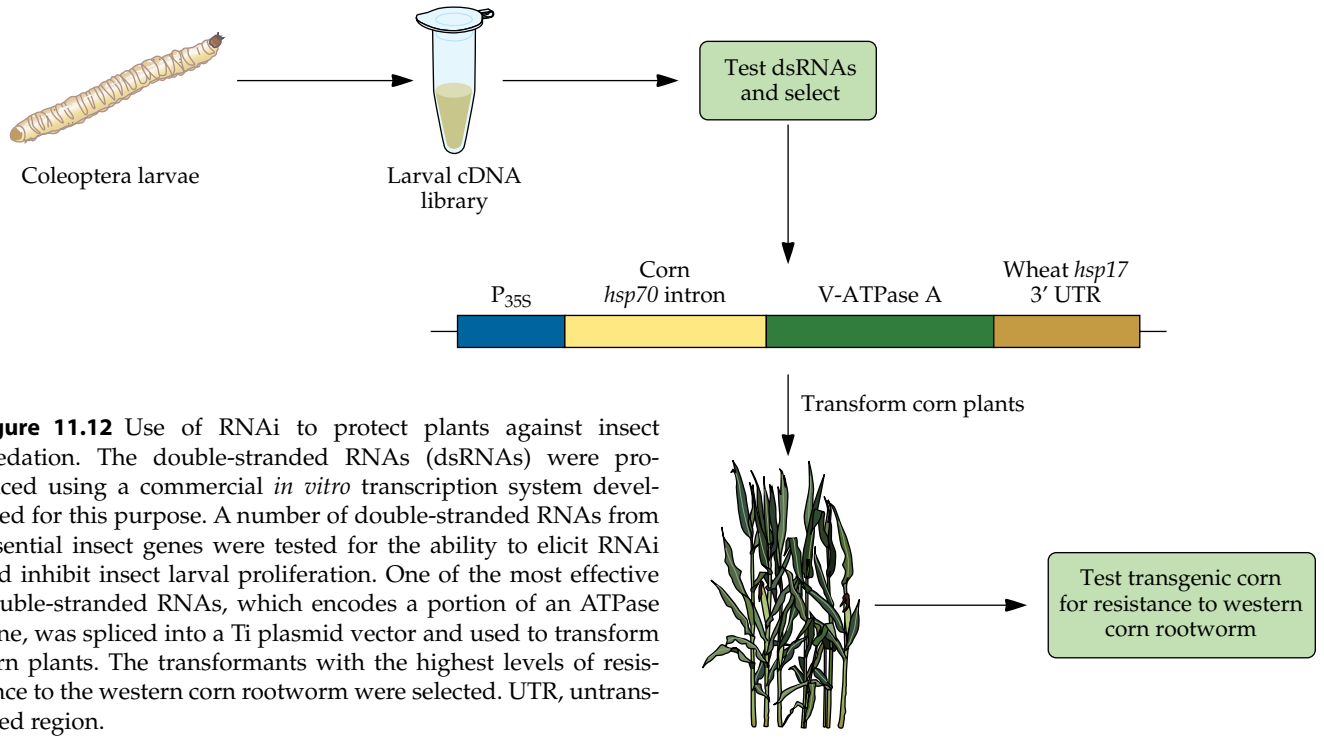


Figure 11.12 Use of RNAi to protect plants against insect predation. The double-stranded RNAs (dsRNAs) were produced using a commercial *in vitro* transcription system developed for this purpose. A number of double-stranded RNAs from essential insect genes were tested for the ability to elicit RNAi and inhibit insect larval proliferation. One of the most effective double-stranded RNAs, which encodes a portion of an ATPase gene, was spliced into a Ti plasmid vector and used to transform corn plants. The transformants with the highest levels of resistance to the western corn rootworm were selected. UTR, untranslated region.

itself from the toxic effects of gossypol by inactivating the gossypol with the enzyme cytochrome P450 monooxygenase. To prevent insects from inactivating gossypol, transgenic plants were constructed to synthesize an RNAi molecule that would silence the insect's gene for the cytochrome P450 monooxygenase. By preventing the expression of cytochrome P450 monooxygenase, the insect was exposed to the full toxic effects of the plant-produced gossypol, so that it was killed or at least debilitated, and the damage to the plant was limited (Fig. 11.13). Plants produce a myriad of allelochemicals to protect themselves against insects, and many insects have developed strategies to overcome the toxic effects of many of these compounds. Therefore, the mechanisms utilized by insects to overcome the toxicity of the plant-produced compounds are attractive targets for developing insect-resistant plants in the future.

A number of transgenic plants have been constructed to produce small interfering RNAs (siRNAs) that are targeted to essential insect mRNAs, with the idea that insects that ingest these siRNAs will be killed. However, not all transgenic plants expressing siRNAs are as effective at killing insect pests as expected. Recently, one group of researchers observed that transgenic plants

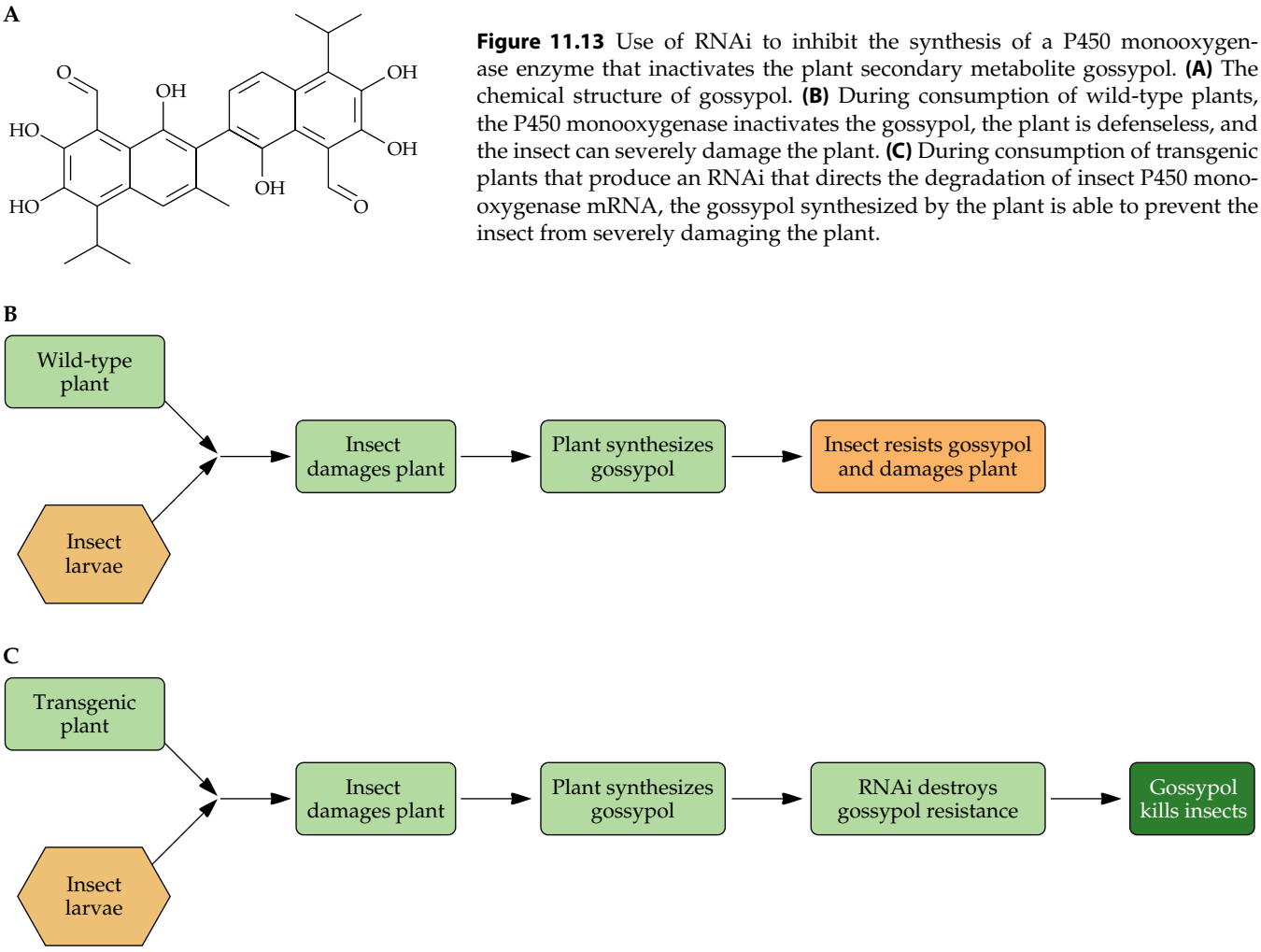


Figure 11.13 Use of RNAi to inhibit the synthesis of a P450 monooxygenase enzyme that inactivates the plant secondary metabolite gossypol. **(A)** The chemical structure of gossypol. **(B)** During consumption of wild-type plants, the P450 monooxygenase inactivates the gossypol, the plant is defenseless, and the insect can severely damage the plant. **(C)** During consumption of transgenic plants that produce an RNAi that directs the degradation of insect P450 monooxygenase mRNA, the gossypol synthesized by the plant is able to prevent the insect from severely damaging the plant.

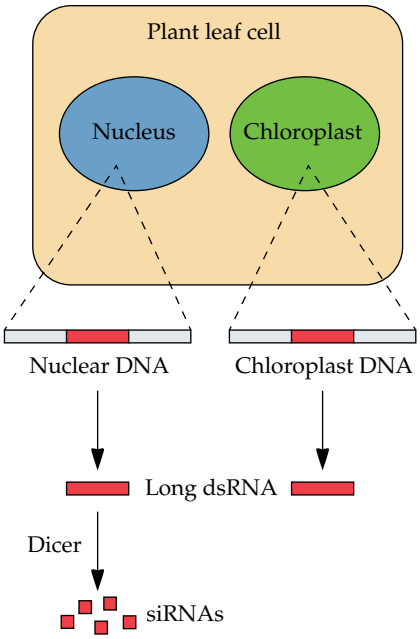


Figure 11.14 A plant leaf cell showing the plant cell nucleus and a single chloroplast (not drawn to size). In each case, a section of the nuclear or chloroplast DNA is shown (in gray) with an RNAi transgene (in red) in the middle. In both cases, the RNAi transgene is converted into a long double-stranded RNA (dsRNA). In the nucleus, the dsRNA is processed by Dicer to form siRNAs.

that produced long dsRNA from which siRNAs are derived (see chapter 6) were much more effective at protecting plants against insect predation. Moreover, it was observed that long dsRNAs that were produced in plant chloroplasts were more effective than dsRNAs produced in the cytoplasm. This is because chloroplasts lack the cellular machinery to process interfering RNAs, so that the dsRNAs produced in chloroplasts were not prematurely cleaved by the enzyme Dicer to form siRNA (Fig. 11.14). While the precise reason why dsRNA is more effective than siRNA is not definitively known, it may be that dsRNA is more stable than siRNA.

Preventing the Development of *B. thuringiensis*-Resistant Insects

There is little doubt that insects have the genetic potential to develop resistance to *B. thuringiensis* insecticidal toxins, and the more that *B. thuringiensis* insecticidal toxins are used, the greater the likelihood that populations of target insects will accumulate resistant individuals. Numerous experimental strategies have been devised to prevent transgenic plants that express the *B. thuringiensis* protoxin gene from acting as selection agents for resistant insects. Some of these strategies include the following.

- **Spatial refugia:** Using spatial-refuge strategies (refugia). In this approach, a certain fraction of each farmer's land, generally around 20%, is planted with a nontransgenic crop, while the remaining 80% of the land is planted with a transgenic version of the same crop expressing a high level of the *B. thuringiensis* insecticidal protoxin (Fig. 11.15). The idea behind this strategy is that the very small number of insects that are able to survive on the transgenic insecticidal crop—a high dose of toxin kills ~99.9% of susceptible insects—will mate with the much larger number of toxin-sensitive insects from the nontransgenic crop so that the next generation will contain mostly homozygous sensitive insects and a very small number of hemizygous sensitive insects. Thus, the gene for resistance is effectively diluted, and the pest population remains sensitive to the insecticide. This approach has been used in the field for many years, with all the available evidence indicating that little to no resistance to any *B. thuringiensis* insecticidal toxins has developed.
- **Refugia-like strategy:** A variation of the above-described refugia strategy includes crossing transgenic plants that synthesize a *B. thuringiensis* insecticidal protoxin with nontransgenic plants that do not produce the protoxin and then crossing the resultant first-generation (F1) hybrid plants and sowing the second-generation (F2) seeds. With this strategy, in the field there will be a random mixture of plants that produce the protoxin and those that do not, and the ratio of protoxin-producing plants to nonproducing plants will be 3:1. Thus, 25% of the plants will be insect sensitive, while 75% will be insect resistant (Fig. 11.16). When this strategy was used in the field in China with transgenic cotton for a period of 11 years, insect resistance to the toxin did not develop. The advantage of this strategy compared to the refugia strategy *per se* is that, once the F2 seeds have been developed, only one group of seeds needs to be prepared and planted.
- **Gene stacking:** Using two or more different *B. thuringiensis* insecticidal toxins in the same transgenic plant (sometimes called gene

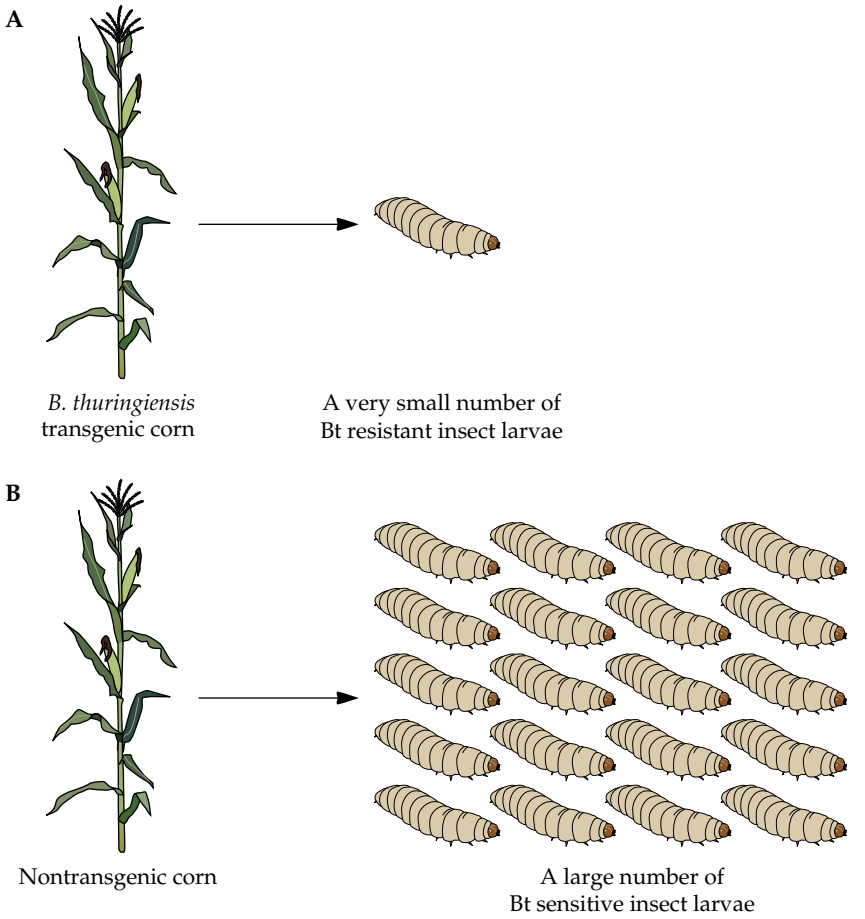


Figure 11.15 A refugia. In this case, the transgenic crop expressing a *B. thuringiensis* (Bt) insecticidal protoxin is corn. The very small number of insects that survive on the transgenic insecticidal crop **(A)** will mate with the much larger number of toxin-sensitive insects from the nontransgenic crop **(B)** with the result that the next generation of insects will be mostly homozygous and sensitive to Bt, with only a very small number of heterozygous sensitive insects.

stacking or pyramiding) or fusing portions of the active regions of two different toxin genes to generate novel hybrid protein insecticidal toxins (Box 11.2). This approach assumes that the different toxins bind to different insect midgut receptor proteins and that it is extremely unlikely that mutations in two different receptors will occur at the same time. This approach has been found to be effective in the field, especially when it is combined with spatial refugia. A variation on the stacking strategy might include the use of two *B. thuringiensis* insecticidal toxins, one of which has had its toxin gene modified so that it binds to a different receptor than the other toxin (Fig. 11.17).

- **Different toxins:** Transforming plants with both a *B. thuringiensis* insecticidal toxin gene and another form of biological insecticide (e.g., an α-amylase inhibitor gene). This strategy assumes that resistance to two very different control methods is much less likely to develop simultaneously.

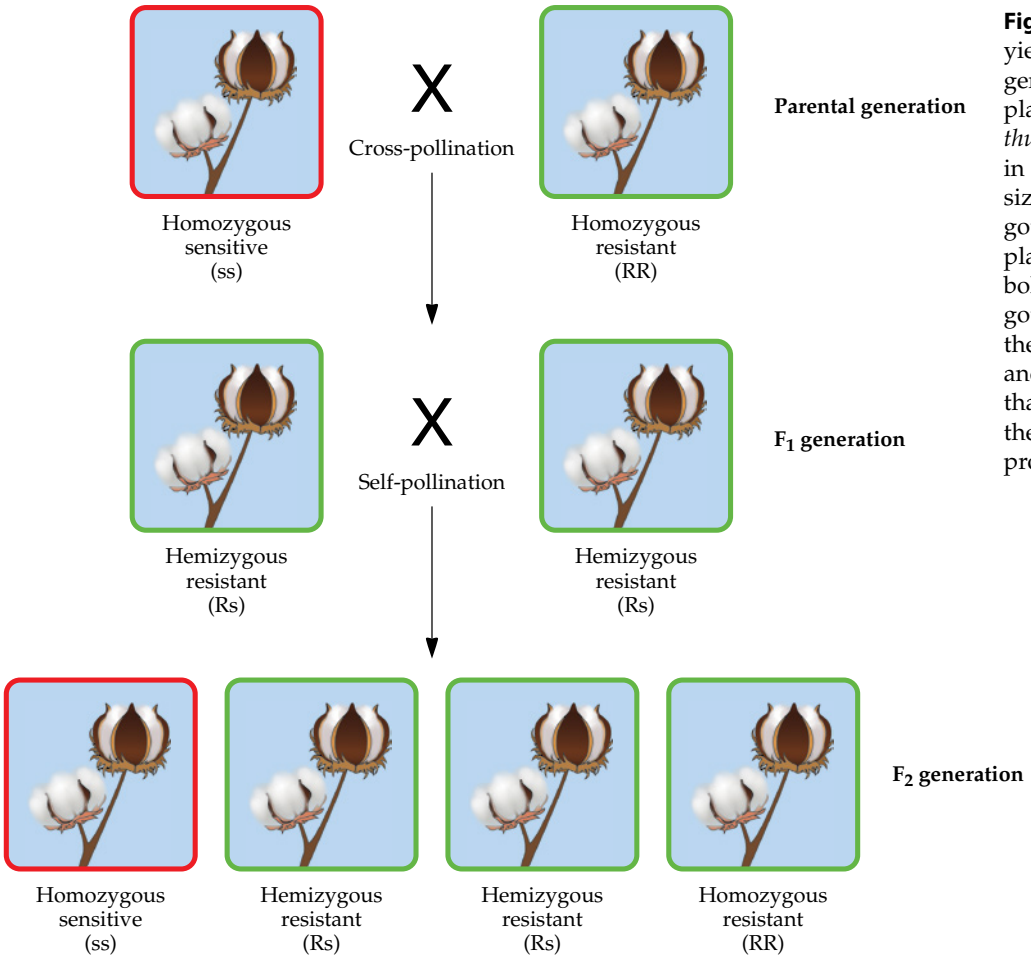


Figure 11.16 A procedure that yields a 3:1 mixture in the F_2 generation of transgenic cotton plants that synthesizes a *B. thuringiensis* insecticidal protoxin and one that does not synthesize the protoxin. The homozygous (RR) protoxin-synthesizing plants are resistant to pink bollworm, while the homozygous plants (ss) that do not synthesize the protoxin are sensitive and the hemizygous (Rs) plants that have only a single copy of the *B. thuringiensis* insecticidal protoxin gene are resistant.

- **Mimic *B. thuringiensis* subsp. *israelensis* behavior:** In contrast to what has been observed with other strains of *B. thuringiensis*, no instances of field resistance of mosquitoes to *B. thuringiensis* subsp. *israelensis* have ever been reported. This lack of insect resistance may reflect the fact that, in addition to synthesizing at least three different Cry proteins—Cry4A, Cry4B, and Cry11A—*B. thuringiensis* subsp. *israelensis* also produces Cyt1A, a secreted highly hydrophobic endotoxin that is not at all homologous to any of the Cry proteins and has a completely different mode of action. While Cry proteins bind to glycoproteins on the insect midgut epithelial membrane, the primary affinity of Cyt1A is to the lipid component of the membrane, especially the unsaturated fatty acids. Cyt1A acts synergistically with the Cry proteins. In one series of experiments, it was demonstrated that with the addition of the Cyt1A protein, insects that had become resistant to Cry4A, Cry4B, and Cry11A (all of which are encoded by *B. thuringiensis* subsp. *israelensis*) were killed when they were treated with *B. thuringiensis* subsp. *israelensis*. Moreover, following the binding of Cyt1A to the midgut epithelial membrane, the protein can act as a receptor for some of the Cry proteins encoded by *B. thuringiensis* subsp. *israelensis* (Fig. 11.18). The reason that *B. thuringiensis*

Managing Insect Resistance to *B. thuringiensis* by Gene Stacking

box 11.2

Although various strains of *B. thuringiensis* produce hundreds of different bacterial insecticidal proteins, transgenic field crops engineered to produce these proteins have utilized only a few types of insecticidal toxin. This has led to concerns that the expression of a single toxin throughout the growing season could result in insects evolving resistance to the toxin's effects. One of the strategies that has been employed to prevent or delay the development of insect resistance to *B. thuringiensis* insecticidal toxins includes transforming plants with two unrelated *B. thuringiensis* insecticidal toxin genes. Recent data indicate that gene stacking (also called gene pyramiding) of two genes encoding insecticidal proteins with different modes of action significantly delays the development of insect resistance to these insecticidal toxins.

Bollgard II is a strain of genetically engineered cotton that was developed by the Monsanto Corporation to produce both the Cry2Ab2 and Cry1Ac insecticidal proteins. This strain was produced by retransformation of the previously commercialized Bollgard cotton, which produces only the Cry1Ac insecticidal protein. The commercial use of Bollgard cotton began in 1996, and by 2003, it had been grown globally on more than 32 million acres with the benefits of reduced insecticide use, improved control of target insect pests, increased yield, and reduced production costs accruing to farmers. The two insecticidal proteins produced by Bollgard II provide protection against several major lepidopteran pests of cotton, including the cotton bollworm, tobacco budworm, pink bollworm, and armyworm. In addition to an expanded insecticidal range, Bollgard II is expected to significantly delay (or prevent) the development of insect resistance in the field. Nevertheless, the use of Bollgard II requires the concomitant employment of refugia.

More recently (in 2014), Bollgard III was commercialized. In this case, in addition to the Cry2Ab2 and Cry1Ac insecticidal proteins, the transgenic cotton plants also produce the *B. thuringiensis* vegetative insecticidal protein vip3A. The presence of an additional insecticidal protein is designed to provide improved control of beet and fall armyworms. Moreover, it is expected that the transgenes in Bollgard III will eventually be stacked with genes that confer herbicide tolerance as well as insect resistance.

subsp. *israelensis* is a highly effective insect pathogenic bacterium may be because the strain not only carries several insecticidal proteins, but also contains a protein that acts as the receptor for these insecticidal proteins.

- **Increase binding targets:** With this in mind, a fusion protein consisting of an N-terminal *B. thuringiensis* insecticidal toxin and a C-terminal peptide consisting of the nontoxic B-chain of the protein ricin was constructed. Ricin is a protein toxin that is extracted from castor beans. It consists of an A-chain of 267 amino acids that contains the toxin activity and a B-chain of 262 amino acids that is catalytically inactive but serves to mediate entry of the complex into the cytosol. Prior to being insecticidal, the *B. thuringiensis* toxin must bind to a receptor located within the membrane of the insect midgut (Fig. 11.19). Normally, since each insecticidal toxin interacts with a single receptor, the loss or modification of the receptor leads to resistance to the insecticidal toxin. However, since the ricin B-chain binds with very high affinity to *N*-acetylgalactosamine residues (which are often adjacent to the *B. thuringiensis* insecticidal toxin receptor), the fusion protein has two separate and independent means by which it is targeted to the receptor. With this fusion protein, it becomes extremely unlikely that both targeting mechanisms will cease to be effective at the same time. It has been suggested that this approach may be most effective in field situations where it is difficult or impossible to implement a spatial-refuge (refugium) strategy (such as for transgenic rice).

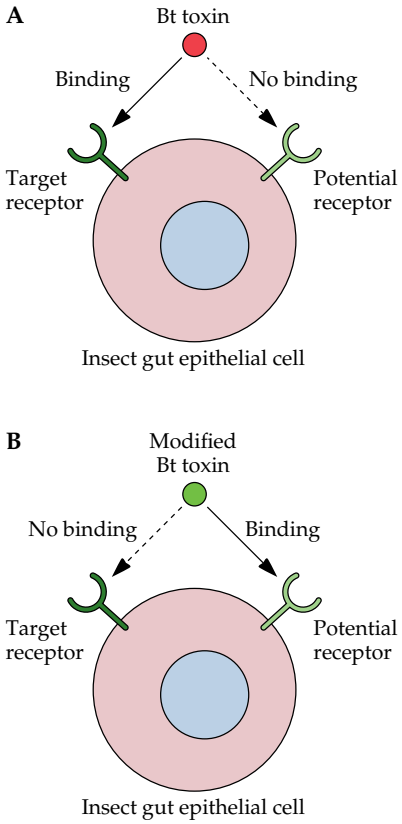


Figure 11.17 Engineering a *B. thuringiensis* (Bt) toxin gene so that instead of the Bt toxin binding to its target receptor on insect gut epithelial cells **(A)**, the modified Bt toxin now binds to a different receptor on the same cells but not to the original receptor **(B)**. When only the native form of the Bt toxin is used, it is possible for a mutation to occur within the target receptor that prevents the native form of the Bt toxin from binding, rendering it completely ineffective. On the other hand, when both the native and modified forms of a particular Bt toxin are added at the same time, i.e., stacked, it is extremely unlikely that both receptors will be mutated at the same time.

- **Use chemical insecticide with a *B. thuringiensis* toxin:** The idea here is that almost no insect survives these two very different treatments, and resistance does not develop to either. In fact, transgenic plants that produce a *B. thuringiensis* insecticidal toxin are commonly treated with chemical insecticides in the field. However, the number of chemical insecticide treatments is significantly reduced when the plants produce a *B. thuringiensis* insecticidal toxin. For example, in Florida, nontransgenic corn plants often require as many as 10 sprayings of chemical insecticides per growing season. Plants that produce a *B. thuringiensis* insecticidal toxin are more likely to be sprayed with chemical insecticides only about three or four times a season.
- **Limit expression time of insecticidal toxin:** The gene for the *B. thuringiensis* protoxin was cloned downstream of the promoter of a gene from tobacco called the pathogenesis-related protein 1a (PR-1a) gene. The expression of the PR-1a gene is part of a natural defense mechanism against various pathogens and is normally induced by any one of a variety of pathogenic organisms or by chemicals, such as salicylic acid and polyacrylic acid. When transgenic plants with the *B. thuringiensis* protoxin gene under the control of the PR-1a promoter are treated with a chemical inducer, they synthesize insecticidal toxin within 1 day of application, which protects the plants against insect attack. Therefore, it is conceivable that the protoxin could be induced by the administration of an inexpensive and safe chemical inducer only when it is required during the growing season, e.g., when insect larvae are feeding. Such periodic production should lower the selection pressure for resistant insects.

The various insect resistance management strategies that have been used up to now appear to have been quite successful. For example, in one large study, researchers monitored the level of resistance of the pink bollworm (*Pectinophora gossypiella*) to *B. thuringiensis* in cotton fields over the course of 8 years, from 1997 to 2004. They found that the frequency of resistance did not change over this period and attributed this result to the use of refugia, the recessive inheritance of the resistance, gene stacking with two unrelated insecticidal toxin genes, the fact that the resistance that developed was incomplete, and the fitness costs associated with the development of insect resistance. It is nevertheless essential that new strategies to prevent the development of insect resistance continue to be developed.

Notwithstanding some of the potential concerns regarding the large-scale employment of transgenic plants expressing *B. thuringiensis* toxins, studies indicate that between 1991 and 2015 there was a major reduction in the global use of chemical insecticides in agriculture with a concomitant increase of crop yield and profits to farmers of 15 to 25%. The amount of land that is devoted to the growth of these transgenic crops continues to increase rapidly worldwide. The types of crops that have been engineered to express a *B. thuringiensis* insecticidal toxin are also continuing to expand. Thus, a responsible approach dictates that we continue to be careful and vigilant in avoiding the development of insect resistance so as not to waste this valuable resource.

Targeting Aphids

Sap-sucking insects (Hemiptera), including aphids, planthoppers, whiteflies, and stink bugs, are not affected by any of the insecticidal proteins produced by *B. thuringiensis*. Moreover, Hemiptera-infested plants have stunted growth, wilted foliage, and significantly decreased crop yields. Depending upon the year, crop damage caused by the soybean aphid costs growers in the United States between $300 million and $700 million in lost productivity annually. In addition to the damage that these insects cause, aphids are hosts for a variety of plant viruses, and aphids carrying plant viruses can inflict severe damage to a crop plant (Fig. 11.20A). To protect plants against the damage caused by aphid predation, scientists constructed transgenic plants that produced a fusion protein consisting of a viral coat protein and a spider neurotoxin (Fig. 11.20B). Aphids typically ingest intact virus when they feed on the phloem (the plant vascular tissue that conducts sugars and other metabolites from the leaves) of infected plants. The ingested virions are eventually released into the insect hemocoel (the insect body cavity where hemolymph circulates). After the virions enter the hemolymph (a fluid in invertebrates equivalent to blood), they circulate and penetrate the salivary glands. This process is limited to only certain combinations of viruses and insects. Thus, fusing the spider neurotoxin with a viral coat protein ensures that the toxin will enter the hemolymph and paralyze and eventually kill the insect.

Resistance to Two Types of Insects

Cotton cultivation is practiced worldwide in more than 70 countries, with the annual crop being worth more than $500 billion. Unfortunately, nearly a fifth of the global cotton crop is lost to insect pests every year even though cotton farmers spend around $1.5 billion per year on chemical insecticides in an effort to control this problem. For the past 10 to 15 years farmers have increasingly utilized transgenic cotton expressing *B. thuringiensis* in an effort to limit insect predation and crop damage. The majority of these transgenic plants include the *B. thuringiensis* protoxin gene under the transcriptional control of the constitutive 35S promoter from *Cauliflower mosaic virus*. However, in one recent report, scientists expressed the

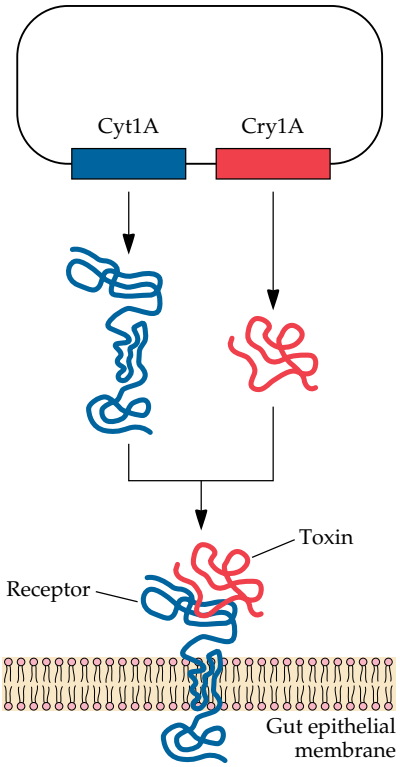
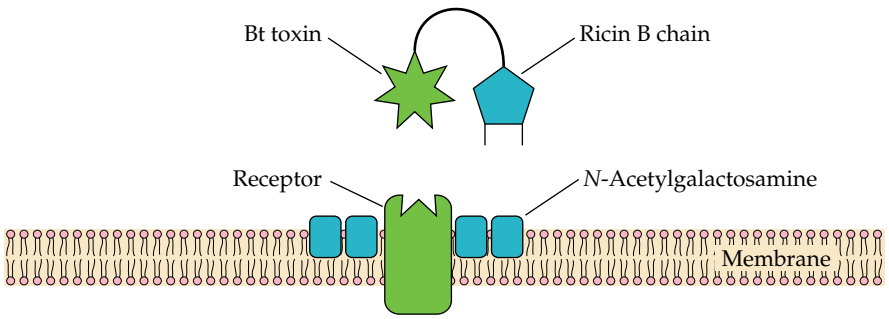


Figure 11.18 An engineered strain of *B. thuringiensis* encoding both Cyt1A (which is from *B. thuringiensis* subsp. *israelensis* and becomes the membrane-bound receptor) and Cry1A (the lepidopteran-specific insecticidal toxin) proteins.

Figure 11.19 A hybrid protein consisting of the Cry1Ac insecticidal toxin protein (at the N terminus), the B-chain of ricin (at the C terminus), and an insect midgut insecticidal-protein receptor. The Cry protein recognizes and binds to the insect midgut receptor, while the B-chain of ricin acts as an *N*-acetygalactosamine-specific lectin that binds very tightly to these residues, which are located adjacent to the receptor. Bt, *B. thuringiensis*.



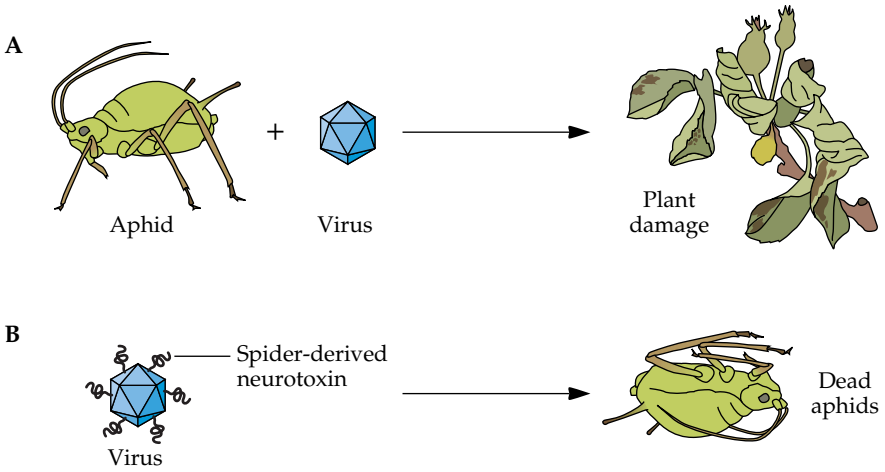


Figure 11.20 Schematic representation of **(A)** an aphid carrying a plant virus significantly damaging a plant and inhibiting its growth and **(B)** a fusion protein of a plant virus coat protein and a spider-derived neurotoxin produced by transgenic plants that targets and kills aphids (and thereby prevents them from damaging plants).

B. thuringiensis cry1Ac gene in transgenic cotton plants under the control of the wound-inducible AoPR1 promoter from *Asparagus officinalis*. Use of the AoPR1 promoter led to higher levels of *cry1Ac* gene expression than the 35S promoter, and insecticidal protoxin expression was limited to periods when plant leaves were being wounded by feeding insect larvae. Limited periods of protoxin expression potentially decrease the possibility of insect pests such as *Spodoptera littoralis* (African cotton leafworm) developing resistance to Cry1Ac. In addition, cotton plants were also transformed with the *G. nivalis* lectin gene (from snowdrop bulbs), *GNA*, under the control of the 35S promoter. The expression of this lectin (a carbohydrate binding protein) causes partial resistance to sap-sucking hemipteran pests and also acts synergistically with Cry1Ac.

Virus Resistance

Plant viruses often cause considerable crop damage and significantly reduce yields. Therefore, in the absence of effective chemical treatments, plant breeders have attempted to transfer naturally occurring virus resistance genes from one plant strain (cultivar) to another. However, resistant cultivars often revert to virus sensitivity, and resistance to one virus does not necessarily confer resistance to other, similar viruses. Natural virus resistance can be achieved in different ways: viral transmission can be blocked, establishment of the virus in host plants can be prevented, or viral symptoms can be bypassed or resisted. On the other hand, genetic engineering has been used to develop nonconventional types of virus-resistant transgenic plants.

Viral Coat Protein-Mediated Protection

When transgenic plants express the gene for a coat protein (which usually is the most abundant protein of a plant virus particle) of a virus that normally infects those plants, the ability of the virus to subsequently infect

the plants and spread systemically is often greatly diminished. For a long time, the precise mechanism by which the presence of coat protein genes inhibits viral proliferation was not understood; however, it is now thought that it likely works through RNAi. Moreover, the antiviral effect occurs early in the viral replication cycle and, as a result, prevents any significant amount of viral synthesis. This feature is an advantage because it decreases the probability of selecting for spontaneous viral mutants that can overcome this resistance and replicate in the presence of the viral coat protein gene. The viral coat protein gene approach has been used to develop virus-resistant transgenic plants for several different crops (Table 11.3). Although complete protection is not usually achieved, high levels of virus resistance have been reported. In addition, a coat protein gene from one virus sometimes provides tolerance for a broad spectrum of unrelated viruses. The utility of this strategy is supported by the observation that transgenic plants that encode viral coat proteins do as well in field trials as in the laboratory studies.

In both eukaryotes and prokaryotes, an RNA molecule that is complementary to a normal gene transcript (mRNA) is called antisense RNA. The mRNA, being translatable, is considered to be a sense RNA. The presence of antisense RNA can decrease the synthesis of a protein by forming a duplex molecule with the normal sense mRNA, thereby preventing it from being translated. The antisense RNA–mRNA duplex is also rapidly

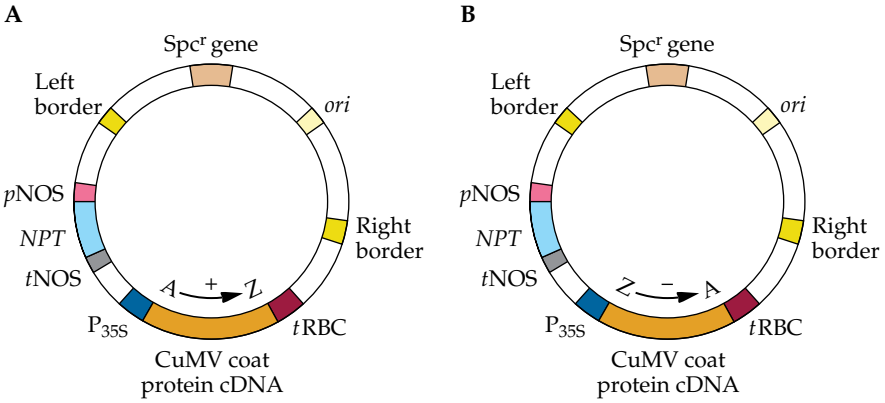
Table 11.3 Some transgenic plants engineered to have viral coat protein-mediated protection against viral infection

Viral source of coat protein	Transgenic plant(s)
Alfalfa mosaic virus	Alfalfa, tobacco, tomato
Arabis mosaic virus	Tobacco
Beet necrotic yellow vein virus	Sugar beet
Cucumber mosaic virus	Cucumber, tobacco
Cymbidium ringspot virus	Tobacco
Grapevine chrome mosaic virus	Tobacco
Maize dwarf mosaic virus	Sweet corn
Papaya ringspot virus	Papaya, tobacco
Plum pox virus	Tobacco
Potato aucuba mosaic virus	Tobacco
Potato leafroll virus	Potato
Potato virus S	Potato
Potato virus X	Potato, tobacco
Potato virus Y	Potato, tobacco
Rice stripe mosaic virus	Rice
Soybean mosaic virus	Tobacco
Tobacco etch virus	Tobacco
Tobacco mosaic virus	Tobacco, tomato
Tomato mosaic virus	Tomato
Tomato rattle virus	Tobacco
Tomato streak virus	Tobacco
Tomato spotted wilt virus	Tobacco
Watermelon mosaic virus 2	Tobacco
Zucchini yellow mosaic virus	Muskmelon, tobacco

degraded, a response that diminishes the amount of that mRNA in the cell. Theoretically, it should be possible to prevent plant viruses from replicating and subsequently damaging plant tissues by creating transgenic plants that synthesize antisense RNA that is complementary to viral coat protein mRNA.

In one of many studies, the efficacies of the viral coat protein gene and antisense RNA approaches were compared by cloning the cDNA for the coat protein of *Cucumber mosaic virus* (CuMV) into tobacco plants in two orientations (sense and antisense; one orientation per plant) and then testing transgenic plants for sensitivity to viral infection (Fig. 11.21). The genome of CuMV consists of three separate single-stranded pieces of RNA, each coding for a specific viral protein. *In vivo*, one of these pieces, RNA3, is processed to remove a portion of its sequence, thereby generating RNA4, which encodes the viral coat protein. To create transgenic plants that either produced normal mRNA and expressed the viral coat protein or produced its antisense RNA, viral RNA4 was isolated and converted into double-stranded cDNA, which was inserted into Ti plasmid binary cloning vectors in both orientations. Each oriented sequence was placed under the control of the 35S promoter sequence from *Cauliflower mosaic virus* and the termination-regulatory sequences from the plant gene for the small subunit of ribulose bisphosphate carboxylase. Subsequently, these plasmids were used to generate transgenic plants carrying the cDNA sequence in one of the two possible orientations. The transgenic tobacco plants that expressed the CuMV coat protein were protected from viral-particle accumulation and did not show symptoms of viral infection, regardless of whether the inoculum of the challenge virus was high

Figure 11.21 Ti plasmid binary cloning vectors containing either the protein-producing sense **(A)** or the RNA-producing antisense **(B)** orientation of the CuMV coat protein cDNA. Each cDNA sequence is under the control of the 35S promoter (P_{35S}) from *Cauliflower mosaic virus* and the transcription terminator–polyadenylation site (*t*RBC) from the gene for the small subunit of ribulose bisphosphate carboxylase. The vector also contains a neomycin phosphotransferase (*NPT*) gene under the control of nopaline synthase transcription signals (*p*NOS and *t*NOS), an Spc^r gene, a T-DNA right-border sequence, a T-DNA left-border sequence, and a broad-host-range origin of DNA replication (*ori*). The protein-producing sense (+) orientation is shown by the A→Z arrow, and the RNA-producing antisense (–) orientation is shown by the Z→A arrow.

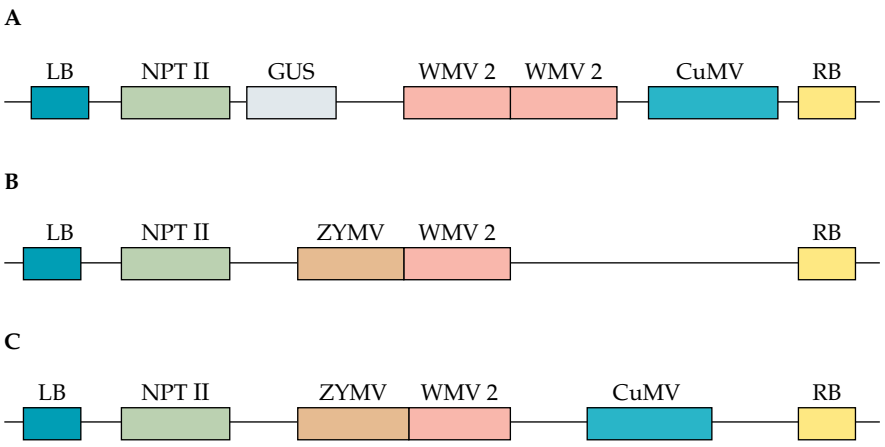


or low, whereas the antisense orientation construct protected transgenic plants only against low viral doses.

Similar results were obtained by several groups of scientists who constructed transgenic plants that synthesize antisense RNA copies of viral coat protein genes and tested whether these plants can withstand a viral challenge. In all instances, the plants were protected against the invading virus only when low concentrations of the virus were used. At high virus concentrations, the plants were damaged by the virus. Antisense RNA copies of viral coat protein genes also generally afforded a much lower level of protection to transgenic plants than did sense versions of the viral coat protein genes.

Often, field crops are exposed to several different viruses, any one of which may damage the plant and lower the final yield. Ideally, transgenic plants should be resistant to more than one virus. With this in mind, Ti plasmid binary vectors expressing one or more coat protein genes for CuMV, *Zucchini yellow mosaic virus*, and *Watermelon mosaic virus 2* were used to transform yellow crookneck squash (*Cucurbita pepo*) plants (Fig. 11.22). Transgenic plants that contained the coat protein genes from all three viruses were resistant to damage by all three viruses under laboratory conditions. Initially, transgenic plants expressing coat protein genes for *Zucchini yellow mosaic virus* and *Watermelon mosaic virus 2* were tested under field conditions by using aphids to naturally transmit these viruses to developing plants. The transgenic plants that expressed both coat protein genes were completely resistant to infection when the two viruses were transmitted at the same time (Fig. 11.23). On the other hand, while transgenic plants expressing only one of the two viral coat proteins were delayed in succumbing to the viruses in comparison with the nontransformed control plants, all of these plants eventually

Figure 11.22 **(A)** A T-DNA construct with a neomycin phosphotransferase (NPT II) gene as a selectable marker, a β-glucuronidase (GUS) gene as a reporter gene, two copies of the coat protein gene from *Watermelon mosaic virus 2* (WMV 2), and the coat protein gene from CuMV. The left and right borders of the T-DNA are indicated by LB and RB, respectively. **(B)** Similar to panel A without CuMV and GUS, with one copy of WMV 2, and with the coat protein gene from *Zucchini yellow mosaic virus* (ZYMV). **(C)** Same as panel B with the addition of CuMV. All of the genes in these constructs include both promoters and transcription terminator regions.



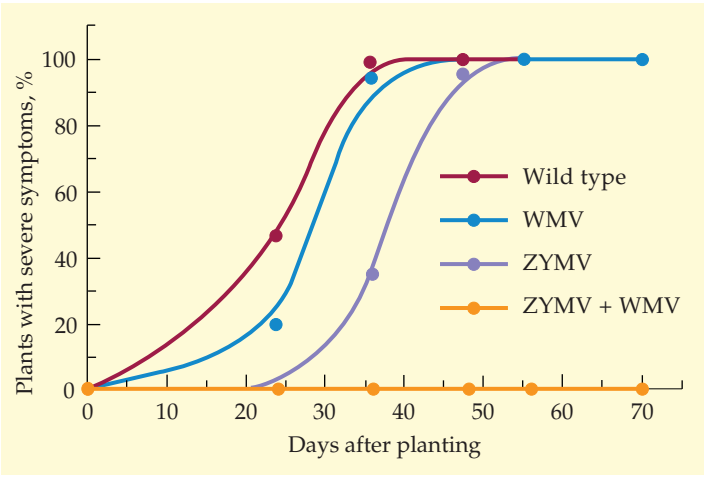


Figure 11.23 Disease frequencies in transgenic and nontransformed (wild-type) yellow crookneck squash in the field. Aphids were used to transmit a mixture of *Zucchini yellow mosaic virus* (ZYMV) and *Watermelon mosaic virus 2* (WMV) to the squash plants. Data from Fuchs and Gonsalves, *Bio/Technology* **13:**1466–1473, 1995.

developed severe symptoms of viral disease, making them unfit for sale to consumers.

Transgenic squash plants that express viral coat protein genes for *Zucchini yellow mosaic virus*, *Watermelon mosaic virus 2*, and CuMV are commercially available. The increase in plant yield resulting from virus protection depends upon which viruses the plants are challenged with, how great the viral pressure is, and the time of the growing season. Despite these variables, one study estimated that transgenic squash with resistance to these three viruses, subjected to severe viral pressure, could produce as much as a 50-fold increase in marketable squash over nontransgenic varieties. Clearly, using more than one viral coat protein gene is an effective strategy that should be useful in developing a range of transgenic plants that are resistant to all of the major viruses that normally inhibit their growth and development. However, it must be borne in mind that to satisfy the variety of consumer tastes, there are a large number of squash varieties, all of which are potentially susceptible to these viruses, and with this approach, all of them would have to be genetically engineered in a similar manner. In addition, those transgenic lines that are resistant to three different viruses are still susceptible to *Papaya ringspot virus* type W, so "complete" viral protection of summer squash will require the introduction of the viral coat protein for this virus as well.

The phenomenon of using a plant-encoded viral gene to disrupt the virus life cycle and thereby confer resistance to the virus is sometimes called homology-dependent gene silencing (formerly called cosuppression). Perhaps the most successful employment of the viral coat protein strategy to prevent damage from a plant virus is the use of the coat protein of *Papaya ringspot virus* to protect papaya plants against the virus. In fact, these transgenic papaya plants are widely credited with having saved the Hawaiian papaya industry (Box 11.3).

box 11.3

Saving the Hawaiian Papaya Industry

Papaya (*Carica papaya*) is an important tropical fruit crop grown in Brazil, India, Mexico, and Thailand (as well as several other tropical countries) and in Hawaii. It is valued as a healthy food because it is rich in vitamins C and A and because it contains large amounts of the proteolytic enzyme papain, which is potentially an aid to the digestion of proteins. The papaya tree is relatively easy to grow from seeds, and the first fruit can be harvested a few months after the seeds are sown. Thereafter, fruit is produced continuously on a year-round basis.

Papaya ringspot virus (PRSV) is a potyvirus that is transmitted by aphids. In addition to papaya, the virus also infects several cucurbits (e.g., squash, watermelon, and cucumber). The viral RNA genome consists of approximately 10,000 nucleotides, which exist as a single strand. The genome is monocistronic, so it is expressed as a single large polypeptide that is subsequently processed into several different functional proteins. When PRSV was discovered to be present on the Hawaiian island of Oahu in the late 1950s, the papaya industry was moved over the course of several years to the area of Puna on the island of Hawaii (sometimes called the Big Island). However, by the 1970s, PRSV was also detected in Puna.

Subsequently, a team of scientists, headed by Dennis Gonsalves, developed (from the commercial cultivar called Sunset) a line of transgenic papaya (called 55-1) that expressed the coat protein gene of PRSV. At the time that this work was undertaken, papaya had not previously been genetically transformed, and thus, the coat protein was introduced by microprojectile bombardment. The homozygous (for the coat protein gene) version of this strain (now called UH SunUP) was shown to be highly resistant to PRSV under field conditions. The UH SunUP strain was then crossed with the nontransgenic Kapoho strain (which is the dominant strain of papaya grown in Hawaii) to create a hybrid strain called UH Rainbow. Importantly, while both the transgenic and nontransgenic papaya strains could be infected with PRSV, the transgenic strains have remained resistant to the virus for up to (at least) 3 years.

The main markets for Hawaiian papaya are the mainland United States, Canada, and Japan. These transgenic plants have now been approved for use in all three of these countries. The success with papaya demonstrates that the approach of using coat protein-mediated protection may be both a safe and an efficacious way to develop protection from a phytopathogenic virus for a range of crops.

Protection by Expression of Other Genes

Engineered resistance to plant viruses—generally as a result of expressing a viral coat protein or another viral gene in the transgenic plant—is usually an effective strategy only against closely related viruses. However, since there are a large number of different viruses that could potentially infect a crop, it would be advantageous if plants could be engineered to be resistant to a broad spectrum of viruses.

Single-Chain Antibodies

One way to protect plants against viral infection is to engineer the plants to produce antibodies that are directed against the invading viruses. This was done, with some success, by expressing single-chain Fv antibodies (see chapter 5) directed against the coat protein of the *Tobacco mosaic virus* in tobacco plants. However, because of the variability between coat proteins from different viruses, this strategy is not useful for providing broad-range resistance against several different viruses.

The majority of plant viruses are RNA viruses, and many of them contain positive-stranded RNA as the genetic material. These viruses all encode RNA-dependent RNA polymerases that are essential for their replication. Thus, a single-chain Fv antibody that recognizes epitopes that are common among the RNA-dependent RNA polymerases from several different viruses should be an effective means of inhibiting the replication of

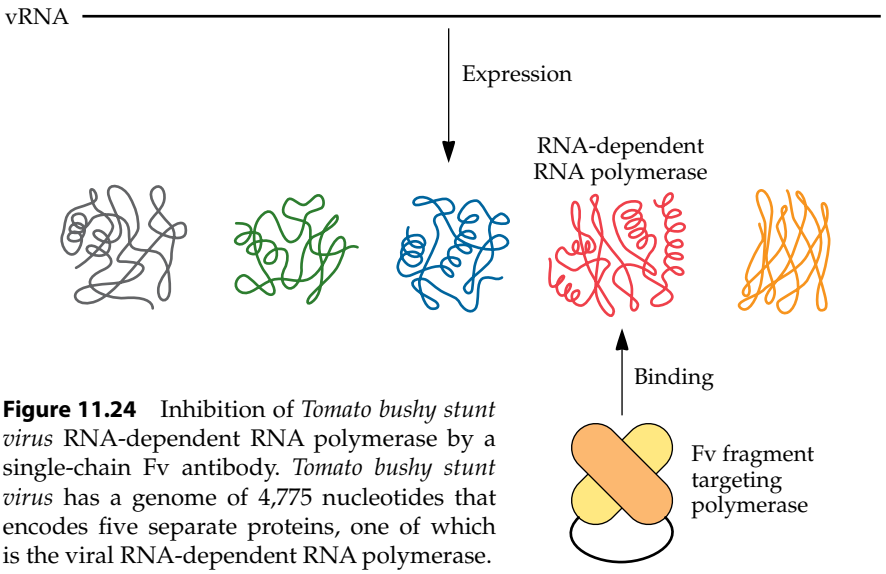


Figure 11.24 Inhibition of *Tomato bushy stunt virus* RNA-dependent RNA polymerase by a single-chain Fv antibody. *Tomato bushy stunt virus* has a genome of 4,775 nucleotides that encodes five separate proteins, one of which is the viral RNA-dependent RNA polymerase.

all of these viruses, thereby making transgenic plants that express these single-chain Fv antibodies resistant to these viruses (Fig. 11.24). That is because the antibody fragment can bind to RNA polymerases and thereby block their activities. A high level of antibody expression is not required because RNA-dependent RNA polymerases are found in only low concentrations in virus-infected cells. When two phage display libraries were screened against purified denatured fragments of the RNA-dependent RNA polymerase of *Tomato bushy stunt virus*, the three single-chain Fv antibodies that displayed the highest affinity were isolated and characterized. A cDNA encoding one of these single-chain Fv antibodies was expressed in *Nicotiana benthamiana* (a close relative of tobacco, native to Australia). The resultant transgenic plants were significantly protected against both *Tomato bushy stunt virus* and *Cucumber necrosis virus* and partially resistant to *Turnip crinkle virus* and red clover necrotic virus. In addition, the single-chain Fv antibody directed against *Tomato bushy stunt virus* bound to the RNA-dependent RNA polymerase of hepatitis C virus, which is much more distantly related. This work represents an important first step in using a simple and straightforward strategy to develop transgenic plants that are resistant to a wide range of viruses.

Micro-RNAs

One approach to developing plants that are resistant to a range of viruses might include engineering the plants to produce micro-RNAs (miRNAs) that interfere with viral replication by targeting the viral RNA (or the viral mRNA) for degradation. miRNAs resemble siRNAs except that miRNAs originate from regions of RNA transcripts that fold back on themselves, while siRNAs originate from longer regions of dsRNAs. In a series of experiments, starting with a 273-nucleotide precursor of a naturally occurring plant miRNA, scientists used PCR to replace a small portion of the existing sequence so that the precursor could be processed to yield a 20- to 24-nucleotide-long miRNA that was complementary to viral RNA (Fig. 11.25). The newly synthesized artificial miRNA (amiRNA)

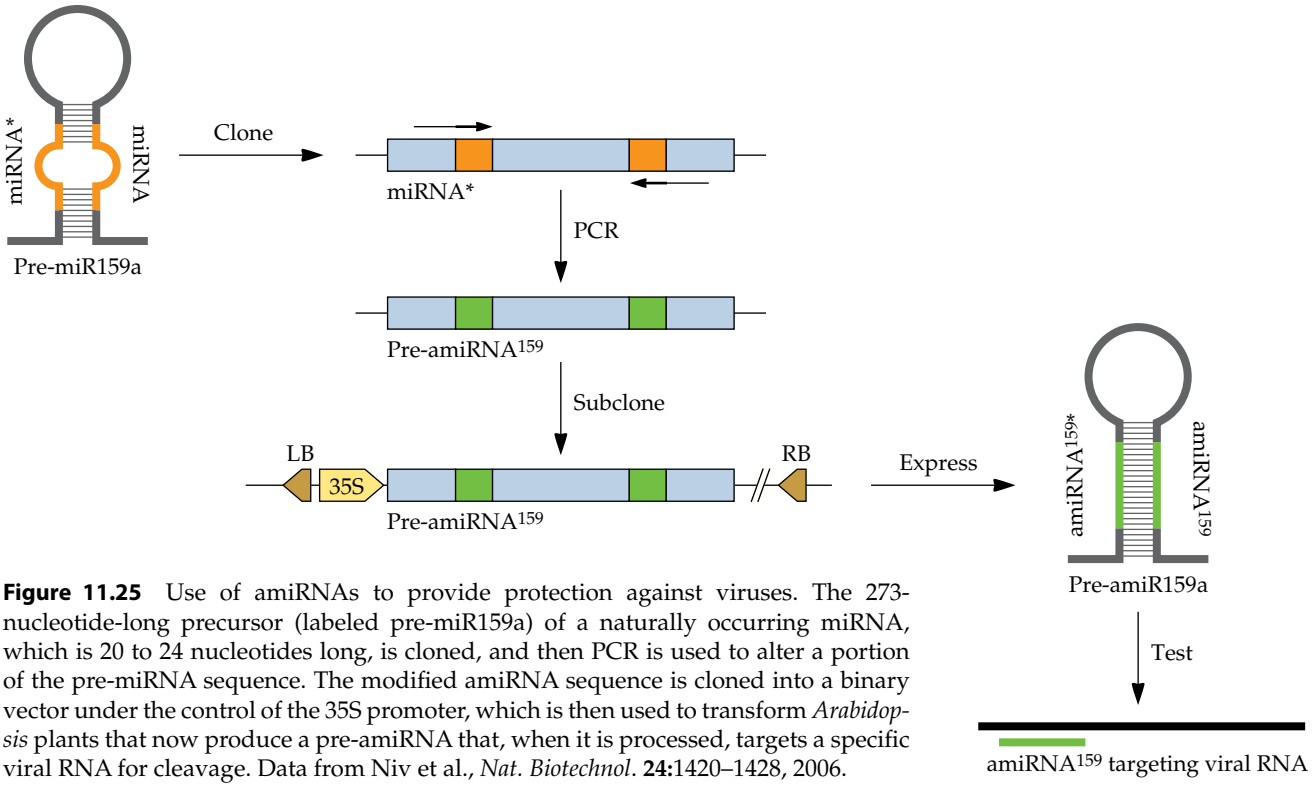


Figure 11.25 Use of amiRNAs to provide protection against viruses. The 273-nucleotide-long precursor (labeled pre-miR159a) of a naturally occurring miRNA, which is 20 to 24 nucleotides long, is cloned, and then PCR is used to alter a portion of the pre-miRNA sequence. The modified amiRNA sequence is cloned into a binary vector under the control of the 35S promoter, which is then used to transform *Arabidopsis* plants that now produce a pre-amiRNA that, when it is processed, targets a specific viral RNA for cleavage. Data from Niv et al., *Nat. Biotechnol*. **24:**1420–1428, 2006.

became part of an RNA-induced silencing complex (see chapter 6) in which the viral RNA (mRNA) was specifically bound and cleaved. It is also possible to clone two or more different pre-amiRNAs in tandem—this was done for *Turnip yellow mosaic virus* and *Turnip mosaic virus*—so that plants transformed with this construct become resistant to two separate viruses. This approach can be made even more effective by targeting more than one portion of each viral RNA using several pre-amiRNAs. Despite its intriguing possibilities, this system still requires a considerable amount of development before it is shown to be effective under field conditions.

CRISPR/Cas

Geminiviruses are a large family (7 genera and 325 species) of circular single-stranded DNA (ssDNA) viruses that replicate within the nuclei of plant cells. These viruses cause significant damage to several dicotyledonous crop plants, including tomato, cassava, cotton, sugar beet, and pepper. These viruses are typically transmitted by leafhoppers, small insects that suck plant sap from grasses, shrubs, and trees, and similar insects. Geminiviruses are nonenveloped viruses with a twinned capsid structure and a relatively small genome (Fig. 11.26). During geminivirus replication, the ssDNA is converted to a double-stranded DNA (dsDNA) intermediate, from which new ssDNA is generated by rolling-circle replication. One group of scientists used CRISPR/Cas9 (see chapter 2) to specifically target the dsDNA of the beet severe curly top virus (a geminivirus) as a way of inhibiting virus replication and conferring virus resistance to host plants (Fig. 11.26). These scientists created transgenic *A. thaliana* and *N. benthamiana* plants that targeted the dsDNA of the replicating viral genome

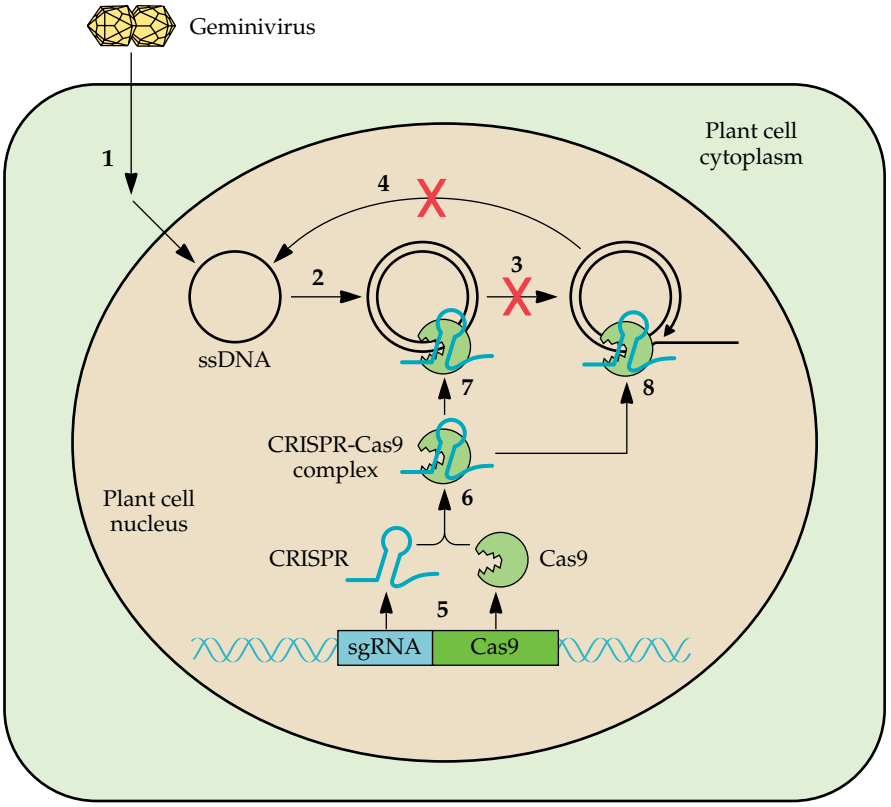


Figure 11.26 Use of CRISPR-Cas9 to provide protection against geminiviruses. (1) The geminivirus single-stranded circular DNA genome enters the plant cell nucleus. (2) The double-stranded DNA replicative form is synthesized. (3 and 4) Additional copies of the single-stranded DNA genomic DNA are produced by rolling-circle replication. (5) sgRNA and Cas9 protein are synthesized within the nucleus from exogenous DNA that had previously been inserted into the plant genomic DNA. (6) The CRISPR-Cas9 complex is assembled. (7 and 8) The CRISPR-Cas9 complex can cleave the target sequence in the double stranded viral replicative form, either before or during DNA replication. Cleavage of the double-stranded viral replicative form thereby prevents the formation of the single-stranded viral DNA (steps 3 and 4).

and showed that, in these systems, both plants synthesized dramatically reduced levels of viral particles. Moreover, these researchers were able to identify several efficient antivirus sgRNAs that recognize sites within the viral genome. While the effectiveness of this approach remains to be demonstrated in the field, this promising result could be used to protect plants against a range of geminiviruses.

Plant RNA potyviruses (nonenveloped filamentous viruses, ~750 nm in length and ~15 nm in diameter with a single-stranded positive-sense RNA genome of ~10 kb encoding 11 proteins) replicate using the mechanisms and components of the plant host, including translation initiation and elongation factors, and ribosomes. In this regard, eukaryotic translation initiation factors play a unique role (Fig. 11.27). With most plant viruses, the eIF4F complex (consisting of eIF4E, eIF4G, and eIF4A) together with other host factors, including the polyA-binding protein (PABP), initiate translation of the viral mRNA by binding to the viral 5′ m7G cap structure and

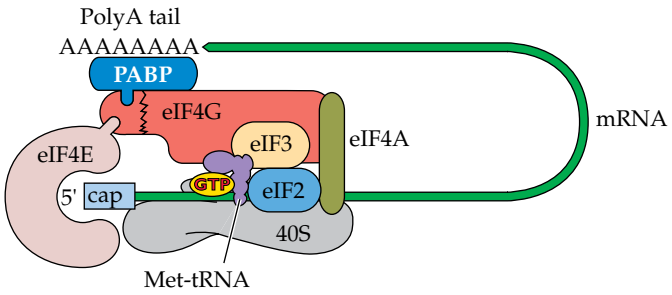


Figure 11.27 Schematic representation of the formation of a eukaryotic translation initiation complex. Potyvirus mRNA does not contain a 5′ cap as part of its mRNA so that eIF4E would not be able to form a translation initiation complex with viral mRNA unless eIF4E first formed a complex with viral protein VPg and the complex then bound to the 5′ end of the mRNA, allowing it to be translated. The unlabeled colored region to the right of the region labeled GTP represents the initiator Met-tRNA. Abbreviations: PABA, polyA binding protein; 40S, the small subunit of the plant ribosome.

the 3′ polyA tail. However, potyvirus mRNAs do not contain the 5′ m7G structures (present on nearly all other eukaryotic mRNAs) to which factor eIF4E normally binds. In order to bind to the 5′ end of the viral mRNA, the eIF4E protein must first interact with potyvirus viral protein VPg. Based on the prior observation that several plant varieties that contain mutant forms of eIF4E are resistant to infection by potyviruses, one group of scientists, using genetic manipulation by CRISPR/Cas9, constructed cucumber plants that were mutated in the genes for eIF4E. These mutated plants (which were homozygous for the introduced mutants) were then immune to *Cucumber vein yellowing virus*, *Zucchini yellow mosaic virus*, and *Papaya ringspot mosaic virus*. In contrast, the heterozygous mutant plants and the wild-type plants were susceptible to these viruses. These mutant genome-edited cucumber plants, which behaved normally in every way except for their susceptibility to potyviruses, may be considered to not be genetically engineered as far as some regulatory agencies are concerned because they do not contain any foreign DNA.

Ribonuclease

Viral infections are one of the main factors limiting soybean yield and quality worldwide. For example, *Soybean mosaic virus* (SMV), the most prevalent and destructive viral pathogen of soybean, is estimated to cause ~10 to 35% yield losses of the global soybean crop each year. SMV-infected plants typically display mosaic patterns, chlorosis, curling, leaf necrosis, malformed pods, mottled seeds, and reduced size. SMV is a member of the plant virus genus *Potyvirus*. The primary mode of transmission of this virus is by aphids, and the main method for the control of this virus is the application of pesticides that control aphid proliferation. To prevent the virus from replicating, soybean plants were engineered to constitutively synthesize the PAC1 RNase from the yeast *Schizosaccharomyces pombe* (which is used in traditional brewing). This simple strategy is effective in protecting plants from damage caused by the virus because the PAC1 RNase targets the dsRNA that is an essential component of the replication complex of this and other eukaryotic positive-strand RNA viruses

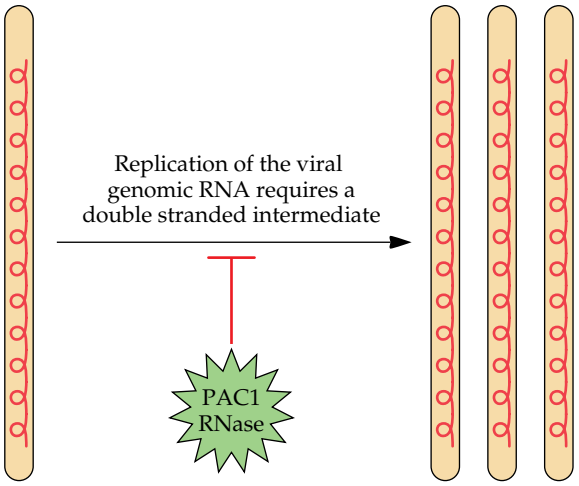


Figure 11.28 Inhibition of the replication of the filamentous *Soybean mosaic virus* by transgenic soybean plants that have been engineered to synthesize PAC1 RNase.

(Fig. 11.28). The results that have been reported to date include a successful field test of PAC1 RNase transgenic plants where the transgenic plants displayed a significantly lower disease index than that of nontransgenic plants over three successive soybean plant generations. However, whether this approach is as effective with other positive-strand RNA viruses remains to be determined.

Herbicide Resistance

A significant fraction of global crop production is lost through weed infestation every year, despite the annual expenditure of ~$35 billion in 2020 on more than 100 different chemical herbicides. In addition, many herbicides do not discriminate weeds from crop plants; others must be applied early, before the weeds take hold; and some persist in the environment. The creation of herbicide-resistant crop plants is one way to overcome some of these drawbacks.

Several different biological manipulations that would cause a crop plant to be herbicide resistant can be envisioned.

1. Inhibit uptake of the herbicide.
2. Overproduce the herbicide-sensitive target protein so that enough of it remains available for cellular functions despite the presence of the herbicide.
3. Introduce a bacterial or fungal gene that produces a protein that is not sensitive to the herbicide but performs the same function as the plant (herbicide-sensitive) protein.
4. Reduce the ability of the herbicide-sensitive target protein to bind to a particular herbicide.
5. Endow plants with the capability to metabolically inactivate the herbicide.

A number of these strategies have been implemented to produce herbicide-resistant transgenic plants (Table 11.4). This approach has been

Table 11.4 Some examples of gene-based herbicide resistance

Herbicide(s)	Mode of development of herbicide resistance
Triazines	Resistance is due to an alteration in the *psbA* gene, which codes for the target of this herbicide, chloroplast protein D-1.
Sulfonylureas	Genes encoding resistant versions of the enzyme acetolactate synthetase have been introduced into poplar, canola, flax, and rice.
Imidazolinones	Strains with resistant versions of the enzyme acetolactate synthetase have been selected in tissue culture.
Aryloxphenoxy-propionates, cyclohexanediones	These herbicides inhibit the enzyme acetyl coenzyme A carboxylase. Resistance, selected in tissue culture, is due either to an altered enzyme that is not herbicide sensitive or to the degradation of the herbicide.
Glyphosate	Resistance is from overproduction of EPSPS, the target of this herbicide. Resistance has been engineered by transforming soybean with the gene for a glyphosate-resistant EPSPS and tobacco with a glyphosate oxidoreductase gene, which encodes an enzyme that degrades glyphosate.
Bromoxynil	Resistance to this photosystem II inhibitor has been created by transforming tobacco and cotton plants with a bacterial nitrilase gene, which encodes an enzyme that degrades this herbicide.
Phenoxycarboxylic acids (e.g., 2,4-D and 2,4,5-T)	Resistant cotton and tobacco plants have been created by transformation with the *tfdA* gene from *Alcaligenes*, which encodes a dioxygenase that degrades this herbicide.
Glufosinate (phosphinothricin)	Over 20 different plants have been transformed with either the *bar* gene from *Streptomyces hygroscopicus* or the *pat* gene from *Streptomyces viridochromogenes*. The phosphinothricin acetyltransferase that these genes encode detoxifies this herbicide.
Cyanamide	Resistant tobacco plants were produced when a cyanamide hydratase gene from the fungus *Myrothecium verrucaria* was introduced. The enzyme encoded by this gene converts cyanamide to urea.
Dalapon	Tobacco plants transformed with a dehalogenase gene from *Pseudomonas putida* can detoxify this herbicide.

so successful that in 2018 more than 50% of the transgenic crops that were planted worldwide were engineered to be herbicide resistant, occupying more than 87 million hectares.

Glyphosate

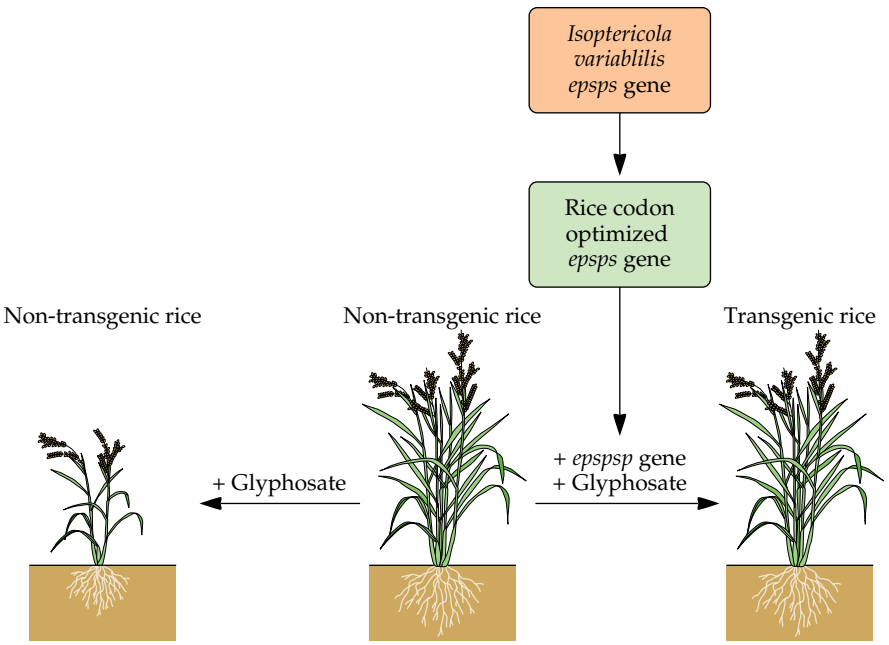
By far the most widely used herbicide is glyphosate, which was first registered in 1974 and is considered to be safe, cheap, effective, and "environmentally friendly" because it is readily degraded to nontoxic compounds in the soil. Glyphosate, trademarked as Roundup by the Monsanto Corporation, inhibits a key enzyme in the shikimate pathway, 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS), which plays an important role in the synthesis of aromatic amino acids in both bacteria and plants. Plants resistant to this herbicide have been developed by putting an EPSPS-encoding gene from a glyphosate-resistant strain of *E. coli* under the control of plant promoter and transcription termination–polyadenylation sequences and cloning the construct into plant cells. Transgenic soybean, corn, canola, tobacco, petunia, tomato, potato, alfalfa, sorghum, sugar beet, Indian mustard, and cotton plants that produce an amount of the resistant *E. coli* EPSPS sufficient to replace the inhibited plant enzyme are resistant to the effects of glyphosate. Thus, in these cases, the crop plant would not be affected by glyphosate treatment, whereas the weeds would be. Crops

that have been engineered to be resistant to glyphosate by this approach are said to be "Roundup ready."

Rice is the major crop of ~50% of the world's population. Moreover, it is essential to significantly enhance rice production to meet the increased demand for food of an ever-increasing global population. As a consequence, there has been a shift from traditional transplanting (where rice is first grown in greenhouses and then transplanted to the field) to direct seeding of rice (which has currently reached around 25% of the total rice crop). Unfortunately, with direct seeding, both the yield and quality of rice are severely affected by weeds, with losses sometimes reaching 80% of the planted crop. Although glyphosate has been used with great success with several different crops, it is not applied directly in rice fields due to its toxicity to rice. To overcome this problem, a gene encoding a glyphosate-tolerant EPSPS was isolated from the bacterium *Isoptericola variabilis*. To ensure that this gene would be expressed at a high level in transgenic rice plants, the gene was altered to reflect the codon usage that is found in rice plants. Following introduction of the modified gene into rice plants by *A. tumefaciens*, the *epsps* gene was expressed under the control of a maize ubiquitin promoter (which is highly active in monocot plants such as rice) at a significant level and provided the plants with a high level of glyphosate tolerance (Fig. 11.29).

Notwithstanding the many years of successful use of glyphosate and Roundup-ready plants, two important factors are now changing people's thinking about this approach. First, the herbicide patent has now expired, and other companies are very actively pursuing the development of plants that are resistant to glyphosate using other approaches. Second, there is

Figure 11.29 Introduction of a codon-modified *Isoptericola variabilis epsps* gene into rice plants so that the plants are resistant to the otherwise inhibitory effects of the herbicide glyphosate.



a realization that worldwide agriculture has become too dependent upon a single herbicide and that alternative strategies need to be developed. In this regard, glyphosate has been applied in some fields since the mid-1990s, and in recent years at least 19 different weeds have acquired naturally occurring resistance to glyphosate.

To find an enzyme that can inactivate glyphosate, one group of researchers assayed a collection of several hundred *Bacillus* sp. strains for the ability to acetylate glyphosate (Fig. 11.30). The assay was based on the ability to measure *N*-acetylglyphosate in the supernatant of permeabilized cells (Fig. 11.31). The three strains (all *Bacillus licheniformis*) that had the highest level of glyphosate *N*-acetyltransferase activity were isolated, and the enzymes were characterized. In all cases, these strains exhibited only a very low level of enzyme activity. Subsequently, the genes from each of the three selected strains encoding glyphosate *N*-acetyltransferase activity were isolated and then shuffled numerous times (see chapter 3), each time selecting the strain with the highest level of glyphosate *N*-acetyltransferase activity. Following 11 iterations of gene fragment shuffling, the enzyme catalytic efficiency (the ratio of k_{cat}/K_m) improved by nearly 10,000-fold. In each iteration of DNA shuffling, about 5,000 gene variants were screened

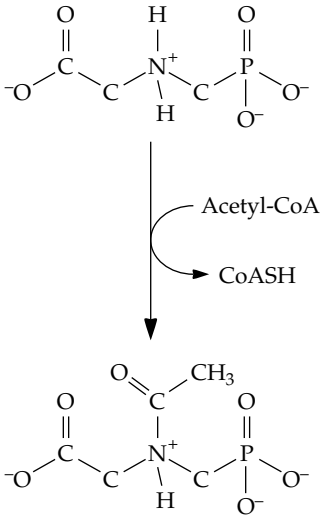
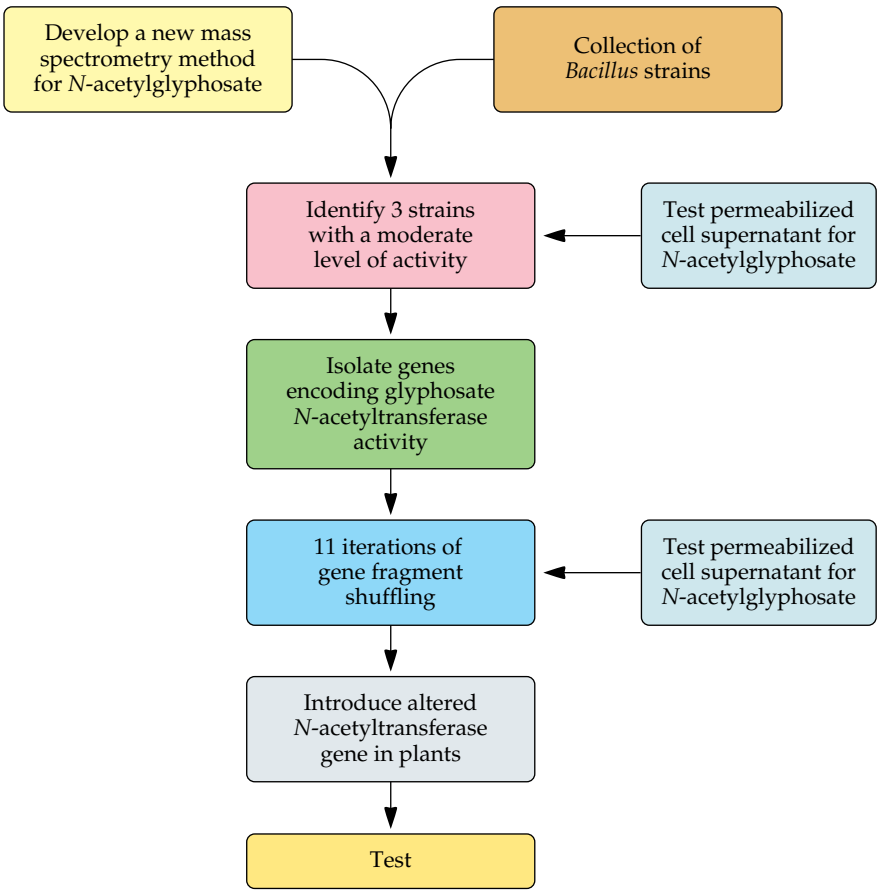


Figure 11.30 *N*-acetylation of the herbicide glyphosate by the bacterial enzyme glyphosate *N*-acetyltransferase. CoASH, coenzyme A.

Figure 11.31 Overview of a scheme to isolate a bacterial enzyme with a sufficient level of glyphosate *N*-acetyltransferase activity to allow it to be used to engineer plants so that they are resistant to high levels of glyphosate.



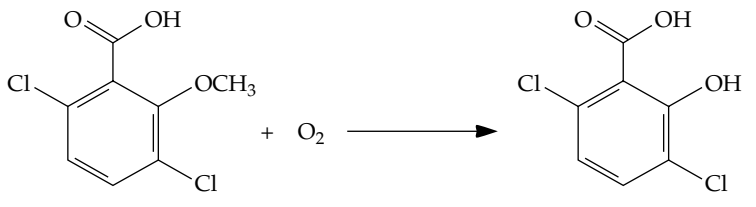
and tested. Interestingly, while the modified enzyme still functions as a glyphosate *N*-acetyltransferase, after so many rounds of modification, the amino acid sequence of the modified enzyme was only 76 to 79% identical to the amino acid sequences of the parental enzymes. Finally, the modified glyphosate *N*-acetyltransferase gene was introduced into *Arabidopsis*, tobacco, and corn plants. The transgenic plants, which expressed the enzyme in the plant cytosol and were both morphologically normal and fertile, were tolerant of approximately six times the dose of glyphosate that killed the parental nontransformed plants. The efficacy of this approach has now been demonstrated in the field, and it is expected that transgenic plants that express a modified glyphosate *N*-acetyltransferase gene will soon become commercially available. Moreover, transgenic tobacco plants that simultaneously overexpressed glyphosate *N*-acetyltransferase and EPSPS have been constructed. These doubly transformed plants, with two different mechanisms to resist the toxic effects of glyphosate, were found to be more resistant to glyphosate than those containing only one resistance gene.

Dicamba

The herbicide dicamba (Fig. 11.32A) has been used since the 1960s to control a wide range of broadleaf weeds. When it is applied to dicotyledonous plants, dicamba acts by mimicking the effects of high levels of the plant hormone indole-3-acetic acid and binding to indole-3-acetic acid receptors, which are essential for normal growth and development of the plant, and thereby inhibits receptor activity. The herbicide is widely used, relatively inexpensive, and environmentally friendly in that it does not persist in soils and has no toxicity to humans or other animals. Moreover, the widespread use of the herbicide has not led to the development of any

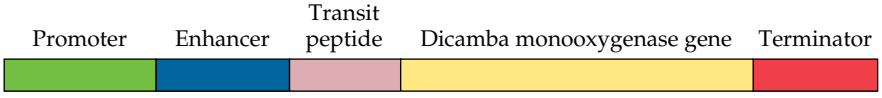
Figure 11.32 **(A)** Conversion of dicamba to 3,6-dichlorosalicylic acid by dicamba monooxygenase. **(B)** The genetic construct used to express the dicamba monooxygenase gene within the chloroplasts of transgenic plants. The promoter was from peanut chlorotic streak virus, the enhancer was from tomato etch virus, the transit peptide was from the small subunit of pea ribulose 1,6-bisphosphate carboxylase, the dioxygenase gene was from the soil bacterium *Pseudomonas maltophilia*, and the terminator sequence was from the small subunit of pea ribulose 1,6-bisphosphate carboxylase. Data from Behrens et al., *Science* **316:**1185–1188, 2007.

A



Dicamba 3,6-Dichlorosalicylic acid

B

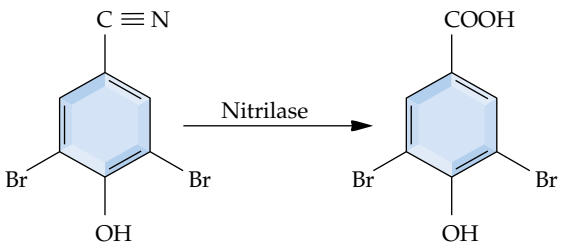


dicamba-resistant weeds. Researchers have therefore sought to develop crop plants that are resistant to dicamba. To do this, a bacterial dicamba monooxygenase gene was isolated and subsequently expressed in *Arabidopsis*, tomato, and tobacco plants (all as test systems) (Fig. 11.32B). The dicamba monooxygenase is part of the three-component enzyme dicamba *O*-demethylase, from the bacterium *Pseudomonas maltophilia*, which converts dicamba to 3,6-dichlorosalicylic acid, a compound without any appreciable herbicidal activity (Fig. 11.32A). In transgenic plants, only dicamba monooxygenase is needed for the inactivation of the herbicide, since the enzyme can be targeted to the chloroplast, where there is a ready source of reduced ferredoxin (the product of the other two enzymes encoded in the dicamba *O*-demethylase complex). The reduced ferredoxin supplies electrons for the monooxygenase reaction. As expected, transgenic plants expressing dicamba monooxygenase are resistant to high levels of the herbicide when grown in both the greenhouse and the field. It should be possible to "stack" plants with genes encoding both glyphosate and dicamba resistance so that farmers can either alternate the use of the two herbicides or else apply them at the same time. In this way, it is anticipated that unwanted weeds are unlikely to develop resistance to both herbicides, while the transgenic crop plant is uninhibited by the herbicide.

Other Herbicides

Resistance due to inactivation of bromoxynil (3,5-dibromo-4-hydroxy benzonitrile), an herbicide that acts by inhibiting photosynthesis, has been achieved for some plants. In this case, resistant plants were created by the introduction of a bacterial gene that encodes the enzyme nitrilase, which can inactivate bromoxynil before the herbicide can act (Fig. 11.33). The gene for the nitrilase was isolated from the soil bacterium *Klebsiella ozaenae* and placed under the control of the light-regulated promoter from the small subunit of the enzyme ribulose bisphosphate carboxylase before it was transferred to tobacco plants. As expected, the transgenic plants expressed nitrilase activity in their shoots and leaves, but not in their roots, and were resistant to the toxic effects of the herbicide.

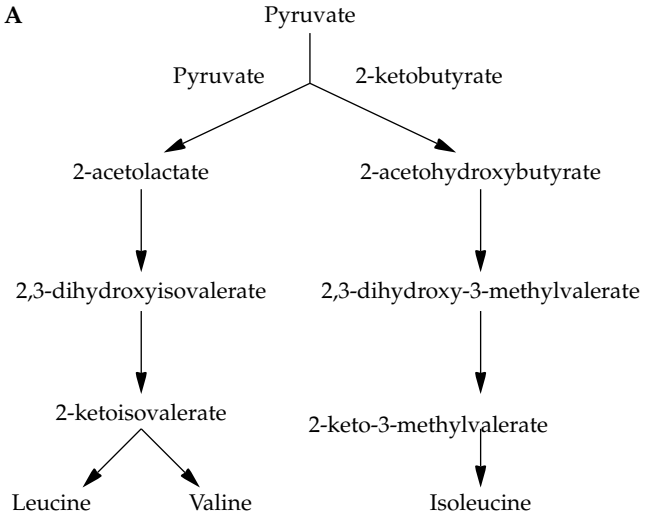
Maize (corn; *Zea mays*) was first domesticated by indigenous peoples in southern Mexico and Central America around 10,000 years ago and has since become a global food staple with total annual production (more than a billion metric tons) surpassing that of both rice and wheat. Given the enormous amount of maize that is lost because of weeds, scientists



Bromoxynil

3,5-Dibromo-4-hydroxybenzoic acid

Figure 11.33 Detoxification of the herbicide bromoxynil by the enzyme nitrilase from *Klebsiella ozaenae*.



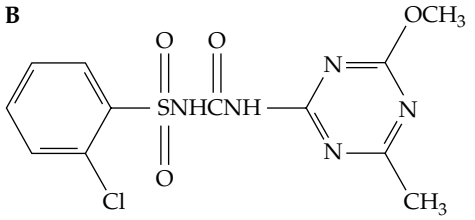
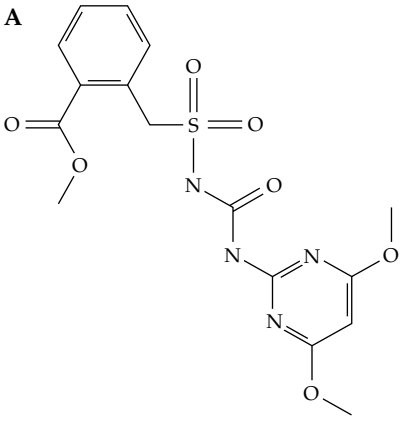
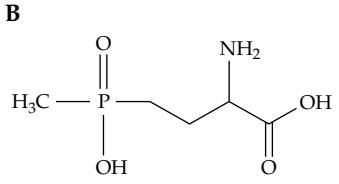


Figure 11.34 **(A)** The biosynthesis of branched-chain amino acids in plants, starting with pyruvate, where the enzyme acetolactate synthase catalyzes the first step in the pathway to leucine and valine (where two molecules of pyruvate are condensed) and the pathway to isoleucine (where a molecule of pyruvate is condensed with a molecule of 2-ketobutyrate). **(B)** The chemical structure of the sulfonylurea herbicide chlorosulfuron.

have devoted considerable efforts to creating transgenic maize varieties that express by the introduction of exogenous herbicide resistance genes, as described above. On the other hand, genome-edited herbicide-resistant plants may be advantageous compared to plants generated by the introduction of mainly foreign genes because, in many jurisdictions, genome-edited plants are not regulated. One group of scientists recently reported using CRISPR/Cas9 to introduce missense mutations into the two maize genes encoding the enzyme acetolactate synthase, which catalyzes the initial step in the biosynthesis of the branched-chain amino acids valine, leucine, and isoleucine (Fig. 11.34A). These mutations produced functional enzymes that were no longer inhibited by high levels of the common sulfonylurea herbicide chlorosulfuron (Fig. 11.34B). This mutation also conferred chlorosulfuron resistance to both *Arabidopsis* and soybean. Thus, this may be a generally effective strategy.

Figure 11.35 The chemical structures of the herbicides **(A)** bensulfuron methyl and **(B)** glufosinate.





The continuous use of any one herbicide generally leads to the evolution of weeds resistant to that herbicide. To get around this problem, farmers often alternate the use of two or more herbicides, a strategy that significantly slows the development of resistant weeds. However, this strategy requires crops that are resistant to multiple herbicides. One group of researchers engineered rice plants using *A. tumefaciens* to introduce foreign genes that conferred resistance to both bensulfuron methyl and glufosinate (Fig. 11.35). The transgenic plants demonstrated significant resistance to both of these herbicides during both seed germination and when the plant was more fully developed. In addition, the transgenic plants did not display any reduced yield as a consequence of expressing resistance to these two herbicides. Whether the strategy of alternating herbicide application in the field is an effective means of avoiding resistant weed development over several growing seasons remains to be seen.

Fungus and Bacterium Resistance

Extensive plant damage with the loss of crop productivity is often caused by phytopathogenic fungi. For example, it has been estimated that one fungal disease of one major crop, i.e., fungal rice blast, a disease that affects

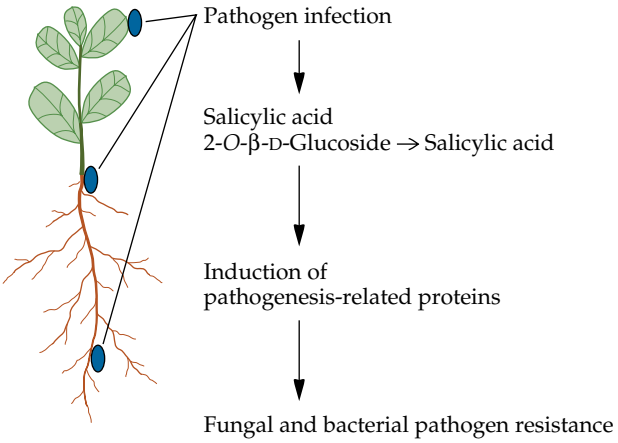
rice plants, costs farmers in Southeast Asia, Japan, and the Philippines more than $5 billion per year. At present, the major way of controlling the damage and losses to crop plants that result from fungal infection is through the use of chemical agents that may persist and accumulate in the environment and that are subsequently hazardous to animals or humans; and in many countries the use of certain chemical fungicides has been banned. It would therefore be beneficial if a simple, inexpensive, effective, and environmentally friendly nonchemical means of preventing fungal damage to crop plants could be found.

Plants often respond to fungal or bacterial pathogen invasion or other environmental stresses by converting a conjugated storage form of salicylic acid (salicylic acid 2-*O*-β-D-glycoside) to salicylic acid, which induces a broad systemic defense response in the plant. This "systemic acquired resistance" to pathogens extends to plant tissues that are far from the site of the initial infection and may last for weeks to months. It results from the synthesis of a group of proteins called pathogenesis-related (PR) proteins (Fig. 11.36). The PR proteins include β-1,3-glucanases, chitinases, and protease inhibitors that protect the plant from invading pathogens. To develop plants resistant to fungal pathogens, researchers have attempted to utilize parts of the systemic acquired resistance system. For example, transgenic plants that constitutively express high levels of one or more PR proteins, such as chitinase, which can hydrolyze the β-1,4 linkages of the *N*-acetyl-D-glucosamine polymer chitin, a major component of many fungal cell walls, have been engineered.

Transgenic Plants

The *NPR1* gene encodes a "master" regulatory protein that positively controls the expression of the PR proteins, and it can be activated or induced by the addition of salicylic acid. Overexpression of the *NPR1* gene can lead

Figure 11.36 Overview of systemic acquired resistance in plants. After infection of a plant by a pathogenic fungus or bacterium (shown in blue), the inactive storage compound salicylic acid 2-*O*-β-D-glucoside is converted to salicylic acid and/or salicylic acid is synthesized. The salicylic acid activates or induces the *NPR1* gene, whose protein product acts as a "master" regulatory protein to turn on the expression of the PR proteins, which have enzyme activities directed against various pathogenic organisms.

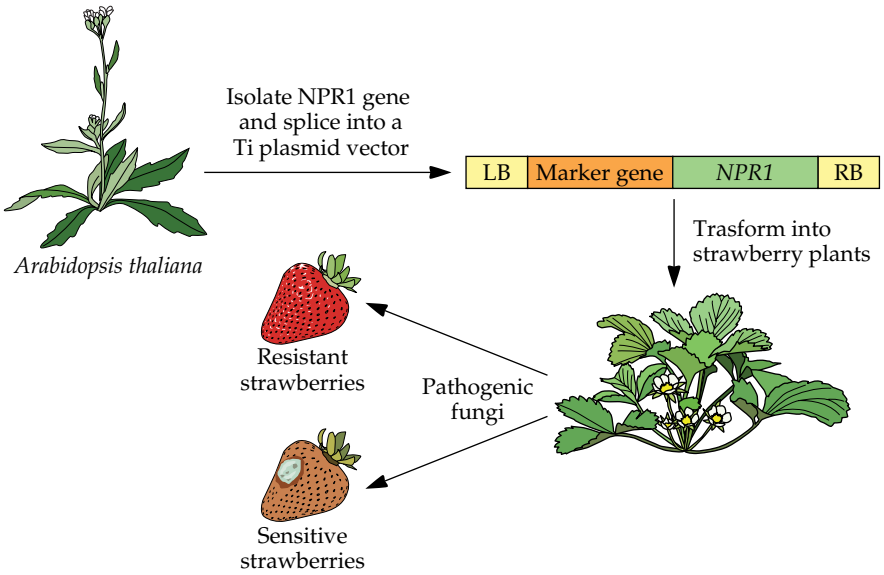


to the generation, in a range of different plants, of broad-spectrum disease resistance against both fungal and bacterial pathogens. For example, transformation of strawberry plants with an *NPR1* gene from *A. thaliana* yields transgenic plants that are protected against several different fungal pathogens that would otherwise make the strawberries unsuitable for human consumption (Fig. 11.37).

Another approach to engineering plants with broad-spectrum disease resistance involves overproducing salicylic acid. This can be done by transforming plants with bacterial genes that encode the enzymes isochorismate synthase and isochorismate pyruvate lyase, which catalyze salicylate synthesis (Fig. 11.38A). Salicylate is synthesized from chorismate, which is produced in large amounts in the chloroplast and is also an intermediate in the biosynthesis of the amino acid tryptophan. The two bacterial genes for salicylate synthesis were fused to chloroplast-targeting sequences from the gene for the small subunit of ribulose bisphosphate carboxylase; the small subunit of ribulose bisphosphate carboxylase is encoded within the nuclear DNA, but following its synthesis, this protein is transmitted to the chloroplast (Fig. 11.38B). The result of this genetic manipulation was that both enzymes were localized in the (tobacco) plant chloroplast, and salicylic acid was produced constitutively. Consequently, the plants expressed a number of PR proteins. The plants appeared normal but exhibited enhanced resistance to both viral and fungal pathogens. However, it is not necessarily advantageous to the plant to constitutively express PR proteins, so this strategy is unlikely to confer protection against a broad range of pathogens in the field.

Transgenic plants (including rice, tobacco, and canola) that expressed chitinase under the control of the *Cauliflower mosaic virus* 35S promoter were more resistant to damage by fungal pathogens than control plants, even though the control plants synthesized their own PR proteins in response to the fungal infection. This resistance reflects the higher level

Figure 11.37 Creation and testing of transgenic strawberries that overexpress an *NPR1* gene from *A. thaliana*. Following testing, only transformants that yield strawberry plants that are fungus resistant are retained.



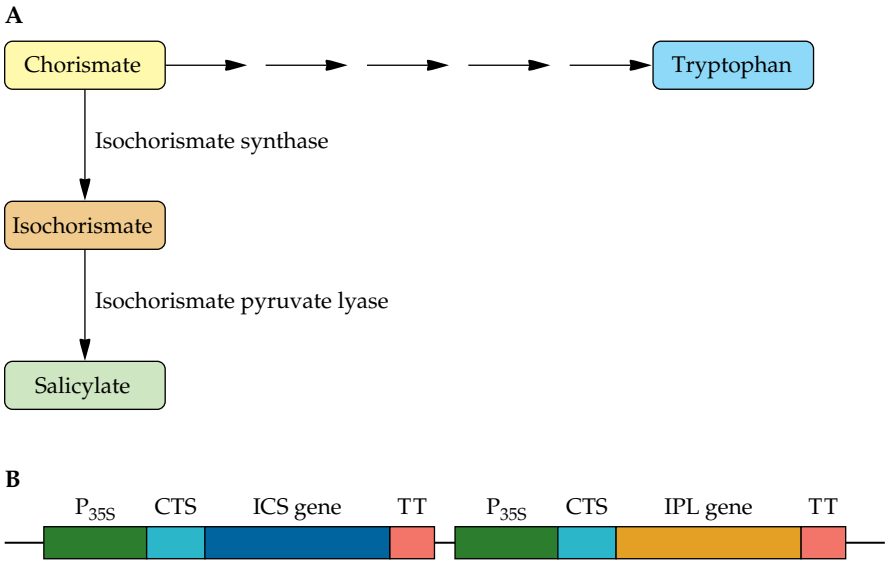


Figure 11.38 **(A)** Use of bacterial enzymes to convert plant chloroplast chorismate to salicylate. **(B)** Construct used to transform plants so that they constitutively overproduce salicylate. P_{35S}, the 35S promoter from *Cauliflower mosaic virus*; CTS, chloroplast targeting sequence; ICS, isochorismate synthase; TT, transcription termination region; IPL, isochorismate pyruvate lyase.

of chitinase expressed by the transgenic plants than by nontransformed plants. In addition, while transgenic plants constitutively expressing chitinase were resistant to fungal pathogens, binding of the beneficial root fungus *Glomus mosseae* to the plant roots was not affected. This was probably a consequence of a difference in the cell wall compositions of different fungi. Importantly, a transgenic plant constitutively expressing chitinase has been found to be effective at resisting fungal damage under field as well as lab conditions. In a separate study, moderate levels of resistance to the phytopathogenic fungus *Rhizoctonia solani* were obtained when rice plants were transformed with a chitinase gene from bitter melon that included the melon promoter. Overall, the overexpression of some PR proteins, such as chitinase, appears to be an effective strategy for protecting plants against damage from several pathogenic fungi.

Fusarium Diseases

Pathogenic fungi belonging to the genus *Fusarium* are the causative agents of some of the most costly and devastating plant diseases in the world; *Fusarium* sp. infections result in losses of around 28 million metric tons of wheat grain, which is worth ~$5.6 billion. Therefore, a strategy that targets various *Fusarium* spp. would be quite important for agriculture worldwide. Several different antimicrobial peptides (which can disrupt the cell membrane) and the enzyme chitinase (mentioned above) are inhibitory to the growth of *Fusarium* spp. However, these biological approaches (regardless of how they are administered) are not as effective as spraying plants with chemical fungicides. To overcome this limitation, workers have fused the coding sequences (cDNAs) for two different antimicrobial peptides (one from the radish *Raphanus sativus* and one from the mold *Aspergillus giganteus*) and a chitinase (from wheat) separately

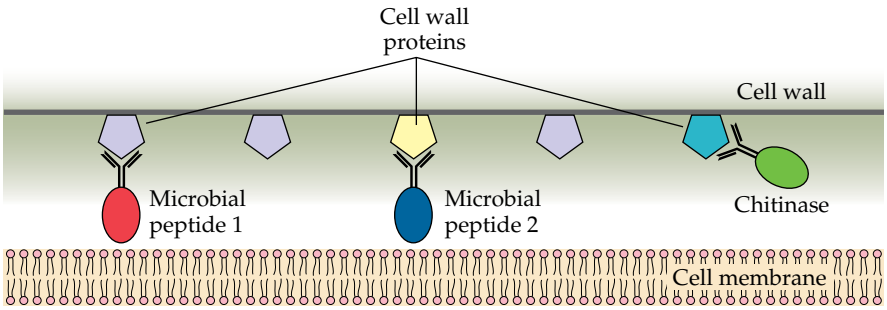


Figure 11.39 Three different antifungal fusion proteins. Each protein includes a single-chain Fv antibody that binds to a *Fusarium* sp. cell wall protein and an antifungal peptide/protein directed against either the cell membrane or the chitin component of the cell wall. Data from Bohlmann, *Nat. Biotechnol*. **22:**682–683, 2004.

to a single-chain Fv antibody that binds to a *Fusarium* cell wall protein (Fig. 11.39). Although the single-chain Fv antibody was originally selected from a library constructed from chickens that had been immunized with *Fusarium graminearum*, the selected antibody cross-reacted with cell wall antigens from nine different species and subspecies of *Fusarium*. Constructs encoding the three fusion proteins were used to transform *A. thaliana*, and the transgenic plants were tested for resistance to infection and growth inhibition by *Fusarium oxysporum*. When either the selected single-chain Fv antibody or any one of the three antipathogenic peptides/proteins was expressed in transgenic plants, it endowed the plants with a low to moderate level of resistance to damage by the pathogen (Table 11.5). However, the three antibody–peptide/protein fusions conferred a high level of resistance to the pathogen, suggesting, in each case, that the two components of the fusion protein were acting synergistically. This is an interesting and potentially quite useful approach that merits further development.

Despite significant worldwide efforts to find resistance genes against *F. graminearum*, no completely resistant plant variety is currently available. *F. graminearum* reduces plant yields by synthesizing the toxin deoxynivalenol (also called vomitoxin), which renders the grain too toxic to use. However, one group of scientists recently discovered a variety of wheat that was resistant to some of the effects of this pathogen. This resistance was based on the wheat containing a *Fusarium head blight 7* (*Fhb7*) gene that encodes a glutathione *S*-transferase that detoxifies deoxynivalenol (Fig. 11.40A). Based on gene homology, the resistance gene is thought to have been acquired by the weed tall wheat grass (*Thinopyrum ponticum*) from the endophytic fungus *Epichloë* sp. many thousands of years ago. The *Fhb7* gene was subsequently transferred to one strain of wheat. The glutathione *S*-transferase detoxifies deoxynivalenol by conjugating it to glutathione (Fig. 11.40B), and since deoxynivalenol is required for *Fusarium* growth on infected tissue, growth of the pathogen ceases. It should now be straightforward to deliberately transfer the *Fhb7* gene to other strains of wheat and to other plants and thereby confer tolerance to *F. graminearum*.

Table 11.5 Resistance of transgenic *A. thaliana* to the phytopathogen *F. oxysporum*[a]

Transgene	Disease index (%)
None (wild type)	100
Fv antibody	55
Peptide 1	50
Peptide 2	60
Chitinase	52
Antibody–peptide 1 fusion	10
Antibody–peptide 2 fusion	0
Antibody-chitinase fusion	5

[a]Data from Peschen et al., *Nat. Biotechnol*. **22:** 732–738, 2004. The extent of disease was assessed 2 weeks after infection with *F. oxysporum*. A disease index of 100% indicates that all of the plants are dead, 50% indicates that the average plant has disease symptoms but is alive, and 0% indicates that all plants are disease free.

Chestnut Blight

Chestnut blight is a fungal disease (caused by *Cryphonectria parasitica*) that has had a devastating effect on American chestnut trees in the United States beginning in the early part of the 20th century. By around 1940, an

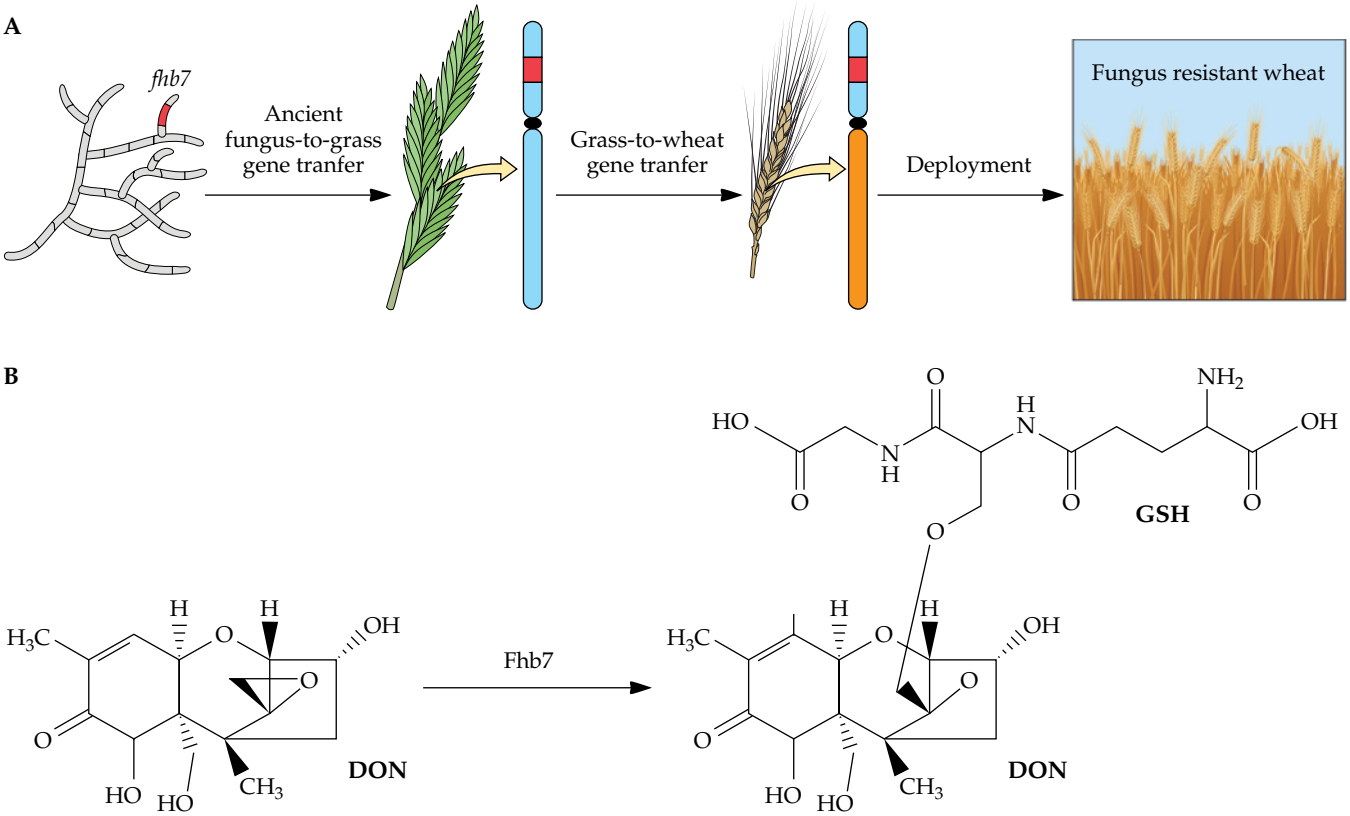


Figure 11.40 (A) The transfer of the *Fnb7* gene from an ancient fungus to a wild grass to a strain of wheat. The putative position of the gene on the grass and wheat chromosomes is shown. **(B)** The Fnb7 enzyme neutralizes deoxynivalenol (DON) by conjugating it with a molecule of glutathione (GSH).

estimated 4 billion chestnut trees had been killed by this fungus, removing a mainstay of small-town life in the eastern United States that had been immortalized by the American poet Henry Wadsworth Longfellow in his poem "The Village Blacksmith." For around 200 years, the American chestnut tree was economically important, providing both a food source and a soft and decay-resistant wood that had a variety of uses, including the making of houses and furniture. Mature trees could grow up to 100 feet high with a trunk diameter of around 14 feet. Since the middle of the 20th century, the main mechanism of creating chestnut trees resistant to this fungus has involved the generation of hybrid trees combining American chestnut trees with some Asian chestnut species that are resistant to the fungus. More recently, scientists have reported creating transgenic chestnut trees that express a wheat oxalate oxidase gene (*oxo*) that can protect the trees from the oxalic acid secreted by the disease-causing fungus (Fig. 11.41). In the absence of the fungus, seeds from the transgenic trees germinated to the same extent and produced a similar level of biomass as seeds from non-transgenic trees. Oxalic acid is toxic to chestnut trees because it lowers the pH of infected plant tissues from around the normal pH 5.5 to around 2.8, which is generally toxic to plant cells. Following these recent experiments, all indications are that this approach appears to be quite effective. Scientists

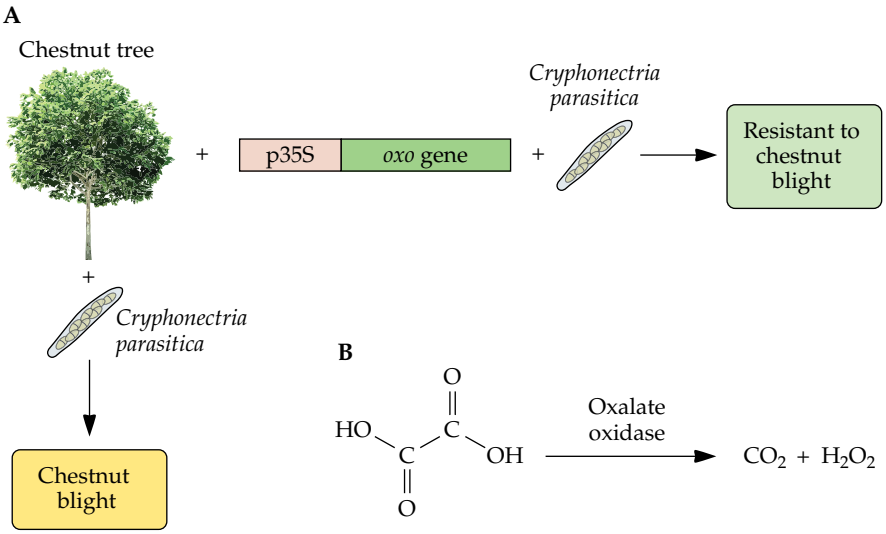


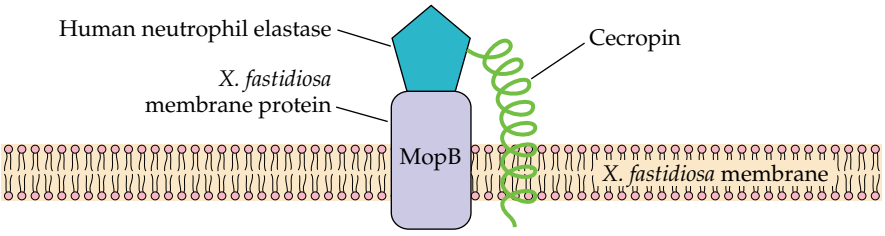
Figure 11.41 Chestnut trees that have been genetically transformed to express the wheat *oxo* gene (encoding the enzyme oxalate oxidase) are resistant to the fungal pathogen that causes chestnut blight **(A)**. The enzyme oxalate oxidase converts oxalate into carbon dioxide plus hydrogen peroxide **(B)**.

are now optimistic that this approach may help them to eventually repopulate the eastern United States with American chestnut trees.

Wine Disease

Pierce disease, which is caused by the Gram-negative bacterium *Xylella fastidiosa*, is a significant economic problem for the wine industry in California. To protect grape plants against this pathogen, scientists have constructed, at the DNA level, a gene that encodes a chimeric protein that targets and destroys the pathogenic bacterium (Fig. 11.42). The fusion protein consists of the human neutrophil elastase, which binds tightly to the MopB protein, the most abundant protein in the *X. fastidiosa* outer membrane, and cecropin, a lytic peptide that targets conserved lipid molecules and creates holes in the bacterial outer membrane. Each portion of the fusion protein by itself provides some protection to grape plants against the bacterial pathogen. That is, the human neutrophil elastase binds to

Figure 11.42 Engineering a targeting and lysis defense system against the plant bacterial pathogen *Xylella fastidiosa*. A fusion protein consisting of the protease human neutrophil elastase, which binds tightly to the MopB protein, the most abundant protein in the *X. fastidiosa* outer membrane, and cecropin, a lytic peptide that targets conserved lipid molecules and creates holes in the outer membrane, was constructed and used to create transgenic grape plants expressing this protein.



and then partially digests the MopB protein on the bacterial outer membrane. Likewise, the cecropin binds to and partially digests the bacterial outer membrane lipid molecules. However, when the two molecules are combined into a single fusion protein, they act synergistically and decrease the damage from the pathogen more than 100 times the amount that either molecule by itself protects the plant. The fusion protein was designed using a plant-specific signal sequence so that the protein was delivered first to the leaf apoplast and then to the xylem (the vascular plant tissue that delivers water and dissolved minerals from the roots to the leaves and shoots). This delivery system is quite effective, as this bacterium enters the plant through the xylem. While this result is still preliminary, it promises to be a highly effective means of dealing with this bacterial pathogen.

RNAi and CRISPR/Cas

A better understanding of some of the fungal and bacterial phytopathogens that infect plants and limit their growth and development, including not only their biology but also their genomic sequences, has allowed scientists to develop the techniques of RNAi and CRISPR/Cas to target specific pathogens (mostly bacteria and fungi but also some viruses). These two systems/techniques, initially viewed as bacterial defenses against viruses, have been developed by researchers to a level where they are commonly employed both in medicine (see chapter 6) and agriculture. Using these approaches is generally technically simpler than developing transgenic plants that target various aspects of pathogen functioning. Moreover, the use of genome-editing techniques such as CRISPR/Cas, which do not introduce any foreign DNA into the plant, obviate many of the concerns that still exist regarding the use of transgenic plants. Some examples of the use of these approaches are summarized in Table 11.6.

Table 11.6 Recent (2015 to 2020) use of RNAi, dsRNA, and CRISPR/Cas techniques that target specific fungal and bacterial phytopathogens

Target organism	Disease	RNAi, dsRNA, or CRISPR/Cas	Gene target	Plant
Blumeria graminis	Powdery mildew	RNAi	Glucanosyl transferase	Barley
Uromyces appendiculatus	Common bean rust	RNAi	4 different targets	Common bean
Sclerotinia sclerotiorum	Stem rot	RNAi	Respiratory burst oxidase homologs	Soybean
Fusarium graminearum	Head blight	dsRNA	Fungal ergosterol biosynthesis	Barley
Agrobacterium tumefaciens	Crown gall	RNAi	Auxin and cytokinin biosynthesis	Tomato, *A. thaliana*, walnut
Xanthomonas citri subsp. *citri*	Citrus canker	CRISPR/Cas	*CsLOB1* promoter	Oranges
Magnaporthe oryzae	Rice blast	CRISPR/Cas	ERF transcription factor	Rice
Oidium neolycopersici	Powdery mildew	CRISPR/Cas	Transmembrane protein	Tomato
Xanthomonas, *Pseudomonas*, *Phytophthora capsici*	Blight and fruit rot	CRISPR/Cas	2-Oxoglutarate	Tomato
Blumeria graminis	Powdery mildew	CRISPR/Cas	Raf-like mitogen-activated kinase kinase kinase	Wheat
Xanthomonas citri subsp. *citri*	Citrus canker	CRISPR/Cas	LOB domain protein	Oranges
Xanthomonas oryzae pv. *oryzae*	Bacterial blight	CRISPR/Cas	SWEET gene promoters	Rice
Verticillium dahliae	Verticillium wilt	RNAi	VdRGS1 gene	Cotton
Oidium neolycopersici	Powdery mildew	CRISPR/Cas	Membrane-associated protein	Tomato

Salt and Drought Stress

Many plants live in environments where growth is severely impaired by either drought or high salinity. With increasing dependence on irrigation in agriculture and (in northern climes) more frequent salting of icy and snowy roads in the winter, increased soil salinity has become a common problem worldwide. Irrigation typically increases the amount of salt present in soil, and approximately one-third of the world's irrigated land has become unsuitable for growing crops because of contamination with high levels of salt. To survive under these conditions, many plants synthesize low-molecular-weight nontoxic compounds collectively called osmoprotectants. These compounds facilitate both water uptake and retention and also protect and stabilize cellular macromolecules from damage by high salt levels. Some well-known osmoprotectants are sugars, alcohols, the amino acid proline, and quaternary ammonium compounds.

Many of the strategies that have been used to engineer plants to become more salt tolerant are also effective at making the plant drought tolerant (Table 11.7); however, some strategies are specific for one stress or the other. In fact, a very large number of different genes have been employed in attempts to create drought-tolerant transgenic plants. These approaches have included introducing genes encoding enzymes for overproduction of various osmolytes (e.g., trehalose, proline, glycine betaine, and polyamines), plant stress proteins (e.g., chaperones and heat shock proteins), reactive-oxygen-scavenging proteins (e.g., superoxide dismutases), hormone biosynthesis and catabolism proteins (e.g., affecting cellular levels of abscisic acid, cytokinin, and ethylene), transcription factors that turn on the synthesis of a host of other proteins, and signaling proteins that activate the synthesis of other proteins.

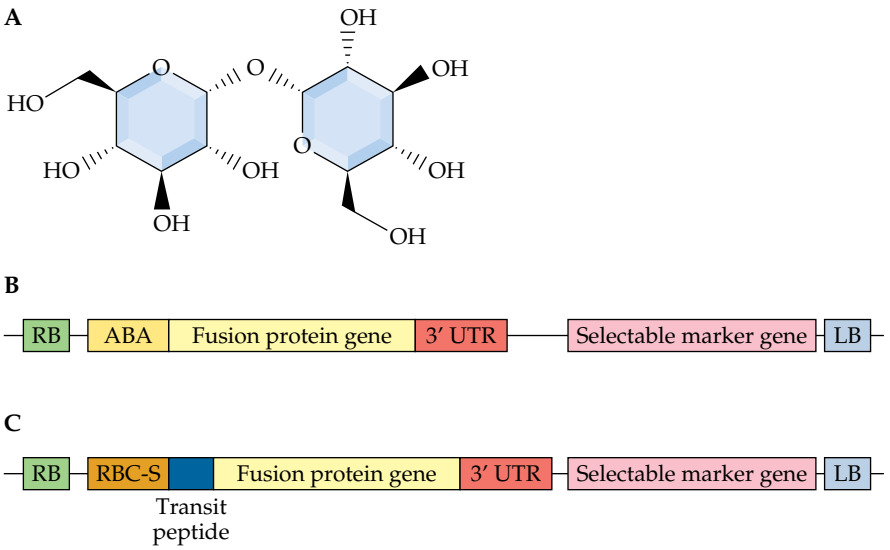
Table 11.7 Genetic engineering of plants to increase their salt and drought resistance

Plant	Introduced gene(s)
Rice	Trehalose biosynthesis
Corn, cotton, rice, tobacco, tomato, canola	Na^+/H^+ antiport protein
Cotton, rice	Glycine betaine biosynthesis
Barley	CIPK16, a protein kinase
Cotton, barley	H^+-pyrophosphatase gene
Rice	Inositol biosynthesis
Canola	ACC deaminase
Date palm	Aquaporins, serine/threonine protein kinases, peroxidases
Tobacco	Mannitol-1-phosphate dehydrogenase
Tobacco	Delta 1-pyrroline-5-carboxylate synthase
Tobacco	Ascorbate peroxidase
Rice	Transcription factor *DREB1A/CBF3*
Tomato	Transcription factor *SlAREB1*
Rice	Transcription factor *ZFP179*
Rice	Calcium-dependent protein kinase
Tobacco	MAP kinase *GhMPK2*

Increasing Trehalose Production

Trehalose is a naturally occurring α,α-1,1-glucoside consisting of two molecules of α-glucose (Fig. 11.43A). It is a highly stable molecule that is resistant to both acid and high temperature and is thought to form a gel phase (replacing water) as cells dehydrate, thereby decreasing the damage to cells from drought and salt. It is possible, by genetic manipulation, to increase the concentration of trehalose in plants and thereby protect the plants against inhibition by drought or high levels of salt in the soil. In *E. coli*, trehalose-6-phosphate is first formed from uridine diphosphate (UDP)-glucose and glucose-6-phosphate, and then the trehalose-6-phosphate is converted to trehalose. To create transgenic plants, a fusion of the genes encoding the two enzymes that normally catalyze the two steps in the biosynthesis of trehalose in *E. coli* was constructed so that a single protein contained both activities. This simplifies the transformation of plants in that only one target gene needs to be introduced and ensures that the two enzyme activities necessary for the synthesis of trehalose are present at identical levels. Rice plants were transformed, using a binary vector, with the fusion protein gene under the control of one of two different promoters (Fig. 11.43B and C). In one genetic construct, the fusion protein gene is under the transcriptional control of an abscisic acid-inducible promoter (where abscisic acid is a plant hormone that plays a key role in the response of plants to drought and salt stress) and is expressed in the cytosol. In the other construct, it is under the control of the promoter for the small subunit of ribulose bisphosphate carboxylase (a very highly expressed plant protein), and the fusion protein is expressed in plant chloroplasts. In

Figure 11.43 Structure of trehalose. **(A)** Two genetic constructs used to engineer rice plants to synthesize trehalose to protect them against growth inhibition by high salt levels. The fusion protein is under the control of an abscisic acid (ABA)-inducible promoter where ABA synthesis increases in the presence of salt **(B)** or the promoter from the small subunit of ribulose bisphosphate carboxylase **(C)**. In panel C, the transit peptide facilitates the localization of the fusion protein in the chloroplasts. RB, right border; LB, left border; UTR, untranslated region.

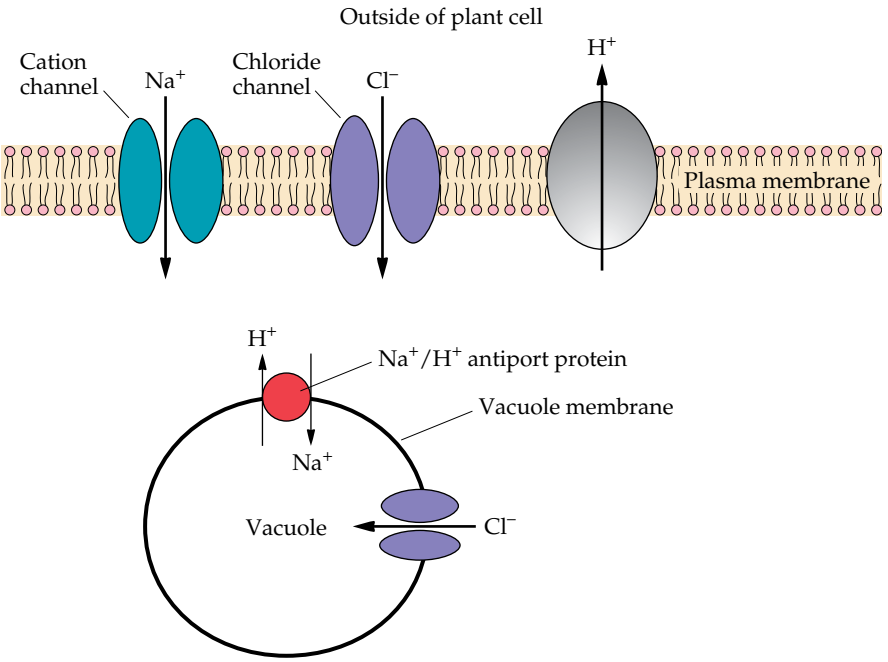


transgenic rice plants that contain either DNA construct, the level of trehalose is 3 to 10 times higher than that in nontransformed rice plants in the presence of salt. Moreover, the biomass of the transgenic plants is four to six times that of the nontransformed plants in the presence of salt. Thus, by increasing the amount of trehalose that a plant synthesizes, the plant acquires increased tolerance for moderate levels of salt in the environment.

Sequestering Sodium Ions

Researchers have engineered the plant *A. thaliana* to be salt tolerant by sequestering sodium ions in the large intracellular vacuole (Fig. 11.44). The strategy consisted of overproducing the endogenous *A. thaliana* gene encoding a Na^+/H^+ antiport protein. The Na^+/H^+ antiport protein transports Na^+ into the vacuole using the electrochemical gradient of protons generated by vacuolar H^+-translocating enzymes. When tested, the transgenic plants that overproduced the Na^+/H^+ antiport protein thrived in soil that was watered with a solution of 200 mM salt. This approach to the manipulation of salt stress in plants is effective with corn, canola, cotton, rice, tobacco, and tomato plants, as well as with *A. thaliana*. In transgenic tomato plants, the salt is localized in the leaves, and therefore, the transgenic tomato fruits do not accumulate salt and are quite normal in all respects, including taste. In addition to Na^+ toxicity, plants that live in saline environments have to contend with water loss caused by osmotic stress. By concentrating the salt in the plant's large vacuole, water that is free of salt is driven into the plant cells, resulting in plants that use water more efficiently. This system has been quite successful in greenhouse trials and in the limited number of field trials where it has been tested. It provides researchers with

Figure 11.44 Ion transport in the plant *A. thaliana* showing the Na^+ ions being sequestered in the large vacuole.

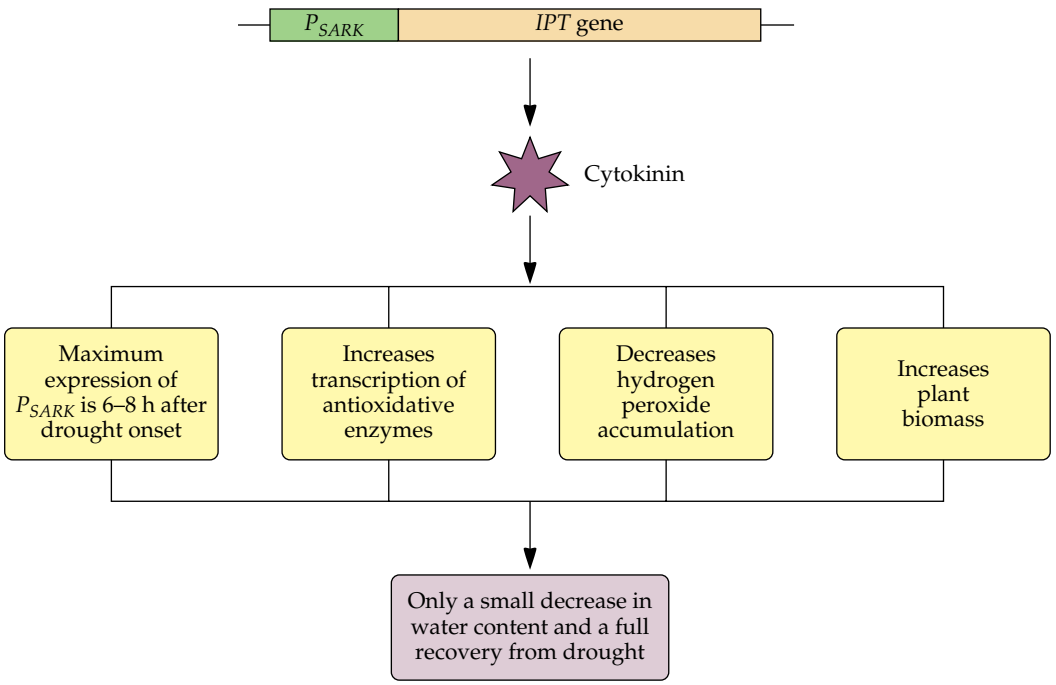


the potential, especially in combination with other approaches, such as the overproduction of certain osmolytes, to engineer a wide range of salt-tolerant crop plants that can be grown on marginal land or possibly watered with seawater or other salt-containing water.

Delaying Drought-Induced Senescence

One group of researchers reasoned that to increase the tolerance of plants for drought, it is necessary to delay the onset of drought-induced senescence during the drought episode. Moreover, by suppressing this response, plants would be more likely to resume normal growth when water became available. Of course, it is necessary to keep in mind that regardless of their genetic makeup, plants cannot exist for indefinite periods in the absence of water. Prior to this work, it had been observed that leaf senescence could be delayed in transgenic plants expressing a foreign gene encoding isopentenyltransferase, an enzyme that catalyzes the rate-limiting step in cytokinin biosynthesis. Therefore, tobacco plants were transformed to express isopentenyltransferase under the control of a *SARK* (senescence-associated protein kinase gene) promoter (Fig. 11.45). This regulatable promoter is induced during late maturation, and its activity was decreased during the development of senescence in the plant. When these transgenic tobacco plants were watered with only 30% of the amount of water used under normal conditions, the suppression of leaf senescence resulted in a 4- to 5-fold-higher level of biomass in the transgenic versus the nontransformed plants. This result suggests that it may be possible for irrigated crops to grow normally with only one-third the amount of water that is usually

Figure 11.45 Overview of how overexpressing cytokinin protects plants against drought. P_{SARK} is a senescence-induced promoter. The *IPT* gene encodes isopentenyltransferase, which catalyzes the rate-limiting step in cytokinin biosynthesis.



used. In fact, more recently, researchers have shown that this strategy is effective with peanut and rice as well as with tobacco and in both laboratory and field conditions.

Hundreds of different genes have been associated with the abiotic stress response in plants (including the response to salt and drought stress). The abiotic stress response is a multilevel physiological and biochemical reorganization that often involves massive changes depending on the severity and duration of one or more stresses. The gene network involved in the signaling, responses, and governing of diverse physiological functions includes various transcription factors that participate in regulating defense and modulate adaptive responses. Since changes in plant hormone levels typically occur prior to the transcriptional remodeling that is initiated to protect the plant against salt and drought stress, many of the strategies to protect plants involve efforts to modulate phytohormone levels, changes that are likely to have a global impact on the stressed plant.

Phytoremediation

Scientists and engineers have developed several methods to remove toxic compounds from polluted environments. Unfortunately, most of these methods include the physical removal of soil to landfill sites or extraction of contaminated soil or water through chemical or physical means, procedures which are prohibitively expensive. Moreover, many contaminated environments contain mixtures of contaminants that are not readily dealt with in a simple manner. These environmental contaminants include one or more metals, radioactive materials, and various organic compounds. Some major metal environmental contaminants include nickel, lead, zinc, cadmium, mercury, and arsenic. Major organic environmental contaminants include polychlorinated biphenyls (PCBs); polycyclic aromatic hydrocarbons (PAHs); trichloroethylene; petroleum products including benzene, toluene, ethylbenzene, and xylene; dichlorodiphenyltricloroethane (DDT); herbicides; explosives; and pesticides.

Some organic compounds can be broken down by bacteria that may either be naturally present in or added to the soil; in the absence of plants, however, this remediation process is usually quite slow and inefficient. On the other hand, the use of plants to facilitate the remediation of polluted soils, i.e., phytoremediation, is a potentially clean, effective, and relatively inexpensive technology that is currently being developed by a large number of researchers worldwide. Notwithstanding the progress that has been made in developing effective phytoremediation protocols, it is important to keep in mind that while organic pollutants can be broken down, and often mineralized, metal environmental contaminants that are taken up by plants are an entirely different problem. Typically, even with plants that can tolerate relatively high concentrations of metals, the metals are mostly localized in the plant roots and are not translocated to plant shoots, meaning that mature plants along with their roots must be harvested and removed.

Phytoremediation can occur in several different ways (Fig. 11.46):

1. In **phytodegradation**, many plants produce, and often secrete to the environment, enzymes that can degrade and mineralize a wide range of organic compounds. However, inorganic pollutants (primarily metals) cannot be degraded.

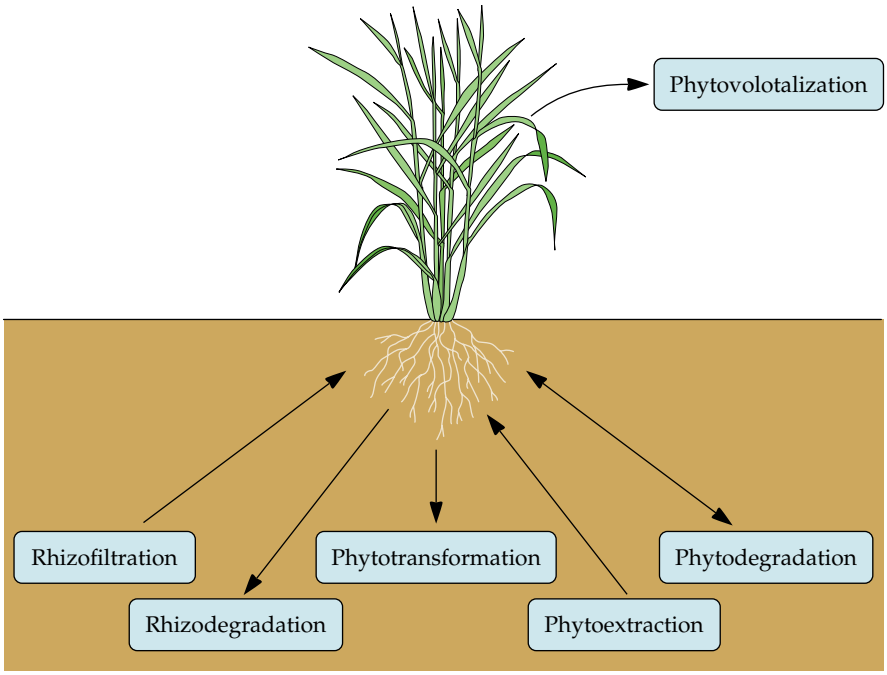


Figure 11.46 Schematic overview of the major processes of phytoremediation. The directions of the arrows indicate where the phytoremediation process most likely takes place.

2. In **rhizodegradation**, the bacteria that live around the roots of plants (i.e., the rhizosphere) produce enzymes that facilitate the degradation of organic pollutants.
3. In **phytoextraction**, environmental contaminants are extracted from soil and transported to and accumulated in plant tissues.
4. In **rhizofiltration**, water-soluble environmental contaminants, often from polluted lakes and rivers, are transported to and accumulated in plant tissues.
5. In the process of **phytotransformation**, environmental contaminants are transformed by plant and/or rhizosphere microorganisms to less harmful substances. This may include immobilizing some contaminants within the root zone and thereby preventing their spreading.
6. In **phytovolatization**, environmental contaminants are taken up by the roots and then transported to the leaves, where they are volatilized through the leaf stroma.

Most phytoremediation experiments have been performed with either wild-type plants or the combination of wild-type plants together with rhizosphere bacteria. However, some of these experiments have included the use of transgenic plants. For the removal of organic pollutants, plants are often engineered to reduce the stress in the plant caused by the presence of these compounds as well as to facilitate their degradation. This might include engineering transgenic plants with ACC deaminase, g-glutamylcysteine synthetase, cytochrome P450 monooxygenase, glutathione *S*-transferase, NADPH-dependent nitroreductase, or atrazine chlorohydrolase. The transgenic plants engineered to be more effective in the uptake of metals from

the environment may express one or more of the following genes: ACC deaminase, phytochelatins, metallothioneins, and metal chelators.

Fruits and Flowers

Flavr Savr Tomato

A major problem in fruit marketing is premature ripening and softening during transport. These changes are part of the natural aging (senescence) process of the fruit. Some of the genes that are induced during ripening encode the enzymes cellulase and polygalacturonase. It has been postulated that by interfering with the expression of one or more of these genes, the ripening process might be delayed. This interference might be achieved by creating transgenic plants with antisense or sense RNA-producing versions of these genes. In fact, when an antisense RNA-producing gene for polygalacturonase was introduced into tomato plants, a ~$1.7 billion-a-year crop in the United States in 2019, both polygalacturonase mRNA and enzymatic activity were reduced by 90%. The lowering of polygalacturonase production inhibited the rate of fruit ripening in tomatoes, permitting the tomatoes to ripen slowly on the vine instead of being harvested while they were still green. These tomatoes were claimed to have a long shelf life while retaining the flavor of the tomato. This genetically engineered tomato is known as the Flavr Savr (pronounced "flavor saver") tomato. In a precedent-setting decision regarding the early commercialization of transgenic plants, on 18 May 1994, the U.S. Food and Drug Administration ruled that the Flavr Savr tomato was as safe for human consumption as tomatoes that were bred by conventional means, and because Flavr Savr tomatoes were essentially the same as other tomatoes, special labeling was not required. Unfortunately, while the antisense polygalacturonase gene had a positive effect on shelf life, it did not affect the fruit's firmness and was not especially popular with consumers. Thus, after a few years, the Flavr Savr tomato was removed from the marketplace.

Lowering Ethylene Levels

The plant growth regulator ethylene induces the expression of a number of genes that are involved in fruit ripening and senescence and in flower wilting. Treatment of plants with chemical compounds that block ethylene production or perception delays these processes. Thus, premature fruit ripening and flower wilting can be avoided by blocking any one of several different steps in the ethylene biosynthesis pathway. For example, transgenic plants that have been engineered to contain antisense RNA versions of *S*-adenosylmethionine synthetase, ACC synthase, or ACC oxidase (Fig. 11.47A) have much lower than normal levels of ethylene. Similarly, the amount of ACC, and hence the amount of ethylene, that can be synthesized may be decreased by transforming plants with the bacterial enzyme ACC deaminase, which converts ACC to α-ketobutyrate (Fig. 11.47B). Each of these genetic manipulations results in a decreased level of ethylene and thereby extends the storage life of fruit and flowers.

By screening a large number of soil bacteria for the ability to utilize ACC as a sole source of nitrogen, strains that degrade ACC were identified. From one of these strains, the gene for the enzyme ACC deaminase was isolated based on the ability of transformed *E. coli* strains that expressed this gene

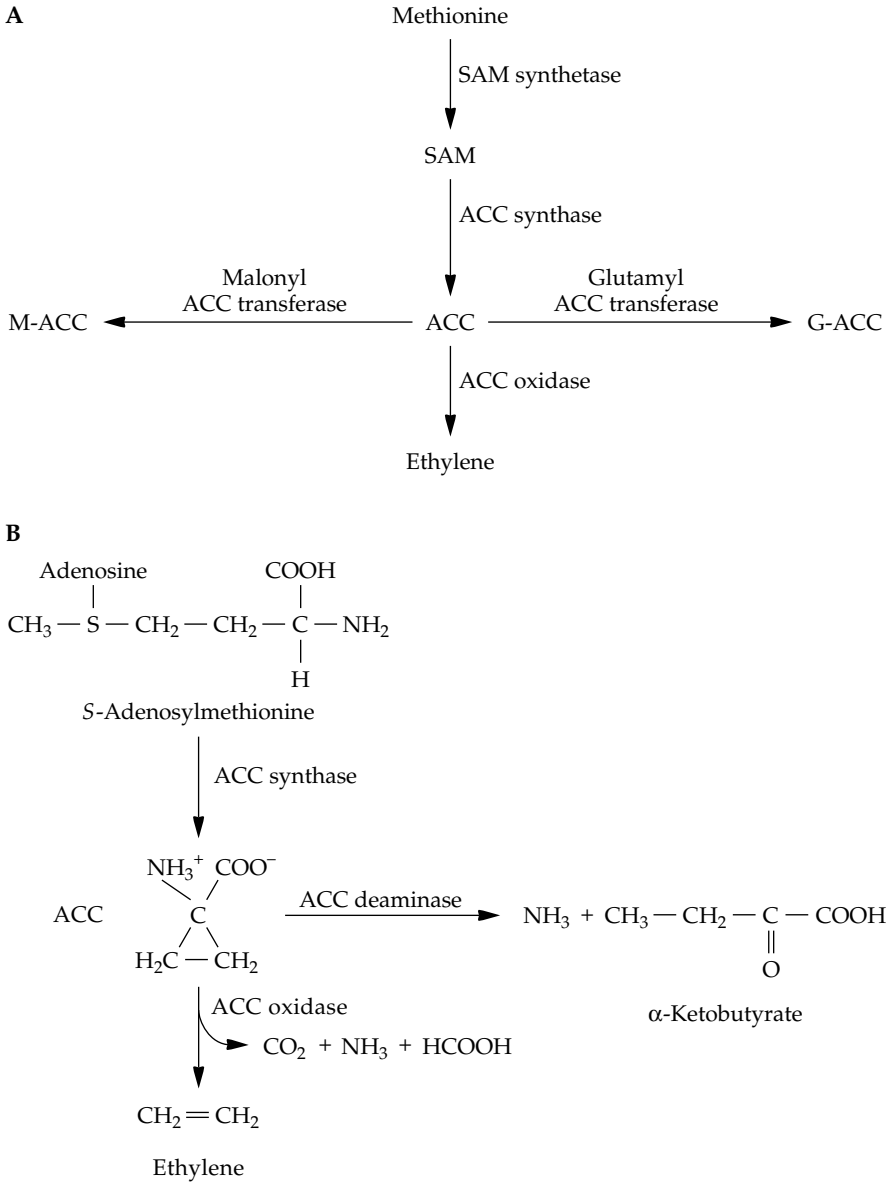


Figure 11.47 (A) Centrality of ACC in the synthesis of ethylene. M-ACC, 1-(m alonylamino)-cyclopropane-1-carboxylic acid; G-ACC, 1-(γ-l-glutamylamino) cyclopropane-1-carboxylic acid. **(B)** Inhibition of ethylene biosynthesis by genetic manipulation. Normally, ACC is synthesized from *S*-adenosylmethionine by the enzyme ACC synthase, and then ACC oxidase converts ACC to ethylene. Ethylene synthesis may be inhibited in transgenic plants either by an antisense mRNA version of ACC synthase, by ACC oxidase, which inhibits the synthesis of these enzymes, or by the enzyme ACC deaminase, which competes with ACC oxidase for the available ACC, producing ammonia and α-ketobutyrate rather than ethylene.

to grow on minimal medium containing ACC as the sole nitrogen source. This gene was subcloned, put under the control of the 35S promoter from *Cauliflower mosaic virus*, and expressed in tomato plants. The transgenic plants synthesized a much lower level of ethylene than did normal plants, and the fruit of the transgenic plants had a significantly longer shelf life.

These bioengineered changes result in fewer losses due to spoilage because of the much lower levels of ethylene. Similar results have been observed with transgenic cantaloupes that have lowered ethylene levels. In fact, this strategy to delay fruit ripening is effective with a range of fruits.

In the future, many plants will be engineered to have lower ethylene levels, primarily so that the rate of fruit ripening and flower wilting, or abscission, is inhibited. Fruits that are likely to be the targets of such genetic manipulation include melons, pineapples, and bananas; targeted flowers might include roses, carnations, tulips, chrysanthemums, and orchids.

CRISPR Mutants

The softening of tomatoes is important for the development of flavor but also affects fruit storage, transportability, and shelf life. Currently, the main method for slowing tomato fruit softening involves the use of hybrids containing nonripening mutations selected through traditional breeding. In one recent study, researchers created several CRISPR/Cas-generated mutants in several ripening-related genes, including pectate lyase, polygalacturonase 2a, and β-galactanase, and then tested these mutants for the physiochemical properties of the fruits. The results indicated that only mutations in pectate lyase resulted in firmer fruit. In contrast to the results obtained with tomato, strawberry softening is inhibited to a much greater extent by mutations in polygalacturonase 2a.

CRISPR/Cas9 may be utilized to generate a wide range of plant mutants. This technique has been used to modify the colors of flowers, including changing Japanese morning glory flowers from a dark violet to a pure white. This was achieved by mutagenesis of the plant's *dihydroflavonol-4-reductase-B* gene encoding. Other researchers reported modifying both copies of the *flavanone 3'-hydroxylase* gene in petunia plants. This modification also has the effect of changing the color of the flowers from dark violet to a pale purplish pink. Employing CRISPR/Cas9 is much simpler and easier than the previously reported methods of altering flower color with antisense RNA.

Modification of Plant Nutritional Content

Over the years, agronomists and plant breeders have been extremely successful in optimizing the useful properties (e.g., protein or oil content) and increasing the productivity (yield) of many crop plants. However, traditional breeding approaches to crop improvement are both difficult and slow, and they are intrinsically limited by the existing genetic content of cross-breeding strains. Conversely, the use of genetic engineering techniques allows scientists to dramatically speed up the process of developing plants with improved characteristics and to introduce traits that would otherwise be impossible to develop by traditional techniques. For example, genetic engineering has been used to improve the nutritional quality of several plants, including corn (maize), rice, and pea, and the fatty acid compositions of both edible and nonedible oil-producing crops.

Amino Acids

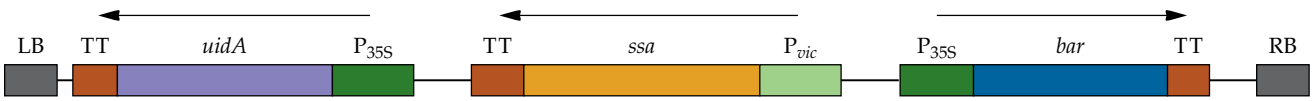
Seed storage proteins, which are used as sources of both carbon and nitrogen during seed germination, contain a limited number of amino acids, which are organized into repeating peptide units. Often, the

nutritional value of these proteins is deficient because they lack one or more of the amino acids, generally lysine or methionine, that are essential for human health. The amino acid composition of the seed storage proteins can be altered to a limited extent by breeding programs, but genetic engineering strategies can also be used.

The bulk of the lupine, a grain legume, that is produced annually in Australia (>800,000 tons) is used to feed cattle, pigs, and chickens. Unfortunately, like most other grain legumes, lupine is deficient in methionine and cysteine. Therefore, lupine feed is supplemented with methionine. To provide animals with more nutritious feed without methionine supplementation, lupines were engineered to express sunflower seed albumin, which is both stable in the rumen and unusually rich in the sulfur-containing amino acids methionine and cysteine (Fig. 11.48). Sunflower seed albumin escapes microbial breakdown in the rumen and is therefore available for digestion and absorption in the lower gastrointestinal tract. The transgenic lupine plants that expressed sunflower seed albumin were used as an animal feed, and as expected, rats that received transgenic lupines as their sole nitrogen source had significantly greater weight gains than did rats fed nontransgenic lupines, comparable to what would be expected if a nontransgenic lupine diet had been supplemented with pure methionine.

One novel way to increase the lysine content of seeds is to increase the production of lysine in transgenic plants by deregulating the lysine biosynthetic pathway. The amino acids lysine, threonine, methionine, and isoleucine are all derived from aspartic acid (Fig. 11.49A). The first step in the conversion of aspartic acid to lysine is phosphorylation of the aspartic acid by aspartokinase (AK) to produce β-aspartyl phosphate, which is then converted to aspartic β-semialdehyde. The condensation of aspartic β-semialdehyde with pyruvic acid to form 2,3-dihydrodipicolinic acid, which is catalyzed by dihydrodipicolinic acid synthase (DHDPS), is the first reaction in the pathway that is committed to lysine biosynthesis. Both AK and DHDPS are feedback-inhibited by lysine. Thus, to overproduce lysine, it is necessary to abolish the feedback inhibition of these two enzymes. This was accomplished by cloning naturally lysine feedback-insensitive genes for DHDPS and AK from *Corynebacterium* and *E. coli*, respectively, fusing each of these genes to a chloroplast transit peptide (to ensure that the two proteins are localized in seed plastids), placing each gene under the control of a seed-specific promoter, and then introducing the two genes on a Ti plasmid binary vector into canola and soybean plants (Fig. 11.49B). Transgenic canola and soybean plants had more than a 100-fold increase in the free lysine in their seeds, with an overall doubling of the total seed lysine content in canola and a 5-fold increase in the total lysine content in

Figure 11.48 The T-DNA construct used to transform lupines to increase the methionine content. The arrows indicate the direction of transcription. LB, left border; P_{35S}, the 35S promoter from *Cauliflower mosaic virus*; *uidA*, a bacterial marker gene encoding the β-glucuronidase; TT, transcription termination region, including a polyadenylation site; *ssa*, the sunflower seed albumin gene; P*vic*, the promoter from the pea vicilin gene; *bar*, a bacterial selectable gene that confers resistance to the herbicide phosphinothricin, which is used as a selectable marker; RB, right border.



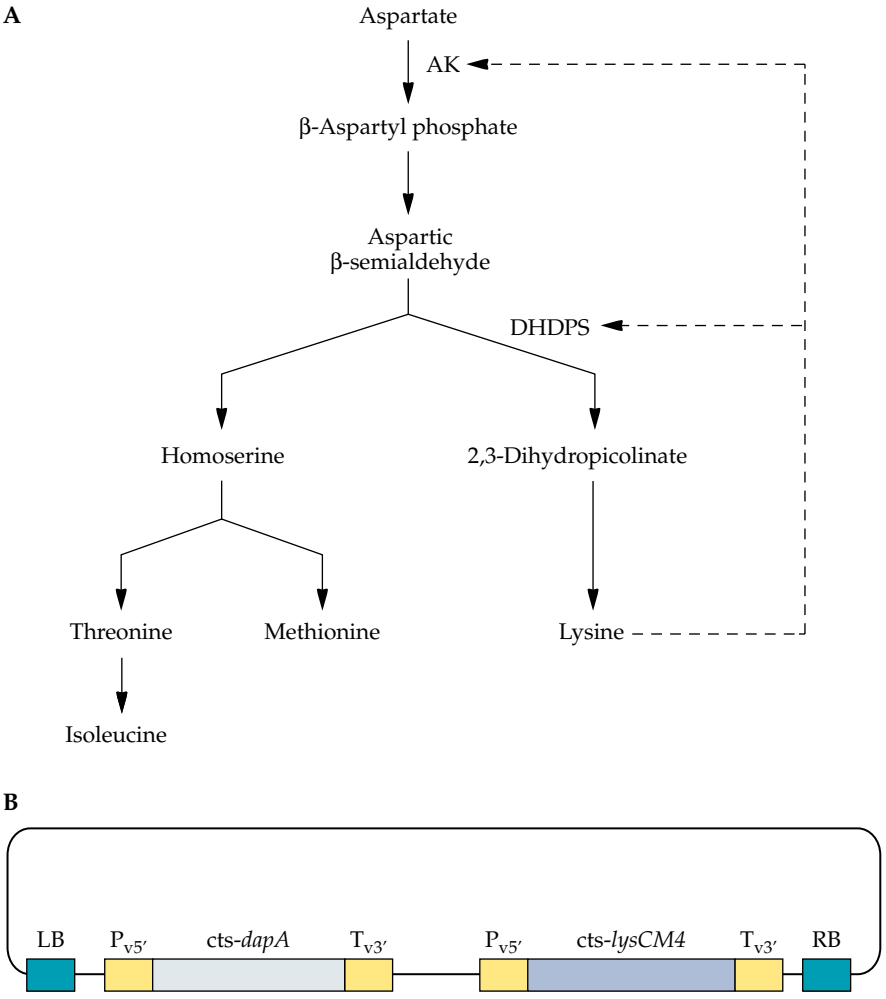


Figure 11.49 **(A)** The biosynthetic pathway for amino acids derived from aspartic acid. Not all of the steps and intermediates are shown. Feedback inhibition is shown by the dashed lines. DHDPS, dihydrodipicolinic acid synthase; AK, aspartokinase. **(B)** A Ti plasmid vector used to transform soybean and canola with genes that modify their lysine contents. $P_{v5'}$ the promoter from the bean β-phaseolin gene; $T_{v3'}$ the transcription termination region from the bean β-phaseolin gene; cts, the coding region for the chloroplast transit peptide from the small subunit of ribulose bisphosphate carboxylase; *dapA*, a gene from *Corynebacterium* that encodes a lysine-insensitive DHDPS; *lysCM4*, a mutant version of the *E. coli lysC* gene that encodes a lysine-insensitive AK. The left and right borders of the T-DNA are indicated by LB and RB, respectively.

soybean. Based on the approach described here, scientists have recently produced transgenic corn with high levels of lysine that has been approved (in several countries) for commercial use in the feeding of livestock.

Lipids

It has been estimated that annual global plant oil consumption was ~204 million metric tons in 2019/2020. More than 90% of this production is for human consumption in margarines, shortenings, salad oils, and frying

oils. Together, soybean, palm, canola (previously called rapeseed), and sunflower account for approximately 80% of worldwide plant oil production. For the most part, these oils consist of palmitic ($C_{16:0}$), stearic ($C_{18:0}$), oleic ($\Delta 9C_{18:1}$), linoleic ($\Delta 9,12C_{18:2}$), and linolenic acids ($\Delta 9,12,15C_{18:3}$), where the number after the C denotes the number of carbon atoms; the number after the colon is the number of C═C bonds; the number after the Δ indicates the position of the first carbon atom in C═C bonds, with the carboxyl group numbered as 1 (Fig. 11.50). In an effort to eat healthier foods, consumers have become concerned about the nutritional content of various edible oils. As can be seen in Table 11.8, the fatty acid content of edible oils can vary dramatically. Moreover, it is deemed desirable to have as low a level as possible of saturated fats, a high level of monounsaturated fats such as oleic acid (which lowers the undesirable low-density lipoproteins, or LDLs, without affecting the desirable high-density lipoproteins, or HDLs), and as high a level as possible of omega-3 fatty acids.

Canola and olive oil are considered to be among the healthiest and most nutritious edible oils. Olive oil is a key component of the so-called healthy Mediterranean diet and has been produced in southern Europe and northern Africa for ~2,500 to 3,000 years. In 1989, canola oil received an award from the American Health Foundation as the Health Product of the Year. Notwithstanding the enormous success that has been achieved

Figure 11.50 Chemical structures of common plant lipids.

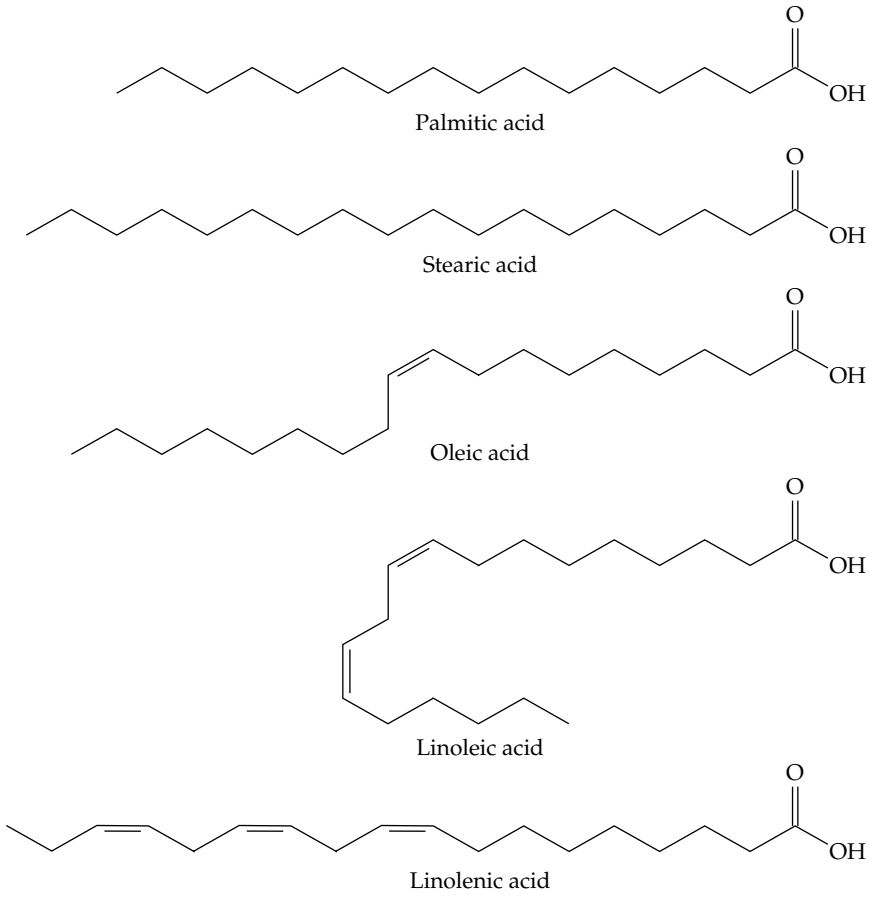


Table 11.8 Dietary fats present in various oils

Oil	Saturated fat (%)	Polyunsaturated fat (%)	Monounsaturated fat (oleic acid) (%)	Omega-3 fatty acids (%)
Canola	7	21	61	11
Safflower	10	76	14	Trace
Sunflower	12	71	16	1
Corn	13	57	29	1
Olive	15	9	75	1
Soybean	15	54	23	8
Peanut	19	33	48	Trace
Cottonseed	27	54	19	Trace
Palm	51	10	39	Trace
Butterfat	66	3	28	1
Coconut	91	2	7	None

in developing canola by conventional breeding, it is possible, by genetic engineering, to rapidly modify several genetic traits to continue to improve canola. For example, approximately 90% of the Canadian canola crop has been engineered to be herbicide resistant, and some varieties of canola have been engineered to be insect resistant. It is currently feasible to change the degree of unsaturation, i.e., the number of carbon–carbon double bonds, and to modify the chain lengths of fatty acids in canola plants by genetic manipulation, as it is desirable to have as low a level as possible of saturated fats and a high level of oleic acid. Several transgenic varieties of canola, each producing a different modified oil, have been created. Each transgenic variety contains one additional gene. For example, the production of shortening, margarine, and confectionery goods requires large amounts of stearate. One variety of transgenic canola contains an antisense copy of a *Brassica* stearate desaturase gene, which inhibits the expression of the normal canola gene and leads to the accumulation of stearic acid rather than the desaturation of stearic acid to oleic acid. Progress on the production of transgenic canola varieties with modified seed oil properties has been both rapid and impressive.

In 2019, a company in the United States, the world's largest producer of soybeans, announced that it had developed a genetically modified soybean oil that contained ~80% oleic acid, no trans fats, and much lower levels of saturated fatty acids compared to nonengineered soybean oil. The company that developed this product advertises that "no foreign DNA has been added to the product," as they used genome editing technology (chapter 6) in its development. In addition to the claimed increased health benefits of this modified soybean oil, the engineered product purportedly has a significantly extended shelf life compared to nonengineered soybean oil.

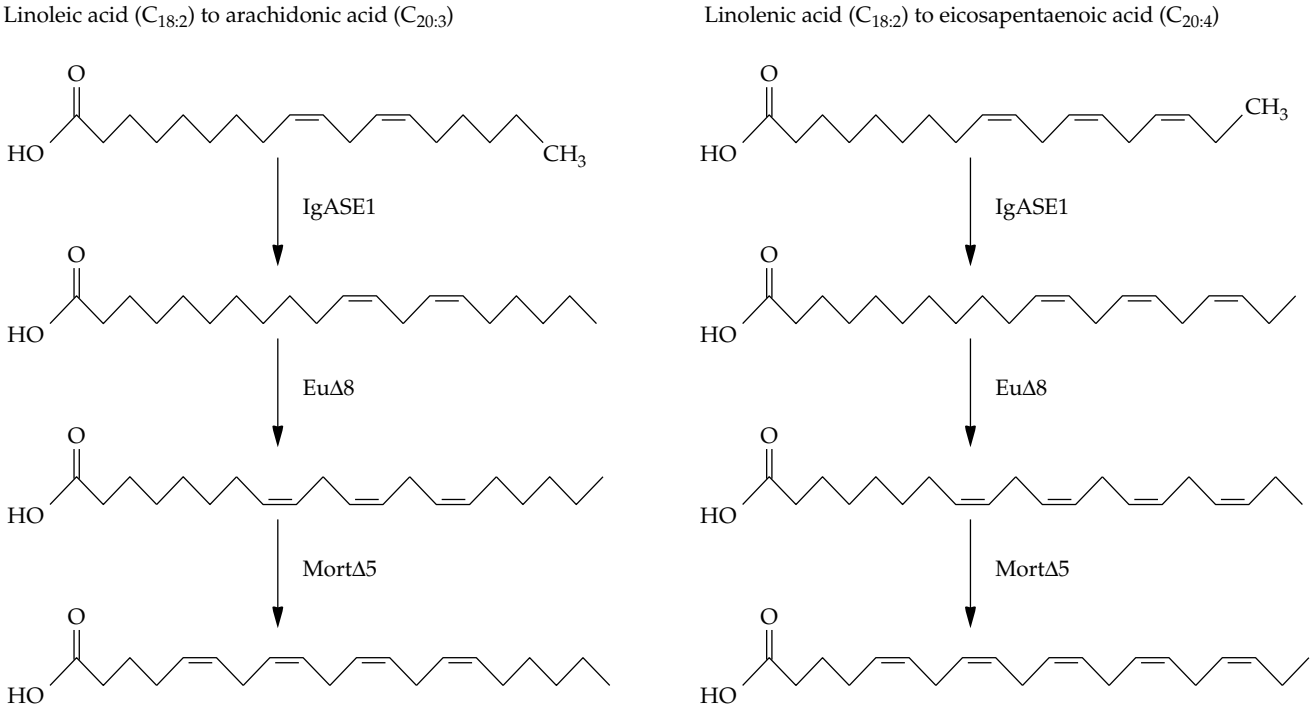
The omega-3 and omega-6 fatty acids are precursors for many prostaglandins and are therefore directly responsible for regulating several important human metabolic functions. They have both been implicated in lowering blood pressure, lowering LDL, and reducing the risk of heart disease. Until now, the major source for these important fatty acids has been marine and fish oils. However, for a variety of reasons, including the fact that global fish resources have declined dramatically in recent years, researchers have turned their attention to genetically engineering plants to produce safe, affordable, and renewable alternatives to the traditional sources. Given

the fact that plants can synthesize linoleic and α-linolenic acid, C-18 precursors of the long-chain omega-3 and omega-6 fatty acids, it may be possible to engineer plants to transform linoleic and α-linolenic acids into long-chain omega-3 and omega-6 fatty acids. In fact, when *A. thaliana* plants were transformed (in three separates stages) with three additional genes, the linoleic and α-linolenic acids were converted to arachadonic acid (an omega-6 fatty acid) and eicosapentaenoic acid (an omega-3 fatty acid), respectively (Fig. 11.51). The three introduced genes included a Δ9-specific elongase from the marine microalga *Isochrysis galbana*, a Δ8-desaturase from the protist *Euglena gracilis*, and a Δ5-desaturase from the fungus *Mortierella alpina*. Even though the engineered plants contained higher than normal levels of arachidonic acid and eicosapentaenoic acid, i.e., 7% and 3%, respectively, in their leaf tissue, plant growth and development were normal. Thus, it is possible to engineer pathways in plants to produce long-chain polyunsaturated fatty acids that are vital for human health. It is now necessary to engineer this pathway so that the three introduced genes are expressed in a seed-specific manner in an oilseed crop, such as soybean or canola.

Vitamins

It is estimated that micronutrient malnutrition, the lack of adequate vitamins and minerals, affected ~2 billion people globally in 2017. Micronutrient malnutrition has long-term effects on human health, learning ability,

Figure 11.51 Overview of the conversion of linoleic acid to arachidonic acid and linolenic acid to eicosapentaenoic acid in transgenic *A. thaliana* expressing three foreign genes (all under the control of the *Cauliflower mosaic virus* 35S promoter). IgASE1, Δ9-specific fatty acid elongase from *I. galbana*; EuΔ8, Δ8-desaturase from *E. gracilis;* MortΔ5, Δ5-desaturase from *M. alpina*.



and productivity. It is therefore a major impediment to socioeconomic development throughout the world. One way to address this problem is through the biofortification of staple crops such as rice, corn, wheat, cassava, and potato. Biofortification, improving the nutritional quality of existing crops, may be achieved in a number of different ways, including the genetic engineering of those crops to produce plants with increased levels of provitamin A; vitamins B_1, B_6, B_9, C, and E; as well as iron and iodine. To date, most of the published studies describe increasing the concentration of a single nutrient in a single crop species. However, this is just a first step, as many plants are deficient in multiple nutrients.

Vitamin A

Although rice (*Oryza sativa*) is the staple food of approximately half of the world's population, it is a poor source of several nutrients and vitamins, including vitamin A. About 250 million people worldwide, approximately half of whom are children, are deficient in vitamin A; this deficiency leads to 1 million to 2 million deaths per year and is a leading cause of vision impairment, including night blindness and total blindness for an estimated 500,000 children per year. One way to address the myriad health problems that result from vitamin A deficiency would be to engineer rice to produce the vitamin A precursor provitamin A (β-carotene). Mammals synthesize vitamin A from β-carotene, which is a common carotenoid pigment normally found in plant photosynthetic membranes. In the year 2000, an international consortium of scientists reported using *Agrobacterium*-mediated transformation to introduce the entire β-carotene biosynthetic pathway into rice (Fig. 11.52). The phytoene synthase and phytoene desaturase genes were introduced on a construct that did not contain any selectable marker. The lycopene β-cyclase gene was part of a separate construct that contained a selectable marker. The frequency of insertion of all three genes into the rice genome and their subsequent expression were quite high. Thus, the engineered rice produces β-carotene, which, after ingestion, is converted to vitamin A. It was thought that this strategy would facilitate the eventual development of transgenic strains of rice that not only produced high levels of β-carotene but did not contain any antibiotic resistance marker genes (removed as described in chapter 10). The transgenic rice that produces β-carotene has a yellow or golden color and has been called "Golden Rice" by the scientists involved in its development. Unfortunately, the initial version of Golden Rice, now called Golden Rice 1, synthesized only 1.6 mg of β-carotene per gram of rice, so that individuals would have had to consume around 3 kg of Golden Rice 1 each day to reach the recommended minimal daily requirement of vitamin A. However, in 2005, scientists reported replacing the daffodil phytoene synthase gene (Fig. 11.52) with a similar gene from corn that encodes an enzyme with a higher level of activity, resulting in a variety called Golden Rice 2 that produces a 23-fold-higher level of β-carotene than Golden Rice 1. In contrast to vitamin A, there are no harmful effects when individuals consume excess amounts of dietary β-carotene. The research that culminated in the development of Golden Rice was funded by several nonprofit agencies, and the companies that held the patents on the technologies that made this work possible agreed to forgo their usual royalties. Therefore, the rice was made freely available to farmers in the world's poorest countries.

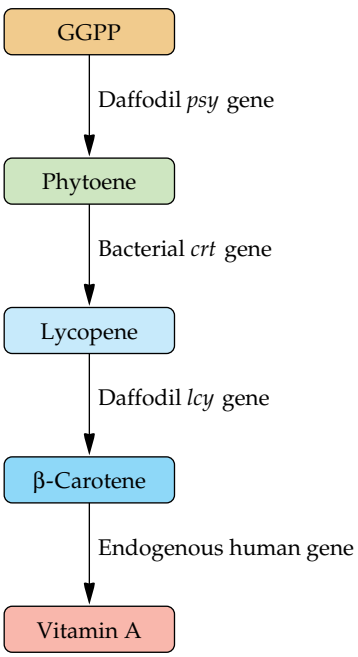


Figure 11.52 Biosynthesis of β-carotene in rice and vitamin A in humans. The daffodil phytoene synthase gene (*psy*) was controlled by a promoter from the rice seed storage protein glutelin. The phytoene desaturase (*crt*) gene was from the bacterium *Erwinia uredovora* and was controlled by the 35S promoter. The lycopene β-cyclase (*lcy*) gene originated from daffodil and was controlled by the rice glutelin promoter. All three genes were fused to transit peptides so that the proteins that they encoded would be transported into the plastid. GGPP, geranylgeranyl pyrophosphate.

One of the problems for the more widespread acceptance of Golden Rice 2 in Asia has been a general mistrust of genetically modified foods by some consumers. In addition, both Golden Rice 1 and Golden Rice 2 were produced from subspecies *japonica* cultivars of rice, which are popular with scientists but do not do well in the field in Asia. To remedy this situation, researchers have introduced the traits from Golden Rice 1 and Golden Rice 2 into the more popular subspecies *indica* varieties of rice by traditional genetic crossing of the *indica* strains with the engineered *japonica* cultivars. The first successful field trials of an *indica* variety of Golden Rice took place at the International Rice Research Institute in the Philippines in 2008 to 2009, with subsequent trials conducted in Taiwan and Bangladesh.

In November of 2013, Pope Francis gave his personal blessing to Golden Rice despite the fact that the Vatican did not provide an official statement of support for this technology. In April of 2015, the United States Patent and Trademark Office awarded the Golden Rice Project the Patents for Humanity award. This award is conferred upon patent owners who bring life-saving technologies to the underserved people of the world. In 2018, Golden Rice was approved as food in Australia, New Zealand, Canada, and the United States. In 2019, Golden Rice 2 was approved for both human and animal consumption in the Philippines.

Responding to the fact that delays in approving Golden Rice 2 in many countries was a political and not a scientific issue, in mid-2016, an international group of more than 110 winners of Nobel Prizes in Science signed an open letter to Greenpeace, a nongovernmental agency opposed to the use of Golden Rice; the United Nations Food and Agriculture Program; and governments around the world (Box 11.4). In this strongly worded letter, they argued that Golden Rice was both safe and needed, especially within the developing world.

Recently, based on the extensive studies previously reported with rice, other groups of scientists have developed golden bananas, golden cassava, and golden potatoes. Interestingly, the golden potatoes were also engineered to overproduce vitamin E so that a 150-gram serving of boiled golden potatoes provided children with ~42% of their daily requirement of provitamin A and ~34% of their daily requirement of vitamin E.

Folate

Tetrahydrofolate, or vitamin B9, is an essential micronutrient that is a necessary component of human diets. While the recommended daily intake of this vitamin is only 400 μg, the diets of many individuals in developing countries are chronically deficient in tetrahydrofolate. Tetrahydrofolate deficiency can result in anemia and severe birth defects, including neural tube defects. For approximately 3 billion people, most of them in developing countries, rice provides around 80% of the daily caloric intake; however, rice is a poor source of tetrahydrofolate, as well as many other micronutrients. To increase the nutritive value of rice, the pathway for the production of tetrahydrofolate and tetrahydrofolate polyglutamates may be genetically manipulated.

Folates are tripartite molecules; that is, they consist of three separate parts: pteridine, *p*-aminobenzoic acid, and one or more glutamate moieties (Fig. 11.53A). In plants, the pteridine precursors are synthesized from GTP in the cytosol, *p*-aminobenzoic acid is synthesized from chorismate in the chloroplasts, and then both the pteridine precursor and *p*-aminobenzoic

Nobel Laureates Supporting Transgenic Plants – 2016

box 11.4

To the Leaders of Greenpeace, the United Nations, and Governments around the World

The United Nations Food & Agriculture Program has noted that global production of food, feed and fiber will need approximately to double by 2050 to meet the demands of a growing global population. Organizations opposed to modern plant breeding, with Greenpeace at their lead, have repeatedly denied these facts and opposed biotechnological innovations in agriculture. They have misrepresented their risks, benefits, and impacts, and supported the criminal destruction of approved field trials and research projects.

We urge Greenpeace and its supporters to re-examine the experience of farmers and consumers worldwide with crops and foods improved through biotechnology, recognize the findings of authoritative scientific bodies and regulatory agencies, and abandon their campaign against "GMOs" in general and Golden Rice in particular.

Scientific and regulatory agencies around the world have repeatedly and consistently found crops and foods improved through biotechnology to be as safe as, if not safer than, those derived from any other method of production. There has never been a single confirmed case of a negative health outcome for humans or animals from their consumption. Their environmental impacts have been shown repeatedly to be less damaging to the environment and a boon to global biodiversity.

Greenpeace has spearheaded opposition to Golden Rice, which has the potential to reduce or eliminate much of the death and disease caused by a vitamin A deficiency (VAD), which has the greatest impact on the poorest people in Africa and Southeast Asia.

The World Health Organization estimates that 250 million people suffer from VAD, including 40 percent of the children under five in the developing world. Based on UNICEF statistics, a total of one to two million preventable deaths occur annually as a result of VAD, because it compromises the immune system, putting babies and children at great risk. VAD itself is the leading cause of childhood blindness globally affecting 250,000–500,000 children each year. Half die within 12 months of losing their eyesight.

WE CALL UPON GREENPEACE to cease and desist in its campaign against Golden Rice specifically, and crops and foods improved through biotechnology in general; WE CALL UPON GOVERNMENTS OF THE WORLD to reject Greenpeace's campaign against Golden Rice specifically, and crops and foods improved through biotechnology in general; and to do everything in their power to oppose Greenpeace's actions and accelerate the access of farmers to all the tools of modern biology, especially seeds improved through biotechnology. Opposition based on emotion and dogma contradicted by data must be stopped.

How many poor people in the world must die before we consider this a **"crime against humanity"**?

Sincerely,

110 Nobel Laureates

acid are imported into the mitochondria, where the final molecule is assembled and modified (Fig. 11.53B). To express high levels of folate in rice, the first genes in each of the chloroplastic and cytosolic pathways, starting from chorismate and GTP, respectively, were introduced from *A. thaliana* so that the two enzymes encoded by these genes were overexpressed (Fig. 11.53B). In one transgenic line, the level of tetrahydrofolate was approximately 100 times the level found in nontransgenic rice plants, which was more than sufficient to meet the recommended daily dietary amount of this vitamin. In another study by a different research group, similar genetic manipulations led to the enhancement of the level of the vitamin in transgenic tomatoes.

Iron

The World Health Organization has estimated that iron deficiency affects approximately 30% of the world's population and is especially problematic where vegetable-based diets are the primary food source. Although several crops are rich in iron, absorption of iron is often prevented by the phytic acid that is present in many plants. As a first step toward developing food

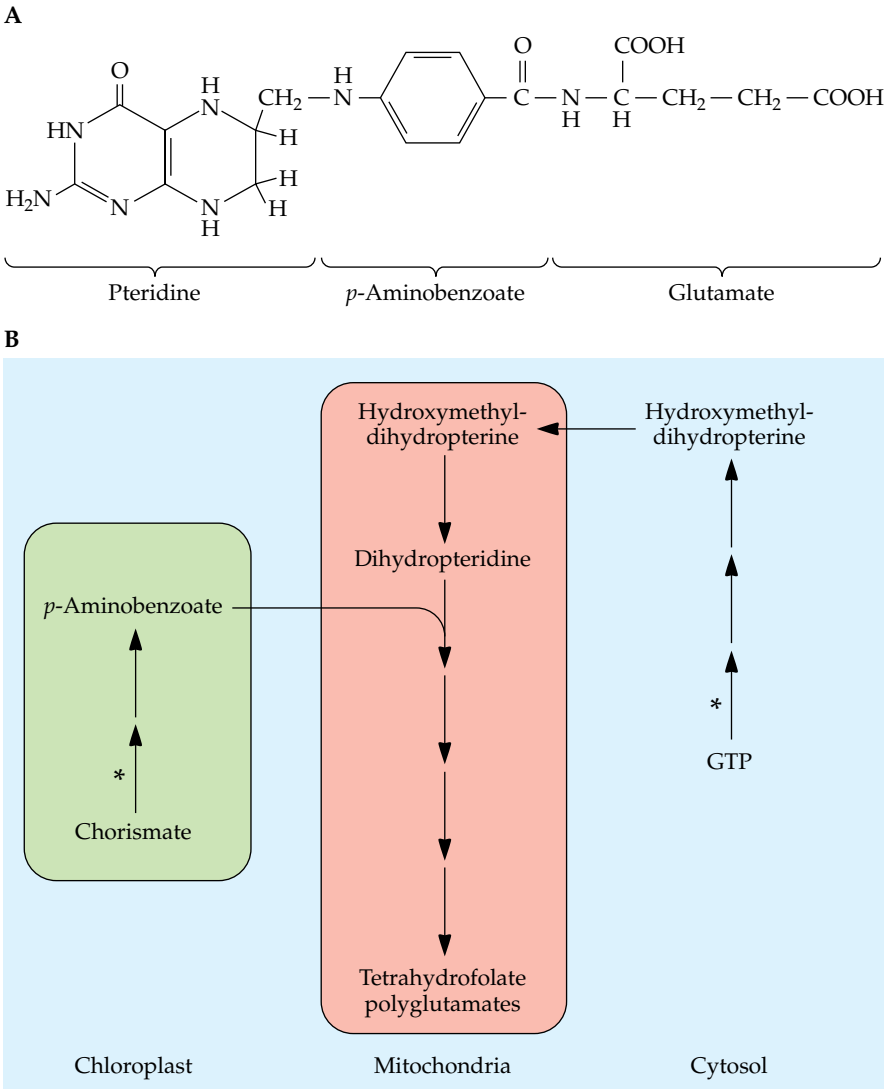


Figure 11.53 (A) Structure of tetrahydrofolate. **(B)** Major steps in the pathway leading to the synthesis of tetrahydrofolate and tetrahydrofolate polyglutamates in rice. A portion of the molecule is synthesized in the chloroplast and a portion in the cytosol, and the molecule is assembled in the mitochondria. Two enzymatic steps (marked by asterisks) that represent rate-limiting steps in the biosynthetic pathway were engineered by the addition of foreign genes to increase the flux through the pathway.

crops with sufficient levels of iron to prevent iron deficiency anemia, scientists engineered rice plants to express the soybean protein ferritin (Fig. 11.54). Ferritin is an iron storage protein that is found in animals, plants, and bacteria and carries up to 4,500 iron atoms in its central cavity, which is formed from the interaction of 24 monomeric ferritin subunits. The soybean ferritin cDNA was cloned into a binary vector under the transcriptional control of the rice seed storage protein glutelin promoter (Fig. 11.54), and the entire construct was introduced into plants by electroporation. In this case, soybean ferritin was expressed in the seeds of rice plants and not in any other tissues. As a result, the iron content of rice seeds per gram

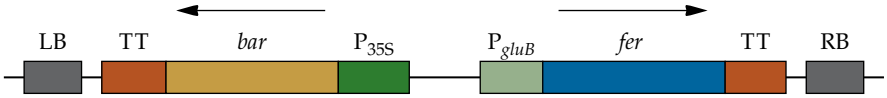


Figure 11.54 The T-DNA construct used to transform rice and produce ferritin in the seeds. The arrows indicate the direction of transcription. LB, left border; TT, transcription termination region; *bar*, the bacterial phosphinothricin acetyltransferase gene used as a marker gene; P_{35S}, 35S promoter from *Cauliflower mosaic virus*; P_{gluB}, the promoter from the rice seed storage protein glutelin; *fer*, the soybean ferritin-encoding cDNA; RB, right border.

(dry weight) of tissue was increased to approximately 2.5 times the original value, while the iron content of leaves, stems, and roots did not change to any significant extent. Based on a typical daily adult portion of 150 grams of rice, the transgenic ferritin rice provides approximately 30 to 50% of the recommended daily adult requirement of iron.

More recent experiments have focused on the *indica* rice strain IR64, which is said to have good market value because of its preferred milling properties and good cooking quality. Rice grains from transgenic lines of this strain had more than a 3-fold increase in iron concentration compared to the wild-type strain. Moreover, the transgenic plants that overexpressed soybean ferritin were constructed to be marker-free, making the transgenic plants more acceptable for environmental release.

Gluten

Celiac disease is a chronic, immune-mediated disease of the gastrointestinal system that is caused by the ingestion of gluten proteins from wheat, barley, rye, and related grains. Celiac disease occurs worldwide, typically affecting approximately 1 to 1.5% of the population. The disease has a high degree of heritability. Symptoms of celiac disease include diarrhea and malabsorption of food as well as lesions in the small intestine. At the present time, the only treatment that is available for celiac disease patients is a strict gluten-free diet. The frequency of celiac disease notwithstanding, in recent years in the United States and Canada, up to 20% of the population have become gluten avoiders, believing that the consumption of gluten is deleterious to their overall health. This gluten-free craze has been popularized by several celebrities who have, without any real evidence, actively promoted a gluten-free diet as a key component in their healthy lifestyle.

Gluten is a mixture of proteins found in wheat and related grains. It is a composite of storage proteins (prolamins) that gives elasticity to dough, helping it to rise and maintain its shape, often giving the final product a chewy texture. Wheat gluten includes two protein families: glutenins and gliadins. The glutenins comprise both high- and low-molecular-weight fractions, while the gliadins can be divided into three structural types: α-, γ-, and ω-gliadins. Celiac disease patients typically have their greatest sensitivity to epitopes that are very rich in proline and glutamine residues, with the majority of these epitopes being found in the gliadin fraction.

A group of scientists has engineered bread wheat that produced only a very low level of gliadins. To do this, scientists had to contend with the fact that gliadins are encoded by a multigene family so that several gliadin genes had to be silenced. Their approach was based on the use of

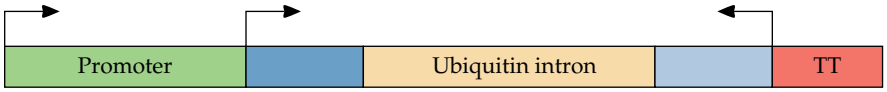


Figure 11.55 A genetic construct used to abolish or decrease gliadin gene expression. The construct includes a promoter (green with the arrow indicating the direction of transcription), the sense version of the 361-bp target region (dark blue with an arrow indicating the sense orientation), a spacer consisting of a ubiquitin intron (yellow), the antisense version of the 361-bp target region (light blue with an arrow indicating the antisense orientation), and a transcription terminator region (red, TT).

RNAi hairpin vectors. These vectors consist of a promoter and terminator region flanking an inversely repeated sequence of the target gene, with a spacer region between the repeats such that the transcribed product forms a hairpin by intrastrand complementary base pairing (Fig. 11.55). For this purpose, workers engineered a set of hairpin constructs containing a 361-bp DNA fragment that is highly conserved among wheat α-, γ-, and ω-gliadins. Following the introduction of these constructs, the resultant transgenic wheat plants produced significantly decreased levels of these proteins. Transgenic wheat lines with low levels of gliadins could be used as a source of food that is tolerated by many celiac disease patients. This work shows that it is possible to engineer gluten-reduced varieties of wheat. The hope is that it may be possible in the future to reduce the amount of gluten in some grains even further. More recently, another group of scientists engineered the wheat genome using CRISPR/Cas to produce a decreased level of gliadins. In these genome-edited strains of wheat the immunoreactivity of gliadins was reduced by 85%.

Modification of Food Plant Taste and Appearance

Preventing Discoloration

The postharvest discoloration of fruits and vegetables is a considerable problem for the food industry. A lack of acceptance of discolored foods by consumers has been dealt with by the food industry through the use of chemical additives in a wide range of foods. However, the safety of some of these food additives, in particular sulfites, has been questioned.

The enzymes that are thought to be responsible for the initial step in the discoloration of fruits and vegetables, the oxidation of monophenols and *o*-diphenols to *o*-quinones, are polyphenol oxidases. These nucleus-encoded enzymes with a molecular weight of around 59,000 are localized in chloroplast and mitochondrial membranes.

The contention that inhibition of the enzyme polyphenol oxidase would decrease the extent of discoloration has been tested with transgenic potatoes carrying several different polyphenol oxidase cDNA constructs. Vectors were constructed with either the full-length or partial potato polyphenol oxidase cDNA in either the sense or antisense orientation under the control of the *Cauliflower mosaic virus* 35S promoter, the granule-bound starch synthase promoter, or the patatin type I promoter, where patatin is an abundant glycoprotein in potato plant tubers (Fig. 11.56). The latter two are strong potato tuber-specific promoters. The two commercial varieties of potato that were transformed with these constructs are considered to

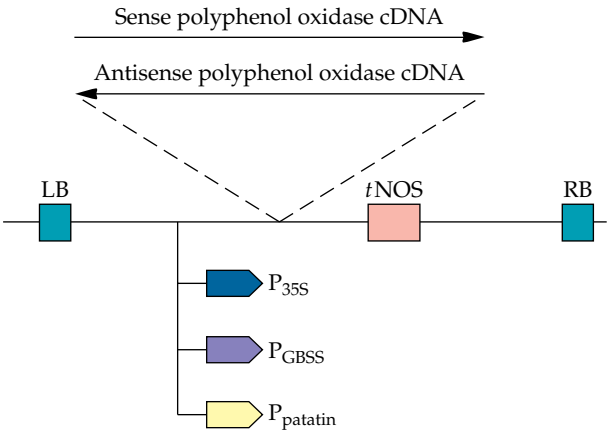


Figure 11.56 Sense and antisense polyphenol oxidase gene constructs. Transcription of either the sense or antisense cDNA is separately under the control of the *Cauliflower mosaic virus* 35S promoter (P_{35S}), the granule-bound starch synthase promoter (P_{GBSS}), or the patatin type I promoter ($P_{patatin}$) with the nopaline synthase transcription terminator region (*t*NOS). RB and LB, right and left borders of the T-DNA, respectively. Data from Bachem et al., *Bio/Technology* **12:**1101–1105, 1994.

have a good level of intrinsic resistance to black spot (enzymatic discoloration by polyphenol oxidase), so any increase in black spot resistance by genetic manipulation would be greater than what could be attained by traditional breeding techniques. Transgenic plants with the polyphenol oxidase cDNA constructs were deliberately bruised, and then the extent of black spot damage was assessed. Most of the transgenic potato plants with an antisense version of the polyphenol oxidase gene under the control of either the *Cauliflower mosaic virus* 35S promoter or the granule-bound starch synthase promoter were significantly more resistant to black spot than the nontransformed potatoes. The construct with the patatin promoter did not prevent polyphenol oxidase accumulation. The sense constructs all synthesized increased amounts of polyphenol oxidase and showed larger amounts of black spot than nontransformed control plants. More recently, scientists have shown that the bulk of the unwanted browning in potatoes may be effectively suppressed using artificial micro-RNAs (amiRNAs) directed against one of the several potato polyphenol oxidase genes.

Another approach to engineering crops to resist browning was reported in the spring of 2016. Scientists engineered the edible fungus *Agaricus bisporus* (the common white button mushroom) to resist browning. This was done by deleting a few base pairs in one of the mushroom's genes that encodes the enzyme polyphenol oxidase using the CRISPR/Cas9 system. By knocking out just one of the six polyphenol oxidase genes in the mushroom's genome, the total activity of that enzyme within modified mushrooms was reduced by 30%. Interestingly, when the genetically modified mushrooms were submitted to the Animal and Plant Health Inspection Service of the U.S. Department of Agriculture, it was ruled that these mushrooms did not need to be regulated. This ruling was based on the fact that the CRISPR/Cas9 system merely edited the mushroom's existing genome without introducing any foreign DNA. This opens up the prospect of rapidly bringing these modified mushrooms, and other CRISPR/Cas9-modified fruits and vegetables, to the marketplace.

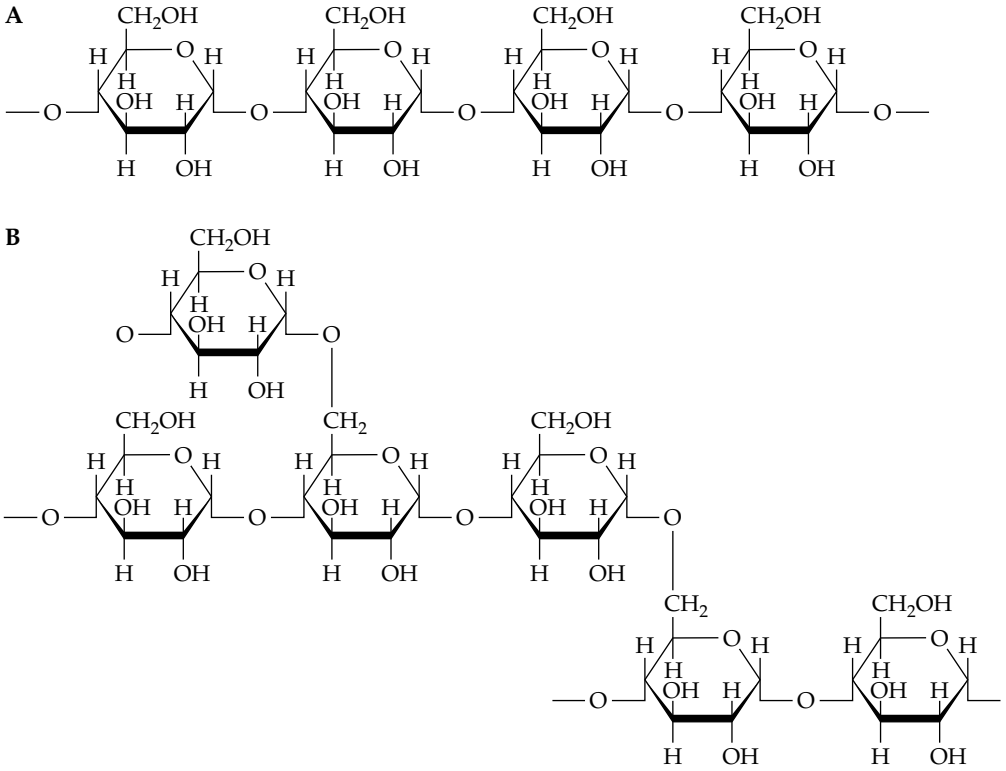
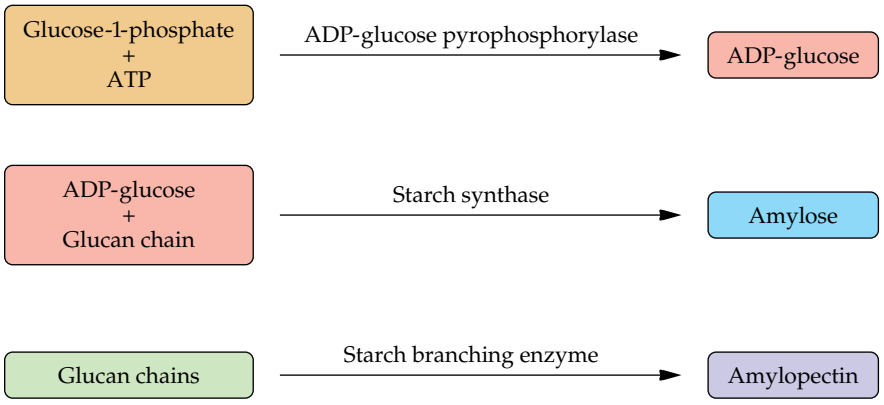


Figure 11.57 **(A)** A portion of an amylose chain with only α-1,4 linkages; **(B)** a portion of an amylopectin chain with α-1,4 and α-1,6 linkages.

Starch

The starch that is found in most crop plants, such as potato, consists of 20 to 30% (straight-chain) amylose and 70 to 80% (branched-chain) amylopectin chains (Fig. 11.57). The ratio of amylose to amylopectin has a large influence on the physical and chemical properties of the starch. For many industrial applications, it would be useful to have starch that is highly enriched in either amylose or amylopectin. Starch is normally synthesized in a stepwise process (Fig. 11.58). The enzyme ADP-glucose

Figure 11.58 The major reactions in the biosynthesis of starch.



pyrophosphorylase catalyzes the transfer of ADP (from ATP) to glucose-1-phosphate to form ADP-glucose. The enzyme starch synthase catalyzes the transfer of glucose from ADP-glucose to the nonreducing end of a preexisting glucan chain, which is a short version of an amylose chain, by means of an α-1,4 linkage. Branching can occur, catalyzed by a starch-branching enzyme, when two glucan chains are joined by an α-1,6 linkage. To generate potatoes with a high percentage of amylose and a corresponding low percentage of amylopectin, plants were transformed by using *A. tumefaciens* with antisense versions of the starch-branching enzyme under the transcriptional control of the 35S promoter. In transgenic lines that exhibited only approximately 1% of the normal amount of the starch-branching enzyme activity, the fraction of amylose increased from around 28% to 60 to 89% of the starch content. This is an initial step toward developing potatoes with unique starch compositions that may be used either for food or as a source of unusual starches with unique industrial properties.

The freezing and thawing that are common in the development and use of frozen foods often lead to unwanted changes in the texture of the starch component, making the product less attractive to consumers. Under these conditions, the starch can separate into a solid phase and a liquid phase, a process known as syneresis. The long, unbranched amylose chains have a much greater tendency to form separate phases following freezing and thawing than the shorter, branched amylopectin chains. One approach to creating a starch that could withstand freezing and thawing would be to modify the starch structure by decreasing the synthesis of amylose chains. This was achieved by transforming potatoes with antisense versions of several different starch synthase genes (Fig. 11.58). This reduced the expression of each of these genes to very low levels and resulted in a dramatic decrease in the amount of amylose and, at the same time, in the length of the amylopectin chains. These modified potatoes are currently being tested to ensure that a variety of physical and chemical properties are unaltered so that they can be used in the production of frozen foods.

Since starch is both inexpensive and very abundant, it is used in a number of industrial processes as a thickener, a gelling agent, or an adhesive. In addition, it is a substrate for the production of high-fructose syrups and ethanol. The production of high-fructose syrups requires the starch to be degraded, usually at high temperature, by the enzyme α-amylase and then treated, also at high temperature, with glucose isomerase, which converts the glucose (released from the starch that was degraded by α-amylase) into fructose. The enzymes that catalyze these reactions account for a major portion of the cost of the industrial conversion of starch into fructose. Therefore, reducing the cost of the enzymes makes the overall process more attractive. With this in mind, researchers constructed a bifunctional enzyme containing the essential regions of both α-amylase and glucose isomerase; PCR was used to amplify the genes for each enzyme from thermotolerant bacteria and to introduce restriction enzyme cut sites (see chapter 2) that, following digestion, were used to fuse the genes together (Fig. 11.59). Of course, care was taken to ensure that the chimeric gene retained the correct reading frames of the two genes. When the bifunctional enzyme was synthesized in *E. coli*, both α-amylase and glucose isomerase activities

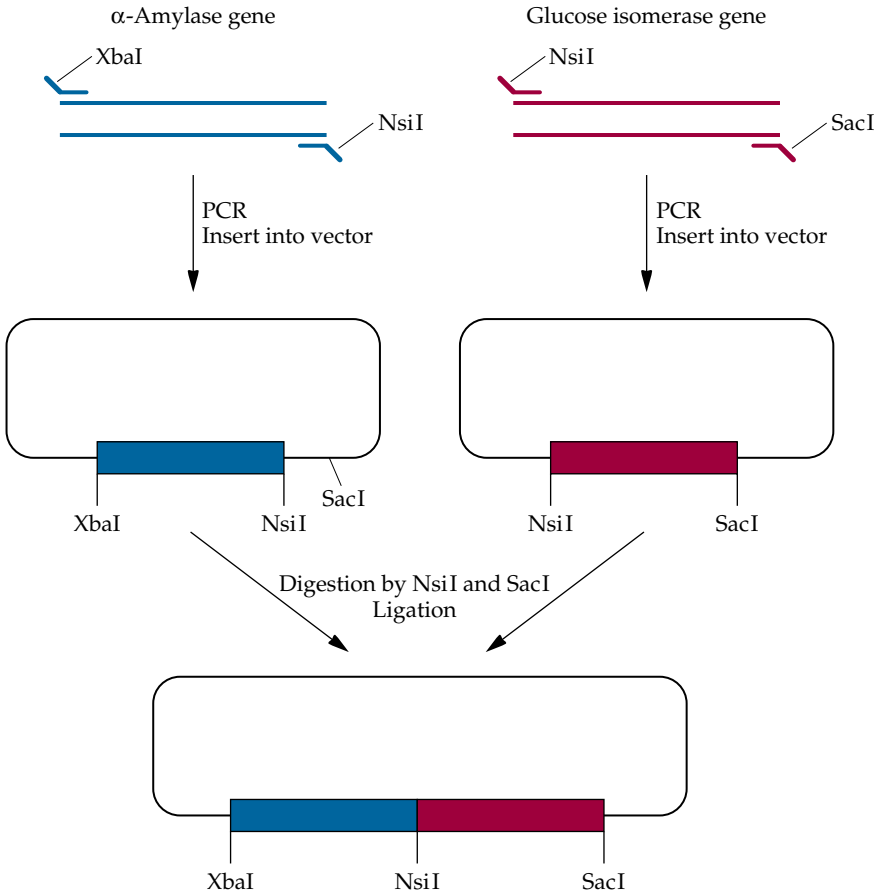


Figure 11.59 Construction of an α-amylase–glucose isomerase fusion gene. The α-amylase and glucose isomerase genes were from *Bacillus stearothermophilus* and *Thermus thermophilus,* respectively. The portions of the PCR primers that do not match the target genes are designed to retain both genes in the same reading frame and to include appropriate restriction endonuclease sites. Data from Beaujean et al., *Biotechnol. Bioeng*. **70:**9–16, 2000.

were found, and both activities were identical to that found in the thermotolerant source organisms. The chimeric gene was introduced into potato plants under the transcriptional control of the granule-bound-starch synthase promoter. Biochemical analysis performed on transgenic potatoes, after treatment at 65°C (the optimal temperature for the α-amylase), revealed a 3.9-fold and a 14.7-fold increase compared with nontransgenic potatoes in the concentrations of glucose and fructose, respectively. At 25°C, the glucose and fructose concentrations were the same in transgenic and nontransgenic potatoes. Thus, both enzymes function optimally at high temperatures and only very little or not at all at low temperatures. This work indicates that it is technically feasible to engineer potatoes to produce their own α-amylase and glucose isomerase and that the enzyme that is produced can then be used directly, without the need for any purification, in a process directed toward producing fructose.

Plants as Bioreactors

Plants are relatively easy to grow and can generate considerable biomass. With these features in mind, research has been carried out to determine whether transgenic plants can be used for the production of commercial proteins and chemicals. Unlike recombinant bacteria, which are grown in large bioreactors, a process that requires both highly trained personnel and expensive equipment, plants can be produced relatively inexpensively by less-skilled workers (Table 11.9). In addition, when proteins that are intended for human use are produced in transgenic plants, there is a significantly reduced risk of mammalian virus contamination in comparison to proteins that are produced in animal cells grown in culture. Ultimately, the biggest hurdle to overcome in the production of foreign proteins in plants is the purification of the product of a transgene from the mass of plant tissue.

Chinese hamster ovary (CHO) cells under optimal conditions, in stirred tank bioreactors, can produce up to 10 grams of monoclonal antibody per liter of culture, with the largest available fermenter having a volume of 25,000 liters and the fermentation requiring ~12 days; assuming perfect conditions, the yield of monoclonal antibodies would be ~7.5 metric tons per year. This amount is insufficient to meet the anticipated demand for some highly effective monoclonal antibodies. On the other hand, the same amount of monoclonal antibody could be produced from transgenic tobacco in a field of less than 1 square km.

To date, on a laboratory scale, plants have been used to produce monoclonal antibodies and antibody fragments (see chapter 10, Table 10.2); the polymer polyhydroxybutyrate, which can be used to make a biodegradable plastic-like material; and a number of potential therapeutic agents (Table 10.1) and vaccine antigens (Table 10.3). It is likely that many of these plant-produced products will be approved for more widespread use within the next few years.

Antibodies

The production of antibodies and antibody fragments in transgenic plants has several potential advantages over their synthesis in recombinant microbial cells (Table 11.9). For example, transformation of plants generally results in the stable integration of the foreign DNA into the plant

Table 11.9 Comparison of recombinant protein production in plants and other systems

Parameter	Bacteria	Yeast	Mammalian cell culture	Transgenic plants
Glycosylation	None	Incorrect	Correct	Generally correct
Assembles multimeric proteins	Limited	Limited	Limited	Yes
Production costs	Medium	Medium	High	Low
Protein-folding accuracy	Low	Medium	High	High
Protein yield	High	High	Medium	Medium
Scale-up costs	High	High	High	Low
Time required	Low	Low	High	Medium
Skill level required for growth	Medium	Medium	High	Low

genome, while most microorganisms are transformed with plasmids that can be lost during a prolonged or large-scale fermentation. In addition, the processing and assembly of foreign proteins in plants are similar to those in animal cells, whereas bacteria do not readily process, assemble, or posttranslationally modify eukaryotic proteins. Moreover, plants are inexpensive to grow on a large scale, and their production is not limited by fermentation capability; it is estimated that it costs approximately $5,000 per gram to produce antibodies from hybridoma cells in culture, $1,000 per gram to produce antibodies from transgenic bacteria, and $10 to $100 per gram to produce antibodies from transgenic plants. However, since most harvested plant tissues cannot usually be stored for long periods, foreign proteins might be produced in seeds, where they will be stable for long periods under ambient conditions. To date, a large number of antibodies, including immunoglobulin G (IgG), IgM, single-chain Fv fragments, and Fab fragments, have been produced in plants (see chapter 10, Table 10.2). Some of these plant-produced antibodies (sometimes called plantibodies) have been purified for possible use as diagnostics and human therapeutics, and others have been used to protect the plant against certain pathogenic agents, such as viruses. In a small number of instances, plant-produced antibodies are being tested in clinical trials to determine whether they are essentially equivalent to antibodies produced in other host cells.

The effective industrial-scale production of antibodies synthesized by plants has until recently been hampered by a very low yield, typically in the range of 1 to 10 μg/gram of fresh biomass. Moreover, with conventional transgenic plant technology, it is estimated that it takes about 2 years from the beginning of the cloning process to produce gram quantities of antibody. The use of transient-expression systems can significantly speed up this process (see chapter 10), but it is still problematic to coordinate the synthesis and assembly of the two different polypeptides that are integral components of antibody molecules (see chapter 2). Traditional transient-expression systems are limited by the low infectivity of the viral vectors, especially those carrying medium-size to large inserts. It is also necessary to ensure that transient-expression systems based on viral vectors do not spread to plants other than the intended host. This is typically done by using mutant plant viruses that cannot produce capsid (coat) proteins and therefore are unable to form functional viral particles. However, these systems generally produce only very low levels of active antibody. Fortunately, an alternative transient-expression system has been developed. This expression system involves coinfection of plant cells with two separate plant virus vectors, one based on *Tobacco mosaic virus* and the other on *Potato virus X* (with both viruses being capable of infecting a wide range of plants), so that the two vectors do not compete with one another but, rather, can coexist within the same cell. These vectors coreplicate within the plant, with each vector expressing a different antibody chain; i.e., one expresses the light chain and the other expresses the heavy chain (Fig. 11.60). This system has been used to increase the amount of IgG antibody that is typically synthesized by a plant transient-expression system by about 100-fold, so that it is possible to produce up to 0.5 mg of assembled IgG antibody per gram of fresh leaf biomass. At this high level of expression, it is possible to grow transgenic plants that produce specific IgG molecules in small areas indoors in controlled greenhouses, thereby avoiding environmental concerns regarding the inadvertent "escape" of

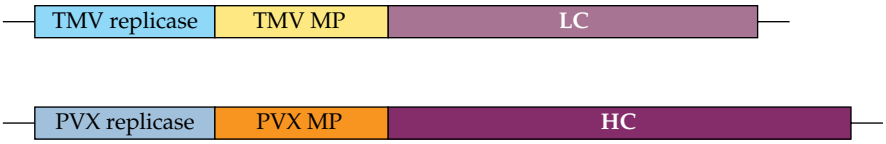
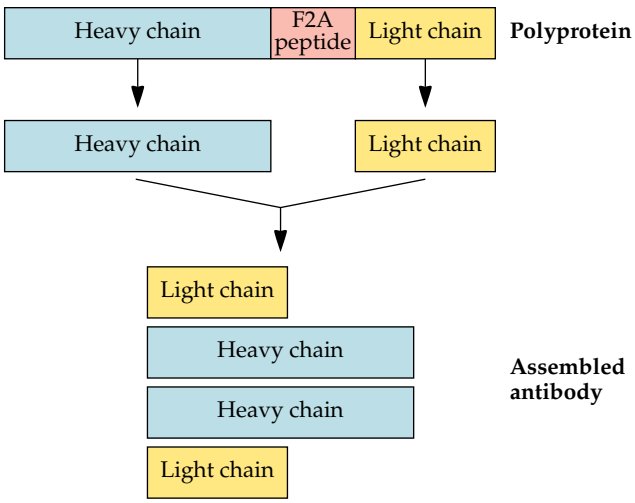


Figure 11.60 A portion of the viral vectors used to produce full-size IgG antibodies in plants. In each case, the viral replicase and movement protein (MP) are cloned together with the gene (cDNA) for either a light chain (LC) or a heavy chain (HC). Each antibody gene is controlled by a promoter, a signal peptide (to ensure secretion), and a transcription termination region from the appropriate virus, none of which are shown. The viruses used included *Tobacco mosaic virus* (TMV) and *Potato virus X* (PVX). In both cases, the viruses were unable to replicate because of the absence of the viral coat protein. Recombinant viral genomes were introduced into plants as part of the T-DNA that is transferred during *A. tumefaciens* infection.

antibody-producing plants. The perceived economic potential of this technology recently provided the incentive for the sale of the company where the technology was originally developed to a larger company that hopes to open a clinical-grade manufacturing plant with the objective of beginning clinical trials of the antibodies produced.

To synthesize monoclonal antibodies in plants requires that the level of expression of the light and heavy chains be identical so that an antibody molecule with two identical light and two identical heavy chains can be assembled (see chapter 4). One way to do this is to synthesize the light and heavy chains as part of a polyprotein where the two polypeptides are linked by a self-cleaving foot-and-mouth disease virus peptide (F2A). Following synthesis of the polypeptide in plant tissue, the F2A peptide is cleaved and the antibody light and heavy chains, produced in identical levels, are assembled into a functional antibody molecule (Fig. 11.61). The F2A peptide had previously been used in mammalian systems; however,

Figure 11.61 Synthesis, in transgenic tobacco plants, of a polyprotein consisting of a monoclonal antibody heavy chain, a short linker peptide (F2A), and the monoclonal antibody light chain. Following synthesis of the polyprotein, the linker peptide is self-cleaved, leaving separate light and heavy chains which then assemble into whole antibody molecules. The polyprotein segments are not drawn to scale. The disulfide bonds between the subunits are not shown.



this was the first reported use of this peptide in a plant-based system. Expression of the polyprotein in tobacco plants was regulated by the heat stress-inducible glyceraldehyde-3-phosphate dehydrogenase (*GAPC*) gene promoter. Antibody molecules were purified from leaf tissue extracts. These plant-produced fully assembled monoclonal antibodies functioned normally and were synthesized at a relatively high level.

Pharmaceuticals and Vaccines

Given the high cost of many therapeutic compounds that are currently produced in animal or bacterial cells in culture, scientists have begun to produce an increasing number of these compounds in transgenic plants (sometimes called pharming). A number of therapeutic compounds are now synthesized in plants, and some are in the midst of clinical trials and a few have been approved for human use (Table 11.10). It is expected that in the next 10 to 15 years more and more pharmaceutical compounds will be produced in transgenic plants. Currently, tobacco plants are a popular host for the synthesis of pharmaceutical proteins because of the large size of the tobacco leaf from which they are extracted. Rice is another popular plant host because the pharmaceutical can be synthesized in rice kernels, which may be ground into a powder and placed into easy-to-consume capsules.

In addition to producing therapeutic agents, plants can also be engineered to synthesize vaccines. For example, current efforts to develop an HIV vaccine have focused on reproducing the structure of the native viral spike protein gp160 trimer. One recent study found that a soluble HIV envelope protein that was transiently expressed in tobacco plants was an attractive vaccine candidate (Fig. 11.62). In this case, the portion of gp41 that is bound to the viral membrane in the mature virus was deleted, while the soluble portion of the spike protein remained intact. Following its purification from tobacco leaves, this soluble spike protein (gp41) was utilized as a subunit vaccine (see chapter 7) and successfully produced antibodies when tested in rabbits. While the yield of soluble spike protein was lower than expected, researchers are currently fine-tuning this system to obtain higher expression levels.

Table 11.10 Some examples of plant-synthesized pharmaceuticals (as of March 2020)

Product	Disease	Plant	Clinical trial status
HIV microbicide MB66	HIV	Tobacco	Phase I
Zmapp	Ebola virus	Tobacco	Phase II/III
VEN BETA	Gastroenteritis	Rice	Preclinical
VEN120	Inflammatory bowel disease	Rice	Phase II
Moss-aGal	Fabry disease	Moss	Phase I
Moss-FH	C3 Glomerulopathy	Moss	Preclinical
Optibumin	Loss of albumin	Rice	On the market
Vibrio cholerae	Cholera	Potato	Phase I
Heat-labile toxin B of *E. coli*	Diarrhea	Potato	Phase I
Norwalk virus capsid protein	Diarrhea	Potato, tomato	Phase I
ISOkine	Growth factor	Barley	On the market
Taliglucerase alpha	Gaucher disease	Carrot	On the market
Griffithsin, cyanovirin-N	HIV	Rice	Preclinical

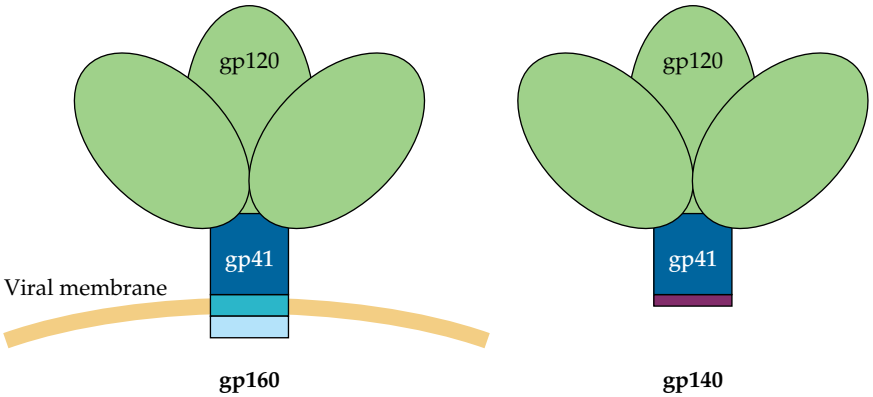


Figure 11.62 The HIV virus spike protein trimer (gp160) embedded in the viral cell membrane and the soluble version of the spike protein (gp140). The major components of these proteins, gp120 and gp41, are indicated.

Several other viral antigens have been successfully synthesized in transgenic plants. These include the spike protein from SARS-CoV-2; the E2 protein from classical swine fever virus; hemagglutinin from H1N1 and H5N1 influenza; rotavirus VP2, VP4, VP6, and VP7; and the B subunit from *E. coli* heat-labile enterotoxin, with several other plant-produced vaccines in very early stages of development.

Poly(3-Hydroxybutyric Acid)

It is costly to produce the polymer poly(3-hydroxybutyric acid), which is used in the synthesis of biodegradable plastics, by bacterial fermentation. Consequently, research has been conducted to determine if the polymer could be produced at a lower cost in plants. In bacteria, such as *Alcaligenes eutrophus*, poly(3-hydroxybutyric acid) is synthesized from acetyl coenzyme A in three steps catalyzed by three enzymes (Fig. 11.63A). Unlike highly valued proteins that are used as therapeutics or specialty chemicals, biological polymers need to be produced at high levels in plants for production to be economically feasible.

Each of the three poly(3-hydroxybutyric acid) biosynthesis genes was fused to a DNA fragment that encodes the chloroplast transit peptide of the small subunit of pea ribulose bisphosphate carboxylase and was placed under the transcriptional control of the *Cauliflower mosaic virus* 35S promoter (Fig. 11.63B). *A. thaliana* plants were transformed separately, with each construct in Ti plasmid binary vectors (Fig. 11.63C). Two resulting transgenic plants, each with a different foreign gene integrated in its genomic DNA, were crossed to form a transgenic plant with two of the poly(3-hydroxybutyric acid) biosynthesis genes (Fig. 11.63C). Then the double-gene-transgenic plant was crossed with a transgenic plant that carried the third biosynthetic gene, and a transgenic plant carrying all three of the bacterial poly(3-hydroxybutyric acid) biosynthesis genes was selected. Mature leaves of some of the triple-gene-transgenic plants produced more than 1 mg of poly(3-hydroxybutyric acid) per gram (fresh weight) of leaf. Unfortunately, *A. thaliana* plants that produced very high levels of poly(3-hydroxybutyric acid) were severely stunted. Nevertheless, this work is a first step in the development of plants that produce large amounts of

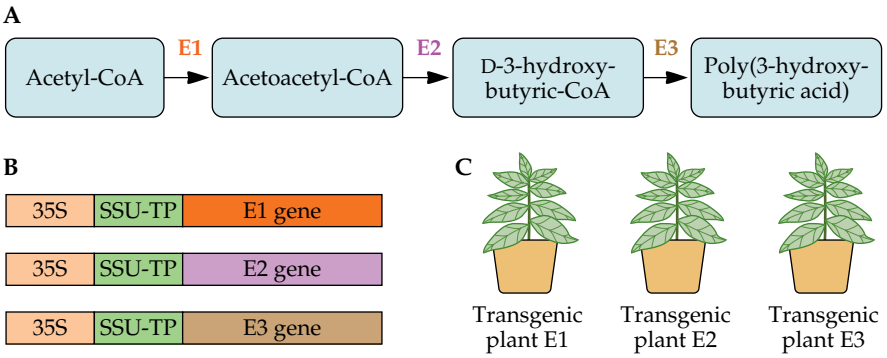


Figure 11.63 **(A)** Synthesis in *Alcaligenes eutrophus* of poly(3-hydroxybutyric acid) from acetyl-CoA. E1 (enzyme 1) is 3-ketothiolase, E2 is acetoacetyl-CoA reductase, and E3 is poly(3-hydroxybutyric acid) synthase. **(B)** The genes for enzymes 1 to 3 are fused to the chloroplast transit peptide of the small subunit of pea ribulose bisphosphate carboxylase so that when the enzyme is synthesized, it will be imported into chloroplasts. The entire construct is placed under the transcriptional control of the *Cauliflower mosaic virus* 35S promoter. **(C)** Separate plants were transformed with each of the constructs shown in panel B. The three transformed plants were crossed with one another until a transgenic plant expressing all three genes was produced.

poly(3-hydroxybutyric acid). However, to realize the commercial potential of this system, these polymers will have to be produced in plants other than *A. thaliana* (where both the plants and the leaves are small) so that there will be a much greater amount of plant biomass produced.

Edible Vaccines

Although considerable progress has been made in recent years in the development of new vaccines, in many countries, either the vaccine itself is too expensive to be used on a large scale or there is a lack of physical infrastructure (e.g., roads and refrigeration) that makes it impossible to disseminate the vaccine. Commercial vaccines are expensive to produce and package and require trained personnel to administer injections. Clearly, it would be advantageous if vaccines could be delivered inexpensively on a broad scale in an edible form, e.g., as part of a fruit or vegetable. When some vaccines are taken orally, they can directly stimulate the immune system (Fig. 11.64). An edible vaccine, in contrast to traditional vaccines, would not require elaborate production facilities, purification, sterilization, packaging, or specialized delivery systems. Moreover, unlike many currently utilized recombinant protein expression systems, plants glycosylate proteins, a factor that may contribute to the immunogenicity and stability of a target protein.

For much of the work on edible vaccines that has been reported so far, the edible vaccines are expressed in potatoes. Potatoes were originally chosen for this work because they were easy to genetically manipulate. However, potatoes were never intended to be the vaccine delivery plant; they require cooking to make them palatable, and cooking destroys (inactivates) most protein antigens. Other plants that are being considered for the delivery of edible vaccines include bananas (although banana trees require several years to mature), tomatoes (although tomatoes spoil readily), lettuce, carrots, peanuts, corn (mainly for "vaccinating" animals), and rice. A number

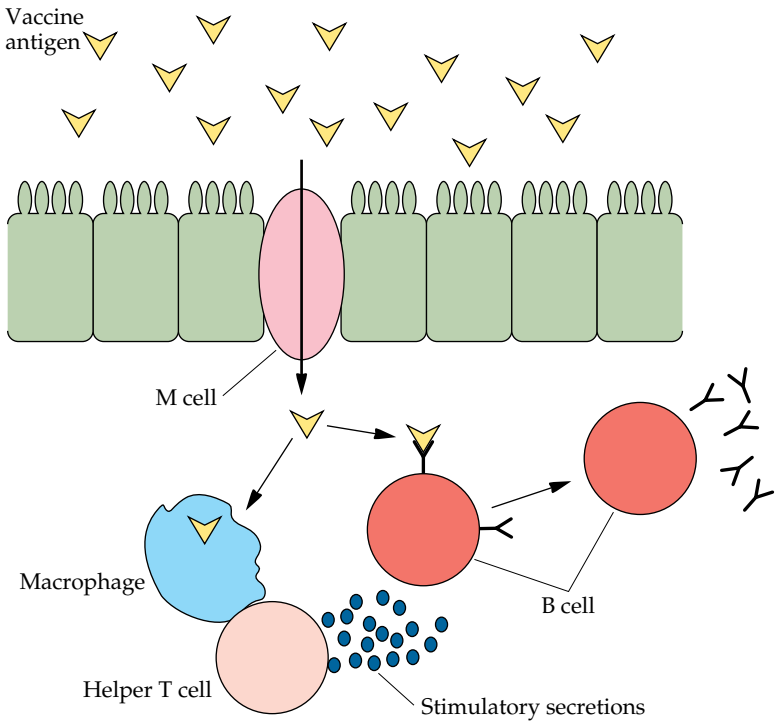


Figure 11.64 An edible vaccine generates an immune response against an antigen from an infectious agent. The ingested antigen, which is expressed by a plant, binds to and is taken up by M cells present in the lining of the intestinal epithelium and is then passed to other cells in the immune system, including macrophages and B cells. The macrophages display portions of the antigen to the helper T cells, which in turn respond by secreting small molecules that activate B cells to synthesize and release antibodies that can neutralize the antigen.

of edible vaccines are currently being developed (Table 11.11), and preliminary data have been reported for vaccines against cholera, HIV, human papilloma virus, tetanus, rotavirus, *Mycobacterium leprae*, *Plasmodium falciparum*, and *Leishmania infantum*. In addition to developing edible vaccines per se, some researchers report positive outcomes in preclinical trials when using edible vaccines as booster vaccines.

Edible Cholera Vaccines

Cholera is an infectious diarrheal disease caused by the enterotoxin produced by the Gram-negative bacterium *Vibrio cholerae* (see chapter 7). Globally, there are more than 5 million cases and 200,000 deaths from cholera each year. While the cholera toxin subunit A1 contains the toxin activity, subunit B, which is connected to subunit A by the A2 peptide, binds to an intestinal receptor. A vaccine comprising subunit B, without subunit A1, would not be toxic and would elicit antibodies against subunit B. On subsequent infection with *V. cholerae*, the antibodies will block the interaction between the cholera toxin and the host cell receptor, thereby preventing the toxin from entering human cells and disrupting ion regulation. To test whether it might be possible to develop an edible vaccine against *V. cholerae*, potato plants were transformed, using *A. tumefaciens*,

Table 11.11 Examples of some recently developed plant-based oral vaccines and oral booster vaccines

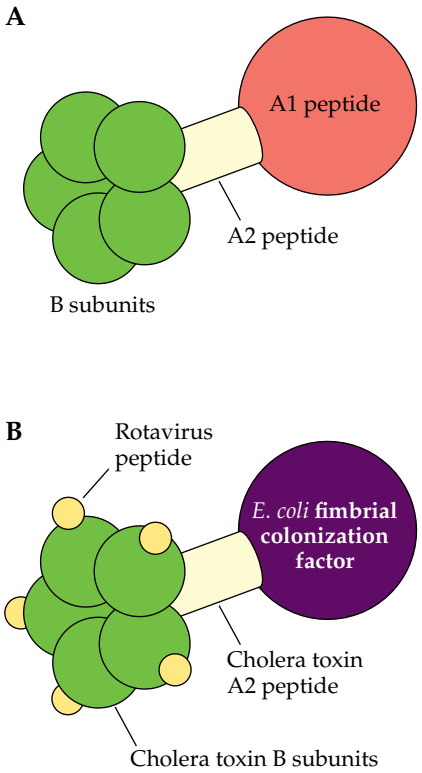
Disease	Target antigen	Plant
Polio	Poliovirus VP1 antigen	Tobacco
Hemophilia A	Coagulation factor VIII, heavy-chain, and C2	Tobacco
Hemophilia A	Factor IX	Tobacco
Pompe disease	α-Glucosidase	Tobacco
Hemophilia B	Factor IX	Lettuce
Enterohemorrhagic *E. coli*	Enterohemorrhagic *E. coli*	Tobacco
Japanese cedar pollen allergy	Cry j 1 and Cry j 2	Rice
Hepatitis B	Hepatitis B surface antigen	Corn, algae, lettuce
Avian influenza	Hemagglutinin H5	*Arabidopsis*
Anthrax	Anthrax protective agent	Indian mustard
HIV	HIV p24 antigen	Wild carrot
Dengue fever	Dengue viral protein	Tobacco
Measles	M-antigen	Lettuce
Type 1 diabetes	Glutamic acid decarboxylase	Red beet
Human papillomavirus	E7 protein	Algae
Diabetes	Glutamic acid decarboxylase	Algae
Infectious bursitis virus	VP2 protein	Quinoa
Cholera	*Vibrio cholerae* toxin B subunit	Rice

with the cholera toxin subunit B gene (Fig. 11.65A). One gram of transgenic potato produced approximately 30 mg of subunit B. After the transgenic potatoes were cooked in boiling water until they were soft enough to be edible by humans, approximately 50% of the subunit B protein remained undenatured. The cooked potatoes were fed to mice once a week for 4 weeks before the mice were tested for antibodies against the subunit B protein and for resistance to *V. cholerae* infection. These tests indicated that the mice had acquired a significant level of protection against *V. cholerae*. Moreover, although mucosal antibody titers declined gradually after the last immunization, they were rapidly restored after an oral boost (an additional feeding) of transgenic potato.

To develop an edible vaccine that protects against other pathogens, the cholera toxin subunit B and A2 genes were each fused to additional antigen genes and then used to generate transgenic potato plants. Two fusion proteins were created: a 22-amino-acid epitope from murine (mouse) rotavirus enterotoxin NSP4 was fused to the C-terminal end of the cholera toxin subunit B protein, and the enterotoxigenic *E. coli* fimbrial colonization factor CFA/I was fused to the N-terminal end of the cholera toxin subunit A2 protein (Fig. 11.65B). Transgenic potatoes that expressed both fusion proteins were fed to mice, which generated antibodies against cholera toxin subunit B protein, murine rotavirus enterotoxin NSP4, and *E. coli* fimbrial colonization factor CFA/I and were protected against rotavirus-caused diarrhea. This approach holds great promise for the development of inexpensive and readily available vaccines for a wide range of diseases.

More recently, the cholera toxin subunit B protein was expressed in banana and rice. To express the cholera toxin B subunit in rice plants, first the B subunit gene was completely resynthesized using codons that are optimized for expression in plants. This gene was introduced into rice plants using a binary vector of the Ti plasmid of *A. tumefaciens* (Fig. 11.66).

Figure 11.65 Cholera toxin **(A)** and an engineered cholera vaccine **(B)**. In the engineered vaccine, the cholera toxin B subunit is synthesized as a fusion protein with the rotavirus peptide, and the cholera toxin A2 subunit is fused to the enterotoxigenic *E. coli* fimbrial colonization factor. The cholera toxin B and A2 peptides are noncovalently attached to one another. Both fusion proteins are expressed in transgenic potatoes and combine spontaneously to form a cholera toxin-like protein, as shown.



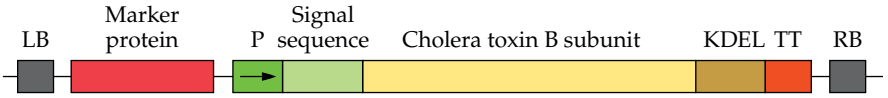


Figure 11.66 The construct used to express the cholera toxin B subunit in rice seeds. LB and RB indicate the left and right borders of the T-DNA construct, respectively; the promoter and transcription termination region of the marker gene are not shown; for the expression of the cholera toxin B subunit, the promoter (P), the signal sequence, and the transcription termination region (TT) are all from the rice seed storage protein gene glutelin *GluB-1*, and the KDEL region is a signal that retains the protein on the endoplasmic reticulum. The arrow indicates the direction of transcription.

The construct included the codon-optimized cholera toxin B subunit gene under the control of the rice seed storage glutelin 2.3-kb *GluB-1* promoter and signal sequence to facilitate secretion. In addition, a DNA sequence encoding the KDEL signal was located at the 3′ end of the cholera toxin B subunit gene. The KDEL sequence (lysine, aspartic acid, glutamic acid, leucine) is expressed as part of the protein and targets the cholera toxin subunit to the endoplasmic reticulum, where the protein is glycosylated. The KDEL sequence is also thought to increase the expression level of some proteins in plants. The glutelin 2.3-kb *GluB-1* transcription termination sequence is also included at the 3′ end of the cholera toxin B subunit gene. When this construct was introduced into two rice varieties, expression levels as high as 30 μg per seed (or ~2.1% of the seed protein) were observed. Mice that were fed a suspension of ground rice powder in water were protected against cholera after ingesting approximately 50 mg of rice powder containing around 75 μg of cholera toxin B subunit protein. It is expected that this type of edible vaccine may become the prototype for a large number of similar edible vaccines. However, it is first necessary for this vaccine to undergo successful clinical trials in humans.

Edible *E. coli* Vaccines

It has been estimated that Shiga toxin-producing strains of *E. coli* cause approximately 100,000 cases of hemorrhagic colitis a year. About 6% of those infections produce severe complications, including kidney failure. Similar to cholera toxin, the Shiga toxin contains one A subunit, which encodes the catalytic, or toxin, activity, per five B subunits, which act (together) to bind to cellular surface receptors. To develop an oral vaccine against type 2 Shiga toxin (type 2 is responsible for the most severe disease in humans), the genes for a genetically inactivated (attenuated) version of the Shiga toxin A and B peptides were both cloned and expressed in tobacco cells (Fig. 11.67). Mice were fed leaves from the transgenic tobacco plants expressing the inactivated Shiga toxin once a week for 4 weeks and then challenged with Shiga toxin-producing *E. coli*; other, control, mice were not fed the transgenic tobacco and were subsequently challenged. One week after the introduction of the toxic *E. coli* strain, all of the mice that were not fed the antigen-producing tobacco had died. In contrast, 2 weeks after treatment with the toxic *E. coli* strain, all of the orally vaccinated mice were still alive. This experiment serves as a proof of the concept that oral administration of the inactivated Shiga toxin is a highly effective means of

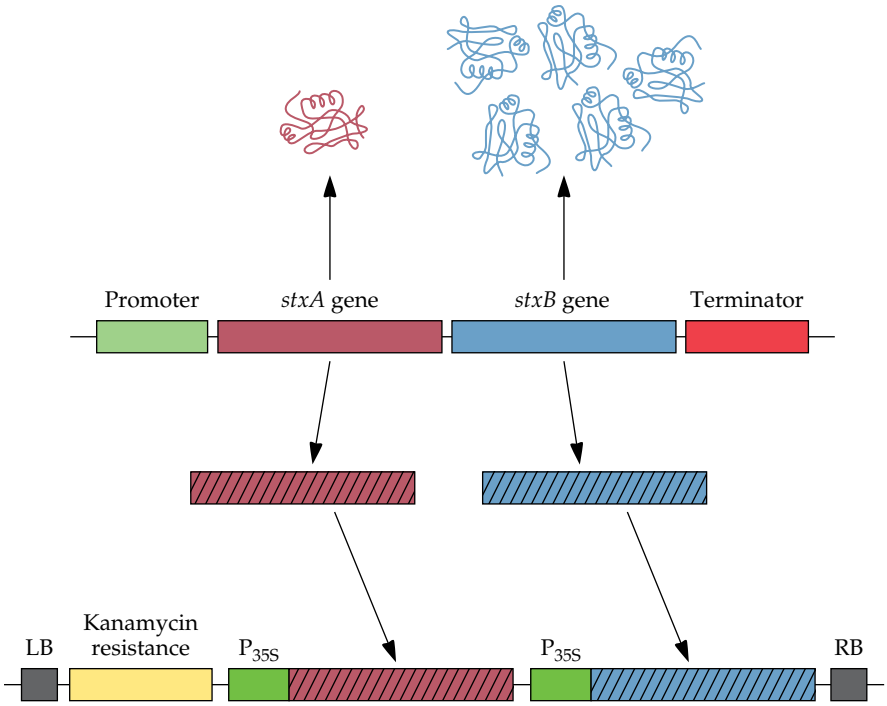
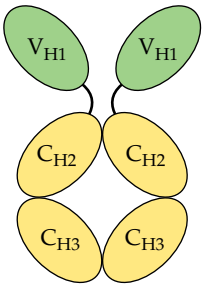


Figure 11.67 Shiga toxin A and B subunits encoded in a bacterial operon (*stxA* and *stxB* genes, respectively) under the control of a single promoter and transcription terminator. These two genes were subsequently isolated and modified, first to inactivate the A subunit and then to eliminate sequences on both genes that might adversely affect their transcription in plant cells. The modified genes were inserted between the left and right borders (LB and RB) of the T-DNA of a binary Ti plasmid-based vector, each under the control of a separate *Cauliflower mosaic virus* 35S promoter. The T-DNA also contained a gene encoding kanamycin resistance under the control of its own promoter. The transcription terminator sequences of the three genes contained within the T-DNA are not shown for the sake of simplicity. The final construct was used to transform tobacco plants.

protecting animals against Shiga toxin-producing *E. coli*. Of course, for a human oral vaccine, a more suitable host plant would be desirable.

More recently, researchers expressed antibodies that included a variable domain, V_{H1}, from llama heavy-chain antibodies against enterotoxigenic *E. coli* fused to the Fc portion of a porcine immunoglobulin IgA molecule, i.e., C_{H2} and C_{H3} regions, with the fusion protein expressed and produced in *A. thaliana* seeds (Fig. 11.68). A llama heavy-chain antibody fragment was used because llama antibodies are considered to be highly stable molecules. When piglets were fed seeds that expressed this chimeric immunoglobulin molecule, they were subsequently protected against gastrointestinal infection caused by enterotoxigenic *E. coli*. This method of providing piglets with passive immunity is highly effective in preventing piglet postweaning diarrhea. However, since *A. thaliana* seeds are very small, to develop this approach so that it is commercially feasible for the oral passive immunization of piglets, the genetic construct encoding the chimeric antibody molecule would need to be transformed into a food crop such as soybean or pea and then tested in pig farms. Following large-scale

Figure 11.68 A hybrid antibody molecule that consists entirely of heavy chains. The variable domain V_{H1} (shown in green) is from llama heavy-chain antibodies against anti-enterotoxigenic *E. coli*. The Fc portion of the molecule (C_{H2} and C_{H3}, shown in yellow) is from a porcine immunoglobulin IgA molecule.



tests, if this approach is successful in limiting bacterial shedding (which causes the disease to spread), the antibody-containing plant would decrease the infection pressure within the herd and reduce the farmer's subsequent dependency on the use of antibiotics.

Plant Yield

In nearly every instance where plants are utilized, it is desirable to optimize the yield of the plant or its components. Theoretically, increases in plant yield may be achieved in several ways. For example, preliminary studies directed toward introducing the C_4 pathway of photosynthesis in plants that normally use the C_3 pathway have been reported; plants have been engineered to take up iron from the soil more efficiently; the supply of oxygen to plant cells has been enhanced, thereby increasing the rate of plant growth; and the lignin content of some trees has been decreased, making it easier to isolate the cellulose. The work of modifying plant yields is still in its infancy. Nevertheless, based on what has been achieved to date, several strategies are likely to be fruitful.

Increasing Grain Yield

With increasing world populations, not only is it necessary to decrease the inhibition of crop productivity by insects, viruses, weeds, and fungal and bacterial pathogens, but it is essential that the world increase the amount of food that is produced. Corn is one of the world's major food and feed crops, and therefore, significant increases in the amount of corn that is produced will go a long way toward feeding the world in the future. Recently, a group of scientists reported increasing corn yields by overexpressing the corn MADS-box (a conserved regulatory DNA motif of 168 to 180 bp) transcription factor gene, *zmm28*. The 142 MADS-box genes present in the corn genome play a wide range of roles in plant growth and development, making them a target for improving plant yield. Increasing the expression of *zmm28* results in increased binding of the encoded protein to the promoter regions of several key genes to increase expression of proteins involved in nitrogen assimilation, photosynthesis, auxin reception, and gibberellin reception (Fig. 11.69). This results in a significantly increased yield of corn compared to nontransformed corn as evidenced in large-scale field trials. Not surprisingly, these researchers have applied for patent protection for this important invention.

Increasing Harvest Index

During the Green Revolution that occurred between 1950 and the late 1960s, agricultural productivity increased significantly as a consequence of the adaptation of a number of technologies, including chemical fertilizers, agrochemicals, mechanization, and newer methods of cultivation. During this time major increases occurred in the harvest index (the fruit or grain yield divided by total plant biomass) of many crop plants, especially rice and wheat, that resulted from the development of dwarf varieties with decreased leaf biomass and increased grain yields. These dwarf varieties resulted from mutations of genes involved in gibberellin biosynthesis and/or signaling. The mutant lines were then crossed with high-yielding

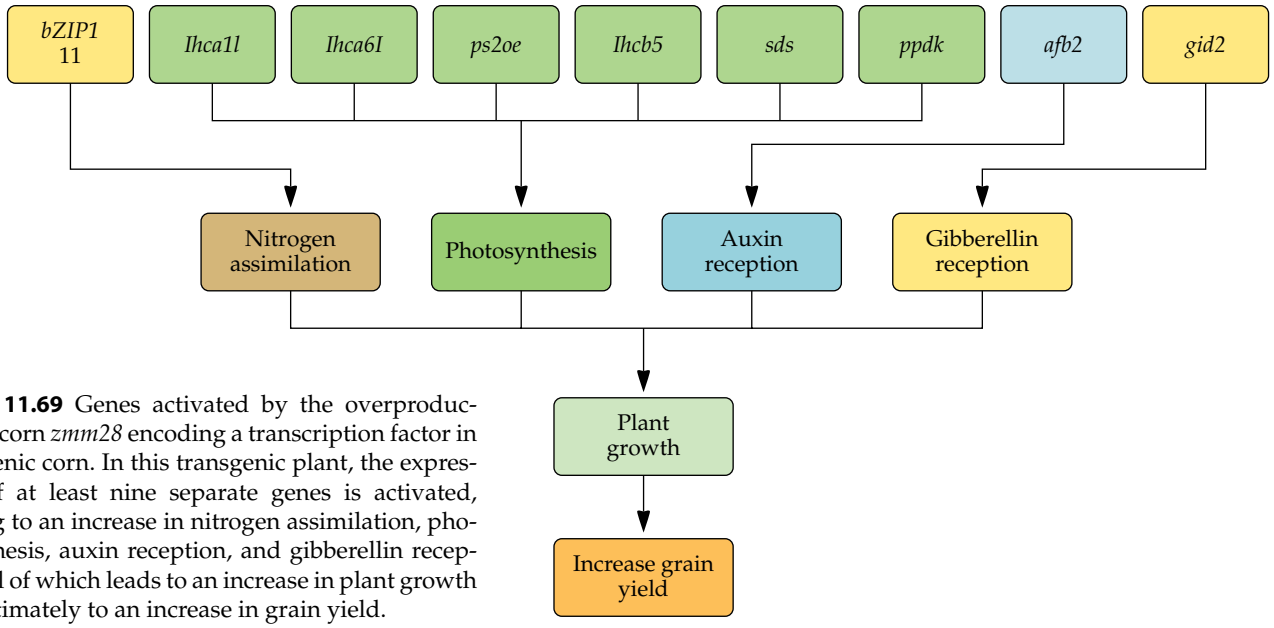


Figure 11.69 Genes activated by the overproduction of corn *zmm28* encoding a transcription factor in transgenic corn. In this transgenic plant, the expression of at least nine separate genes is activated, leading to an increase in nitrogen assimilation, photosynthesis, auxin reception, and gibberellin reception, all of which leads to an increase in plant growth and ultimately to an increase in grain yield.

varieties, leading to an even greater proportion of photoassimilates partitioned into grain. Unfortunately, efforts to determine the precise genetic basis of increases in grain yield were not successful.

However, recently, one group of researchers transformed tomato plants with a gene encoding a cyanobacterial flavodoxin, an electron shuttle protein, directed to plant chloroplasts. Expression of the flavodoxin resulted in a significant decrease in the size of plant stems, leaves, and fruit, all as a consequence of decreased cell size. On the other hand, in the transgenic plants, tomato fruit had a higher pigment content, and there were greater numbers of fruit per plant. The overall yield of tomato fruit could be increased significantly by planting the smaller transgenic tomato plants at a higher density. Importantly, when the transgenic tomatoes were compared to wild-type tomatoes, the levels of sugars, organic acids, and amino acids were similar. Should this preliminary result be verified in the field and in different tomato varieties, this simple genetic manipulation should allow for an increased yield of tomatoes in a small space.

Decreasing Lignin Content

After cellulose, lignin is the second most abundant organic compound on Earth. Depending on the species of tree, it accounts for approximately 15 to 35% of the dry weight of wood. While lignin is important in the mechanical support of trees and in their defense against pathogens, it is a major impediment in obtaining the cellulose that is needed by the pulp and paper industry—releasing cellulose requires harsh physical and chemical treatment. The global market for wood pulp was approximately $240 billion in 2020 and is expected to reach $290 billion by 2027. This process, which removes ~150 million tons of lignin from wood, consumes enormous amounts of energy and harsh chemicals. In addition, high levels of lignin decrease the nutritional value of forage crops that are used for cattle feed.

Lignin is synthesized by the oxidative polymerization of one of three hydroxycinnamyl alcohols: *p*-coumaryl, coniferyl, or sinapyl alcohol. In one approach to reduce the level of lignin produced by quaking aspen trees (*Populus tremuloides*) and to make it easier to harvest the cellulose, one of the steps in the biosynthesis of lignin was altered. The enzyme 4-coumarate:coenzyme A ligase catalyzes the conversion of 4-coumarate to 4-coumarate:coenzyme A. The latter compound is a precursor of both flavonoids and lignin (Fig. 11.70). There are two isoforms of the enzyme 4-coumarate:coenzyme A ligase. One isoform is expressed in the epidermis of stems and leaves and is probably involved in flavonoid biosynthesis. The other isoform is expressed in differentiating xylem and presumably functions mainly in lignin biosynthesis. The gene for the isoform expressed in xylem was isolated, and an antisense version of the gene was placed under the transcriptional control of the 35S promoter and introduced into aspen by *A. tumefaciens*-mediated transformation. Transformed aspen trees exhibited up to a 45% decrease in lignin content and, at the same time, as much as a 15% increase in cellulose. In addition, the transgenic trees were about 25% larger and had thicker stems and larger leaves than the nontransgenic trees. The altered lignin and cellulose content of the transgenic aspen should make it easier to extract cellulose during pulp and paper manufacturing.

In the past 10 to 15 years there have been numerous reports of scientists genetically engineering forest trees to reduce their content of lignin. This has been achieved by various means, including the use of RNAi or antisense RNA to downregulate the expression of key genes, using CRISPR/Cas to inactivate those genes involved in lignin production, and overexpressing some transcriptional repressors. Some of these efforts are summarized in Table 11.12. All of these modifications led to a decreased synthesis of lignin and an increased synthesis of cellulose.

Figure 11.70 Role of 4-coumarate:coenzyme A (CoA) ligase in lignin biosynthesis.

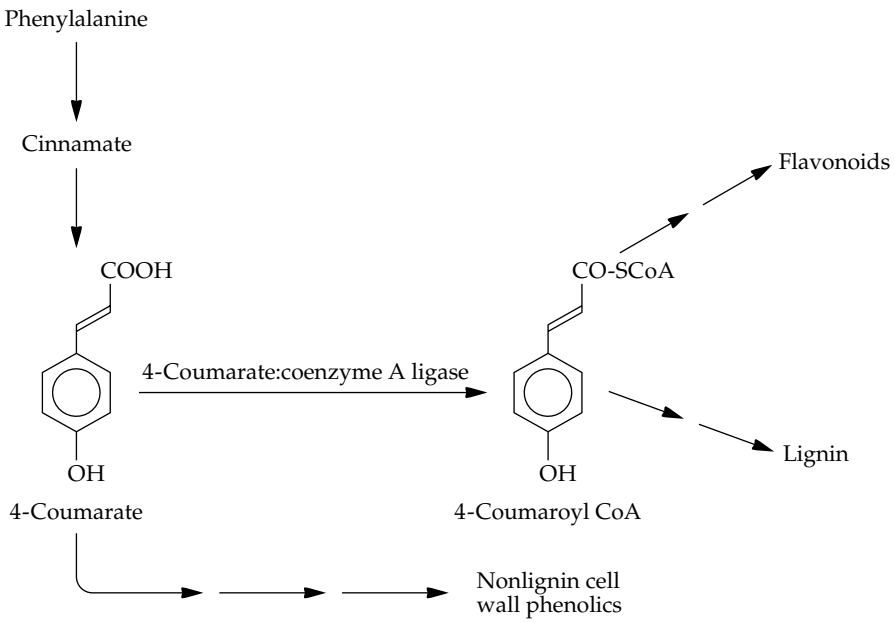


Table 11.12 Some of the trees whose level of lignin has been genetically modified[a]

Species	Gene	Method	Decrease in lignin (%)
Pinus radiata	*HCT*	RNAi	42
Pinus radiata	*CCoAOMT*	RNAi	20
Pinus radiata	*CCR*	RNAi	46
Leucaena leucocephala	*COMT*	Antisense	30
Pinus radiata	*4CL*	RNAi	36
Pinus taeda	*4CL*	RNAi	33
Pinus taeda	*CAD*	Mutant	9
Picea abies	*CCR*	Antisense	8
Eucalyptus urophilla x E. grandis	*C3'H*	Antisense	27
Eucalyptus urophilla x E. grandis	*C4H*	Antisense	36
Populus tremula x Populus tremuloides	*C4H*	Antisense	33
P. tremuloides Michx.	*4CL*	Antisense	45
P. tremuloides	*4CL*	Antisense	40
Paulownia tomentosa	*4CL*	Antisense	42
Populus trichocarpa	*4CL*	Antisense	30
Phyllostachys nigra L. x Prunus maximowiczii	*4CL*	Antisense	55
P. tremula x Populus alba	*4CL1*	CRISPR/Cas	23
P. nigra L. x Prunus maximowiczii	*4CL*	Antisense	47

[a]Abbreviations: HCT, p-hydroxycinnamoyl-CoA:quinate/shikimate-p-hydroxycinnamoyltransferase; CCoAOMT, caffeoyl-CoA O-methyltransferase; CCR, cinnamoyl-CoA reductase; COMT, caffeic acid O-methyltransferase; 4CL, 4-coumarate:CoA ligase; CAD, cinnamyl alcohol dehydrogenase; CCR, cinnamoyl-CoA reductase; C3′H, p-coumaroyl quinate/shikimate 3′-hydroxylase; C4H, cinnamate 4-hydroxylase; 4CL, 4-coumarate:CoA ligase.

A considerable amount of effort has been directed toward finding ways to efficiently utilize plants to produce renewable biofuels. Some of this research has focused on converting materials such as corn stover, wheat straw, grasses, and wood by-products to glucose and then to ethanol. With all of these compounds, it is necessary to pretreat the substrate under harsh conditions to remove the lignin and make the cellulose and hemicellulose more accessible to hydrolytic enzymes. These pretreatment steps are generally costly in terms of materials and energy, and often the harsh conditions employed make it much more difficult to enzymatically digest the cellulose. However, by reducing the lignin content of plant tissues, it may be possible to employ mild pretreatment prior to enzymatic digestion (Fig. 11.71). In fact, by constructing transgenic alfalfa plants that contained an antisense version of the gene for the enzyme skimiate hydroxycinnamoyl transferase, which catalyzes one of the steps in the biosynthesis of lignin, it was possible to reduce the lignin content of the plant by about 50 to 70%. Interestingly, the transgenic plants that had the lowest lignin content also had the highest carbohydrate levels, suggesting that the plant has redirected its resources, compensating for the reduction in lignin. Moreover, when the lignin levels are low, the cell wall carbohydrates are more accessible to cellulose- and hemicellulose-degrading enzymes. Importantly, since the biochemical steps that lead to the synthesis of lignin are highly conserved across the plant kingdom, it may be possible to apply this approach to a variety of plants so that, for example, corn may be engineered to continue to be a food source for both animals and humans, while the parts of the corn plant that are not consumed may be converted into biofuels.

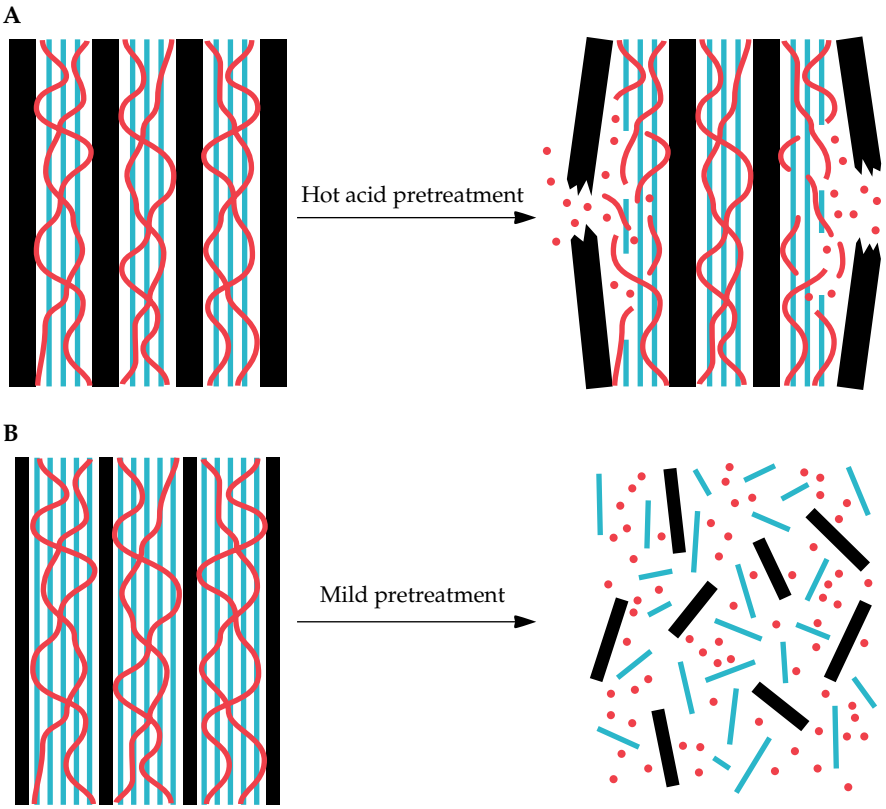


Figure 11.71 Native **(A)** and transgenic **(B)** plant cell walls are treated to partially break down their structure and make the hemicellulose and cellulose components more accessible to enzymatic digestion. Lignin is shown in black, cellulose in blue, and hemicellulose in red.

Researchers have modified a variety of plants, including alfalfa, poplar, switchgrass (*Panicum virgatum* L.), sugarcane, and maize so that their lignin content was decreased, typically from 5 to 25%. In all these instances, one of nine genes involved in lignin biosynthesis was downregulated either by antisense RNA or through the use of RNAi. However, in one instance, the lignin content was decreased by up to 50% in switchgrass by overexpressing the *MYB4* gene. The *MYB4* gene encodes a transcriptional repressor that binds to the promoters of several genes in the monolignol pathway for lignin synthesis (Fig. 11.72). Overexpression of the repressor in transgenic switchgrass results in reduced lignin content. The modified plants had reduced stature, increased tillering (new shoots from the original stalk), and a three-fold increase in sugar release from plant cell wall residues. Plants that have been engineered in this way are likely to become preferred substrates for the production of biofuels.

Decreasing Pectin Content

As described above, one strategy to more easily obtain the sugars in plants and plant residues includes engineering plants with cells walls that are more easily broken down by decreasing the lignin content. However, to release all of the available sugars, hemicellulose, and cellulose, it is also

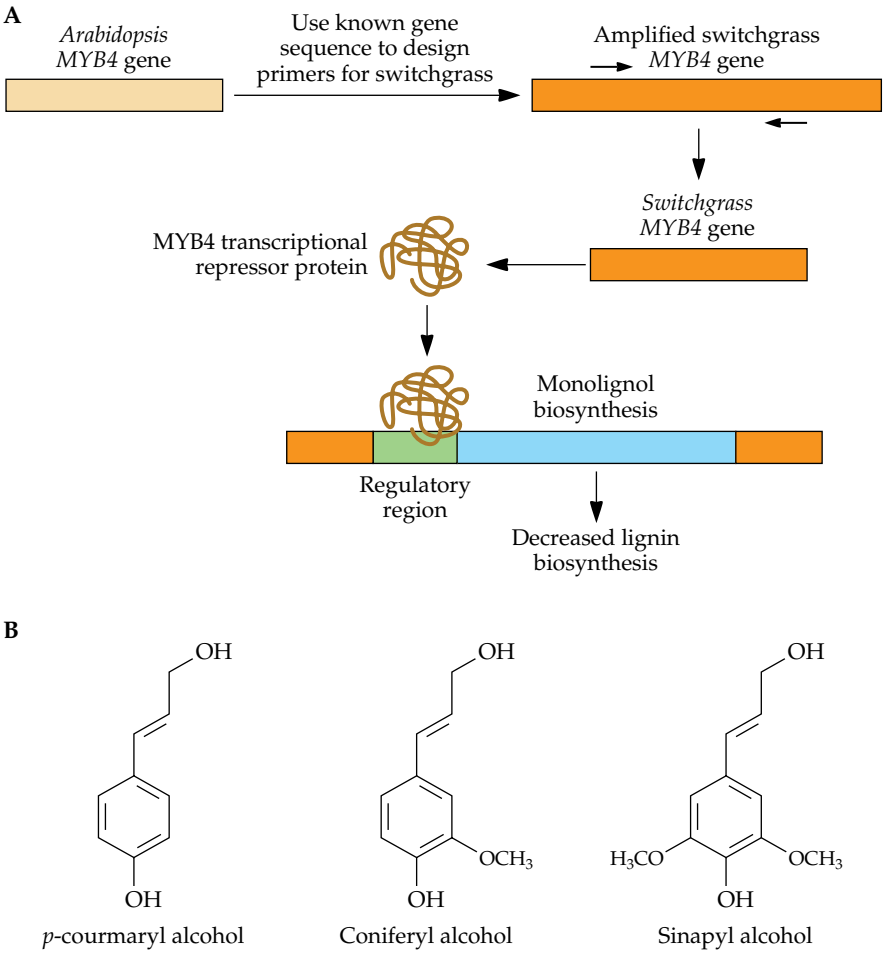


Figure 11.72 Overexpression of the switchgrass *MYB4* gene in order to decrease lignin biosynthesis. **(A)** First, primers based on the known sequence of the *Arabidopsis thaliana MYB4* gene are used to amplify the *MYB4* gene from switchgrass. Second, the *MYB4* gene from switchgrass is used to produce large amounts of the MYB4 transcriptional repressor protein, with the overexpressing construct being expressed in transgenic switchgrass. The excess MYB4 protein binds to the region that regulates the synthesis of monolignols, thereby decreasing their synthesis and hence the synthesis of lignin. **(B)** Structures of the major monolignols, source material for the biosynthesis of lignin.

necessary to remove the pectin. Pectin is a structural acidic heteropolysaccharide whose main component is galacturonic acid and that helps to bind plant cells together (Fig. 11.73A). During fruit ripening, pectin is broken down by the enzymes pectinase and pectinesterase. One group of researchers used RNAi to decrease the expression of the pectin biosynthesis gene *Galacturonosyltransferase 4* (*GAUT4*) in switchgrass, rice, and poplar (Fig. 11.73B). The engineered plants contained an increased amount of available sugar and grew much better than the control plants, producing greater biomass, and the sugars and biomass were much more readily converted to ethanol than in the control plants. Importantly, over a 3-year field trial of engineered switchgrass, the plants retained the engineered RNA-silencing construct. This approach is expected to further increase the

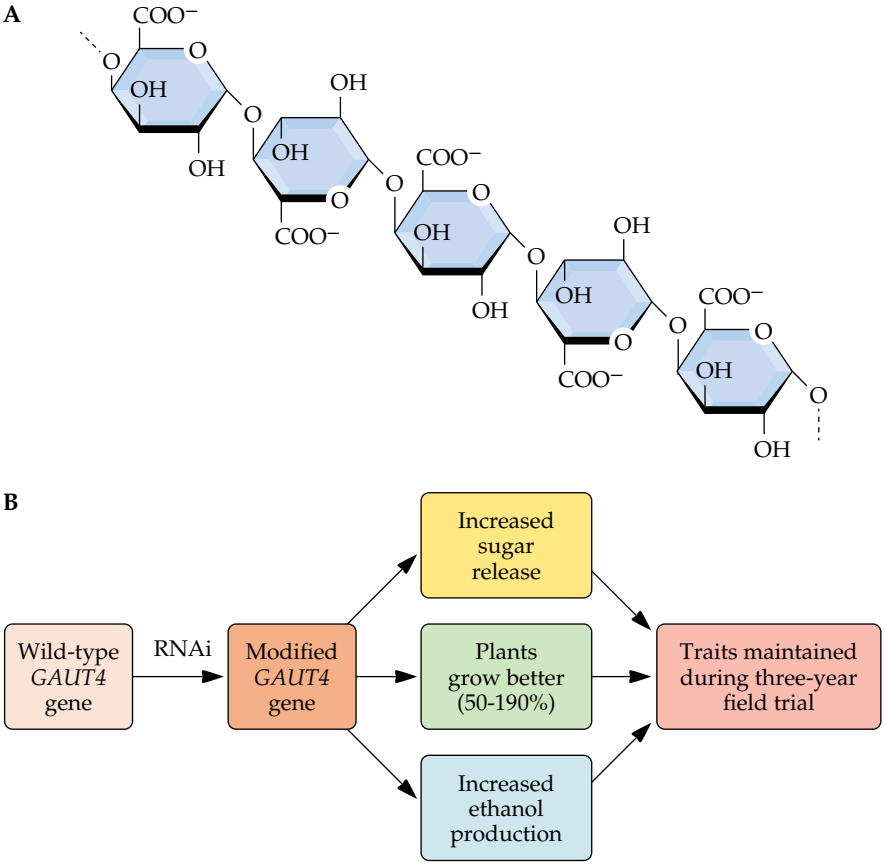


Figure 11.73 **(A)** Structure of a portion of a pectin molecule. **(B)** Flow chart showing RNAi decreasing the expression of the pectin biosynthesis gene GAUT4 and the effect of the decreased expression of pectin on the plant.

productivity of energy crops planted on marginal land as well as facilitate the extraction of biofuels from food crop residues.

Increasing Oxygen Content

Oxygen is an essential substrate for plant respiration, and the amount of available oxygen may limit several plant biochemical reactions. Since it is impractical to increase the external oxygen supply to the plant, one way to increase the oxygen concentration inside a plant is to provide the plant with a protein that can sequester oxygen and provide it where it is needed.

The Gram-negative obligately aerobic bacterium *Vitreoscilla* produces a dimeric hemoglobin that binds oxygen tightly and allows the bacterium to proliferate under oxygen-limited conditions. The gene encoding this hemoglobin has been isolated and expressed in a variety of bacteria, increasing the growth rate, the final cell density, and the yield of cloned foreign genes. When the *Vitreoscilla* hemoglobin gene under the transcriptional control of the 35S promoter from *Cauliflower mosaic virus* was introduced into tobacco by *A. tumefaciens*-mediated gene transfer, transgenic plants produced 80 to 100% more dry weight than nontransgenic plants. In addition, seed

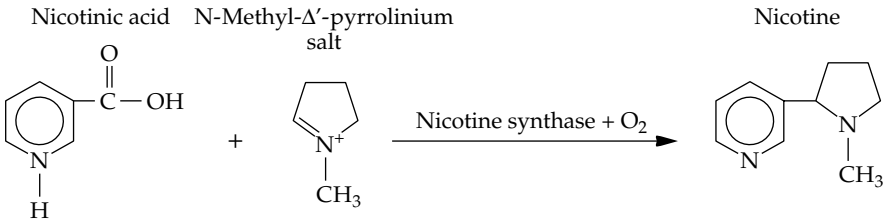


Figure 11.74 The last step in the biosynthesis of nicotine.

germination time was reduced from 6 to 8 days to 3 to 4 days, and transgenic plants contained approximately 35% more chlorophyll and 34% more nicotine than nontransgenic plants. It is thought that the *Vitreoscilla* hemoglobin increases the availability of oxygen and/or energy in the cell. For example, the synthesis of 1 mol of chlorophyll from glutamate requires 4 mol of oxygen and 2 mol of ATP. Similarly, the last step in the synthesis of nicotine, the conversion of nicotinic acid and *N*-methyl-Δ′-pyrrolinium salt to nicotine, is catalyzed by the oxygen-dependent enzyme nicotine synthase (Fig. 11.74). While this sort of genetic manipulation raises more questions than it answers, the possibility of increasing plant yield by manipulating the oxygen concentration is quite exciting.

summary

By using a variety of techniques (see chapter 10), it has become relatively straightforward to transform plants with foreign genes. Plants have been engineered to be resistant to a range of environmental stresses, including insects, viruses, herbicides, pathogens, and oxidative and salt stress.

Several different strategies have been used to confer resistance against insect predators, including introducing a gene encoding an insecticidal protoxin from one of several subspecies of *B. thuringiensis*; plant proteins, such as α-amylase inhibitors, lectins, or protease inhibitors; or other bacterial insecticidal proteins.

Transgenic plants expressing the gene for a viral coat protein are protected against infection by that virus; this occurs through the production of RNAi. They may also be protected against damage from infective viruses by expression of single-chain antibodies, micro-RNAs, and CRISPR sgRNA and Cas9 nuclease.

To permit crop plants to proliferate in the presence of weeds, many plants have been engineered to be resistant to one or more "environmentally friendly" herbicides. This approach has become enormously successful and is the basis for the largest number of transgenic plants that are used in the field.

Different foreign proteins protect plants against different stresses. In addition, compartmentalization of sodium into the vacuole is effective against salt stress, and lowering plant ethylene levels has an impact on many different types of stress.

The expression of foreign genes in plants makes it possible to produce a wide range of new plant varieties. The nutritional content of crops has been enhanced; discoloration of potatoes has been prevented by genetic manipulation; plants have been developed to act as factories for the large-scale production of important foreign proteins, such as antibodies and therapeutics; and plant yield has been increased by increasing the iron and oxygen content and by modulating the lignin content.

Plant nutritional content may be improved in a variety of ways. The amino acid content (specifically, methionine and lysine) can be increased, the lipid composition can be modified to suit the intended end use of the oil, pathways for the synthesis of the precursor to vitamin A have been engineered, and plants with increased levels of available iron have been created.

As bioreactors, transgenic plants can produce, on a laboratory scale, monoclonal antibodies and antibody fragments, the polymer polyhydroxybutyrate, which can be used to make a biodegradable plastic-like material, and a number of potential therapeutic agents. Finally, transgenic plants have also been used as edible vaccines, an approach that could result in a wide range of new and inexpensive vaccines.

REFERENCES

Aida R, **Yoshida T**, **Ichimura K**, **Goto R**, **Shibata M.** 1998. Extension of flower longevity in transgenic torenia plants incorporating ACC oxidase transgene. *Plant Sci* **138:**91–101.

Kumar PA, **Sharma RP**, **Malik VS.** 1996. The insecticidal proteins of *Bacillus thuringiensis*. *Adv Appl Microbiol* **42:**1–43.

Anderson EJ, **Stark DM**, **Nelson RS**, **Tumer NE**, **Beachy RN.** 1989. Transgenic plants that express the coat protein gene of TMV or A1MV interfere with disease development of non-related viruses. *Phytopathology* **12:**1284–1290.

Arakawa T, **Chong DKX**, **Langridge WHR.** 1998. Efficacy of a food plant-based oral cholera toxin B subunit vaccine. *Nat Biotechnol* **16:**292–297.

Aronson AI, **Shai Y.** 2001. Why *Bacillus thuringiensis* insecticidal toxins are so effective: unique features of their mode of action. *FEMS Microbiol Lett* **195:**1–8.

Bachem CWB, **Speckmann G-J**, **van der Linde PCG**, **Verheggen FTM**, **Hunt MD**, **Steffens JC**, **Zabeau M.** 1994. Antisense expression of polyphenol oxidase genes inhibits enzymatic browning in potato tubers. *Bio/Technology* **12:**1101–1105.

Badran AH, **Guzov VM**, **Huai Q**, **Kemp MM**, **Vishwanath P**, **Kain W**, **Nance AM**, **Evdokimov A**, **Moshiri F**, **Turner KH**, **Wang P**, **Malvar T**, **Liu DR.** 2016. Continuous evolution of *Bacillus thuringiensis* toxins overcomes insect resistance. *Nature* **533:**58–63.

Bates SL, **Zhao J-Z**, **Roush RT**, **Shelton AM.** 2005. Insect resistance management in GM crops: past, present and future. *Nat Biotechnol* **23:**57–62.

Baum JA, **Bogaert T**, **Clinton W**, **Heck GR**, **Feldmann P**, **Ilagan O**, **Johnson S**, **Plaetinck G**, **Munyikwa T**, **Pleau M**, **Vaughn T**, **Roberts J.** 2007. Control of coleopteran insect pests through RNA interference. *Nat Biotechnol* **25:**1322–1326.

Beachy RN, **Loesch-Fries S**, **Tumer NE.** 1990. Coat protein-mediated resistance against virus infection. *Annu Rev Phytopathol* **28:**451–474.

Beaujean A, **Ducrocq-Assaf C**, **Sangwan RS**, **Lilius G**, **Bülow L**, **Sangwan-Norreel BS.** 2000. Engineering direct fructose production in processed potato tubers by expressing a bifunctional α-amylase/glucose isomerase gene complex. *Biotechnol Bioeng* **70:**9–16.

Behrens MR, **Mutlu N**, **Chakraborty S**, **Dumitru R**, **Jiang WZ**, **Lavallee BJ**, **Herman PL**, **Clemente TE**, **Weeks DP.** 2007. Dicamba resistance: enlarging and preserving biotechnology-based weed management strategies. *Science* **316:**1185–1188.

Biswal AK, **Atmodjo MA**, **Li M**, **Baxter HL**, **Yoo CG**, **Pu Y**, **Lee Y-C**, **Mazarei M**, **Black IM**, **Zhang J-Y**, **Ramanna H**, **Bray AL**, **King ZR**, **LaFayette PR**, **Pattathil S**, **Donohoe BS**, **Mohanty SS**, **Ryno D**, **Yee K**, **Thompson OA**, **Rodriguez M Jr**, **Dumitrache A**, **Natzke J**, **Winkeler K**, **Collins C**, **Yang X**, **Tan L**, **Sykes RW**, **Gjersing EL**, **Ziebell A**, **Turner GB**, **Decker SR**, **Hahn MG**, **Davison BH**, **Udvardi MK**, **Mielenz JR**, **Davis MF**, **Nelson RS**, **Parrott WA**, **Ragauskas AJ**, **Neal Stewart C Jr**, **Mohnen D.** 2018. Sugar release and growth of biofuel crops are improved by downregulation of pectin biosynthesis. *Nat Biotechnol* **36:**249–257.

Bonning BC, **Pal N**, **Liu S**, **Wang Z**, **Sivakumar S**, **Dixon PM**, **King GF**, **Miller WA.** 2014. Toxin delivery by the coat protein of an aphid-vectored plant virus provides plant resistance to aphids. *Nat Biotechnol* **32:**102–105.

Boonrod K, **Galetzka D**, **Nagy PD**, **Conrad U**, **Krczal G.** 2004. Single-chain antibodies against a plant viral RNA-dependent RNA polymerase confer virus resistance. *Nat Biotechnol* **22:**856–862.

Boulter D, **Gatehouse AMR**, **Hilder V.** 1989. Use of cowpea trypsin inhibitor (CpTI) to protect plants against insect predation. *Biotechnol Adv* **7:**489–497.

Bravo A, **Gill SS**, **Soberón M.** 2007. Mode of action of *Bacillus thuringiensis* Cry and Cyt toxins and their potential for insect control. *Toxicon* **49:**423–435.

Buhari ML, **Sulaiman BR**, **Vyas NL**, **Badaru S**, **Harisu UY.** 2016. Role of biotechnology in phytoremediation. *J Bioremed Biodegrad* **7:**300.

Cao H, **Li X**, **Dong X.** 1998. Generation of broad-spectrum disease resistance by overexpression of an essential regulatory gene in systemic acquired resistance. *Proc Natl Acad Sci USA* **95:**6531–6536.

Castle LA, **Siehl DL**, **Gorton R**, **Patten PA**, **Chen YH**, **Bertain S**, **Cho H-J**, **Duck N**, **Wong J**, **Liu D**, **Lassner MW.** 2004. Discovery and directed evolution of a glyphosate tolerance gene. *Science* **304:**1151–1154.

Cerda H, **Paoletti MG.** 2004. Genetic engineering with *Bacillus thuringiensis* and conventional approaches for insect resistance in crops. *Crit Rev Plant Sci* **23:**317–323.

Chandrasekaran J, **Brumin M**, **Wolf D**, **Leibman D**, **Klap C**, **Pearlsman M**, **Sherman A**, **Arazi T**, **Gal-On A.** 2016. Development of broad virus resistance in non-transgenic cucumber using CRISPR/Cas9 technology. *Mol Plant Pathol* **17:**1140–1153.

Chanoca A, **de Vries L**, **Boerjan W.** 2019. Lignin engineering in forest trees. *Front Plant Sci* **10:**912.

Chapple C, **Ladisch M**, **Meilan R.** 2007. Loosening lignin's grip on biofuel production. *Nat Biotechnol* **25:**746–748.

Chen F, **Dixon RA.** 2007. Lignin modification improves fermentable sugar yields for biofuel production. *Nat Biotechnol* **25:**759–761.

Chi M, **Bhagwat B**, **Lane WD**, **Tang G**, **Su Y**, **Sun R**, **Oomah BD**, **Wiersma PA**, **Xiang Y.** 2014. Reduced polyphenol oxidase gene expression and enzymatic browning in potato (*Solanum tuberosum* L.) with artificial microRNAs. *BMC Plant Biol* **14:**62.

Clark M, **Maselko M.** 2020. Transgene biocontainment strategies for molecular farming. *Front Plant Sci* **11:**210.

Corbin DR, **Grebenok RJ**, **Ohnmeiss TE**, **Greenplate JT**, **Purcell JP.** 2001. Expression and chloroplast targeting of cholesterol oxidase in transgenic tobacco plants. *Plant Physiol* **126:**1116–1128.

Cui Y, **Huang S**, **Liu Z**, **Yi S**, **Zhou F**, **Chen H**, **Lin Y.** 2016. Development of novel glyphosate-tolerant *Japonica* rice lines: a step towards commercial release. *Front Plant Sci* **7:**1218.

Cuozzo M, **O'Connell KM**, **Kaniewski W**, **Fang R-X**, **Chua N-H**, **Tumer NE.** 1988. Viral protection in transgenic tobacco plants expressing the cucumber mosaic virus coat protein or its antisense RNA. *Bio/Technology* **6:**549–557.

Dandekar AM, Gouran H, Ibáñez AM, Uratsu SL, Agüero CB, McFarland S, Borhani Y, Feldstein PA, Bruening G, Nascimento R, Goulart LR, Pardington PE, Chaudhary A, Norvell M, Civerolo E, Gupta G. 2012. An engineered innate immune defense protects grapevines from Pierce disease. *Proc Natl Acad Sci USA* **109:**3721–3725.

De Cosa B, Moar W, Lee S-B, Miller M, Daniell H. 2001. Overexpression of the *Bt cry*2Aa2 operon in chloroplasts leads to formation of insecticidal crystals. *Nat Biotechnol* **19:**71–74.

Dekker J, Duke SO. 1995. Herbicide-resistant field crops. *Adv Agron* **54:**69–116.

Delannay X, LaVallee BJ, Proksch RK, Fuchs RL, Sims SR, Greenplate JT, Marrone PG, Dodson RB, Augustine JJ, Layton JG, Fischhoff DA. 1989. Field performance of transgenic tomato plants expressing the *Bacillus thuringiensis* var. *kurstaki* insect control protein. *Bio/Technology* **7:**1265–1269.

De Steur H, Mehta S, Gellynck X, Finkelstein JL. 2017. GM biofortified crops: potential effects on targeting the micronutrient intake gap in human populations. *Curr Opin Biotechnol* **44:**181–188.

Douglas AE. 2018. Strategies for enhanced crop resistance to insect pests. *Annu Rev Plant Biol* **69:**637–660.

Duan X, Li X, Xue Q, Abo-el-Saad M, Xu D, Wu R. 1996. Transgenic rice plants harboring an introduced potato proteinase inhibitor II gene are insect resistant. *Nat Biotechnol* **14:**494–498.

Dunse KM, Stevens JA, Lay FT, Gaspar YM, Heath RL, Anderson MA. 2010. Coexpression of potato type I and II proteinase inhibitors gives cotton plants protection against insect damage in the field. *Proc Natl Acad Sci USA* **107:**15011–15015.

Díaz de la Garza RI, Gregory JF III, Hanson AD. 2007. Folate biofortification of tomato fruit. *Proc Natl Acad Sci USA* **104:**4218–4222.

Ely S. 1993. The engineering of plants to express *Bacillus thuringiensis* δ-endotoxins, p 105–124. *In* Entwistle PF, Cory JS, Bailey MJ, Higgs S (ed), Bacillus thuringiensis, *an Environmental Biopesticide: Theory and Practice*. John Wiley & Sons, Chichester, United Kingdom.

Falco SC, Guida T, Locke M, Mauvais J, Sanders C, Ward RT, Webber P. 1995. Transgenic canola and soybean seeds with increased lysine. *Biotechnology (N Y)* **13:**577–582.

Fang J, Xu X, Wang P, Zhao JZ, Shelton AM, Cheng J, Feng M-G, Shen Z. 2007. Characterization of chimeric *Bacillus thuringiensis* Vip3 toxins. *Appl Environ Microbiol* **73:** 956–961.

Fartyal D, Agarwal A, James D, Borphukan B, Ram B, Sheri V, Agrawal PK, Achary VMM, Reddy MK. 2018. Developing dual herbicide tolerant transgenic rice plants for sustainable weed management. *Sci Rep* **8:**11598.

Ferry N, Edwards MG, Gatehouse J, Capell T, Christou P, Gatehouse AMR. 2006. Transgenic plants for insect pest control: a forward looking scientific perspective. *Transgenic Res* **15:**13–19.

Fiedler U, Conrad U. 1995. High-level production and long-term storage of engineered antibodies in transgenic tobacco seeds. *Biotechnology (N Y)* **13:**1090–1093.

Fitch MMM, Manshardt RM, Gonsalves D, Slightom JL, Sanford JC. 1992. Virus resistant papaya plants derived from tissues bombarded with the coat protein gene of papaya ringspot virus. *Bio/Technology* **10:**1466–1472.

Fitchen JH, Beachy RN. 1993. Genetically engineered protection against viruses in transgenic plants. *Annu Rev Microbiol* **47:**739–763.

Fuchs M, Gonsalves D. 1995. Resistance of transgenic hybrid squash ZW-20 expressing the coat protein genes of zucchini yellow mosaic virus and watermelon mosaic virus 2 to mixed infections by both potyviruses. *Bio/Technology* **13:** 1466–1473.

Garg AK, Kim J-K, Owens TG, Ranwala AP, Choi YD, Kochian LV, Wu RJ. 2002. Trehalose accumulation in rice plants confers high tolerance levels to different abiotic stresses. *Proc Natl Acad Sci USA* **99:**15898–15903.

Giddings G, Allison G, Brooks D, Carter A. 2000. Transgenic plants as factories for biopharmaceuticals. *Nat Biotechnol* **18:**1151–1155.

Gil-Humanes J, Pistón F, Tollefsen S, Sollid LM, Barro F. 2010. Effective shutdown in the expression of celiac disease-related wheat gliadin T-cell epitopes by RNA interference. *Proc Natl Acad Sci USA* **107:**17023–17028.

Giritch A, Marillonnet S, Engler C, van Eldik G, Botterman J, Klimyuk V, Gleba Y. 2006. Rapid high-yield expression of full-size IgG antibodies in plants coinfected with noncompeting viral vectors. *Proc Natl Acad Sci USA* **103:**14701–14706.

Gonsalves D. 1998. Control of papaya ringspot virus in papaya: a case study. *Annu Rev Phytopathol* **36:**415–437.

Goto F, Yoshihara T, Shigemoto N, Toki S, Takaiwa F. 1999. Iron fortification of rice seed by the soybean ferritin gene. *Nat Biotechnol* **17:**282–286.

Gray JE, Picton S, Giovannoni JJ, Grierson D. 1994. The use of transgenic and naturally occurring mutants to understand and manipulate tomato fruit ripening. *Plant Cell Environ* **17:**557–571.

Grison R, Grezes-Besset B, Schneider M, Lucante N, Olsen L, Leguay J-J, Toppan A. 1996. Field tolerance to fungal pathogens of *Brassica napus* constitutively expressing a chimeric chitinase gene. *Nat Biotechnol* **14:**643–646.

Halpin C. 2005. Gene stacking in transgenic plants: the challenge for 21st century plant biotechnology. *Plant Biotechnol J* **3:**141–155.

Hanin M, Ebel C, Ngom M, Laplaze L, Masmoudi K. 2016. New insights on plant salt tolerance mechanisms and their potential use for breeding. *Front Plant Sci* **7:**1787.

Holmberg N, Lilius G, Bailey JE, Bülow L. 1997. Transgenic tobacco expressing *Vitreoscilla* hemoglobin exhibits enhanced growth and altered metabolite production. *Nat Biotechnol* **15:**244–247.

Hu W-J, Harding SA, Lung J, Popko JL, Ralph J, Stokke DD, Tsai C-J, Chiang VL. 1999. Repression of lignin biosynthesis promotes cellulose accumulation and growth in transgenic trees. *Nat Biotechnol* **17:**808–812.

Ishibashi K, Saruta M, Shimizu T, Shu M, Anai T, Komatsu K, Yamada N, Katayose Y, Ishikawa M, Ishimoto M, Kaga A.

2019. Soybean antiviral immunity conferred by dsRNase targets the viral replication complex. *Nat Commun* **10:**4033.

Ji X, **Zhang H**, **Zhang Y**, **Wang Y**, **Gao C.** 2015. Establishing a CRISPR-Cas-like immune system conferring DNA virus resistance in plants. *Nat Plants* **1:**15144.

Ji X, **Si X**, **Zhang Y**, **Zhang H**, **Zhang F**, **Gao C.** 2018. Conferring DNA virus resistance with high specificity in plants using virus-inducible genome-editing system. *Genome Biol* **19:**197.

Jobling SA, **Westcott RJ**, **Tayal A**, **Jeffcoat R**, **Schwall GP.** 2002. Production of a freeze-thaw-stable potato starch by antisense inhibition of three starch synthase genes. *Nat Biotechnol* **20:**295–299.

Jongedijk E, **de Schutter AAJM**, **Stolte T**, **van den Elzen PJM**, **Cornelissen BJC.** 1992. Increased resistance to potato virus X and preservation of cultivar properties in transgenic potato under field conditions. *Biotechnology (N Y)* **10:**422–429.

Jouanin L, **Bonadé-Bottino M**, **Girard C**, **Morrot G**, **Giband M.** 1998. Transgenic plants for insect resistance. *Plant Sci* **131:**1–11.

Khaabbazi SD, **Khabbazi AD**, **Özcan SF**, **Bakhsh A**, **Basalma D.** 2018. Expression of *GNA* and biting site-restricted *cry1Ac* in cotton; an efficient attribution to insect pest management strategies. *Plant Biotechnol Rep* **12:**273–282.

Klee HJ, **Hayford MB**, **Kretzmer KA**, **Barry GF**, **Kishore GM.** 1991. Control of ethylene synthesis by expression of a bacterial enzyme in transgenic tomato plants. *Plant Cell* **3:**1187–1193.

Knutzon DS, **Thompson GA**, **Radke SE**, **Johnson WB**, **Knauf VC**, **Kridl JC.** 1992. Modification of *Brassica* seed oil by antisense expression of a stearoyl-acyl carrier protein desaturase gene. *Proc Natl Acad Sci USA* **89:**2624–2628.

Kota M, **Daniell H**, **Varma S**, **Garczynski SF**, **Gould F**, **Moar WJ.** 1999. Overexpression of the *Bacillus thuringiensis* (Bt) Cry2Aa2 protein in chloroplasts confers resistance to plants against susceptible and Bt-resistant insects. *Proc Natl Acad Sci USA* **96:**1840–1845.

Langner T, **Kamoun S**, **Belhaj K.** 2018. CRISPR crops: plant genome editing toward disease resistance. *Annu Rev Phytopathol* **56:**479–512.

Lereclus D, **Delécuse A**, **Lecadet MM.** 1993. Diversity of *Bacillus thuringiensis* toxins and genes, p 37–69. *In* Entwistle PF, Cory JS, Bailey MJ, Higgs S (ed), Bacillus thuringiensis, *an Environmental Biopesticide: Theory and Practice*. John Wiley & Sons, Chichester, United Kingdom.

Li Y, **Zhu J**, **Wu H**, **Liu C**, **Huang C**, **Lan J**, **Zhao Y**, **Xie C.** 2020. Precise base editing of non-allelic acetolactate synthase genes confers sulfonylurea herbicide resistance in maize. *Crop J* **8:**449–456.

Liang C, **Sun B**, **Meng Z**, **Meng Z**, **Wang Y**, **Sun G**, **Zhu T**, **Lu W**, **Zhang W**, **Malik W**, **Lin M**, **Zhang R**, **Guo S.** 2017. Co-expression of *GR79 EPSPS* and *GAT* yields herbicide-resistant cotton with low glyphosate residues. *Plant Biotechnol J* **15:**1622–1629.

Lin W, **Anuratha CS**, **Datta K**, **Potrykus I**, **Muthukrishnan S**, **Datta SK.** 1995. Genetic engineering of rice for resistance to sheath blight. *Bio/Technology* **13:**686–691.

Lin Y, **Hung C-Y**, **Bhattacharya C**, **Nichols S**, **Rahimuddin H**, **Kittur FS**, **Leung T**, **Xie J.** 2018. An effective way of producing fully assembled antibody in transgenic tobacco plants by linking heavy and light chains via a self-cleaving 2A peptide. *Front Plant Sci* **9:**1379.

Liu Y, **Cao G**, **Chen R**, **Zhang S**, **Ren Y**, **Lu W**, **Wang J**, **Wang G.** 2015. Transgenic tobacco simultaneously overexpressing glyphosate N-acetyltransferase and 5-enolpyruvylshikimate-3-phosphate synthase are more resistant to glyphosate than those containing one gene. *Transgenic Res* **24:**753–763.

Lodge JK, **Kaniewski WK**, **Tumer NE.** 1993. Broad-spectrum virus resistance in transgenic plants expressing pokeweed antiviral protein. *Proc Natl Acad Sci USA* **90:**7089–7093.

Lomonossoff GP, **D'Aoust M-A.** 2016. Plant-produced biopharmaceuticals: a case of technical developments driving clinical deployment. *Science* **353:**1237–1240.

Lucca P, **Hurrell R**, **Potrykus I.** 2001. Genetic engineering approaches to improve the bioavailability and the level of iron in rice grains. *Theor Appl Genet* **102:**392–397.

MacIntosh SC, **Kishore GM**, **Perlak FJ**, **Marrone PG**, **Stone TB**, **Sims SR**, **Fuchs RL.** 1990. Potentiation of *Bacillus thuringiensis* insecticidal activity by serine protease inhibitors. *J Agric Food Chem* **38:**1145–1152.

Mao Y-B, **Cai W-J**, **Wang J-W**, **Hong G-J**, **Tao X-Y**, **Wang L-J**, **Huang Y-P**, **Chen X-Y.** 2007. Silencing a cotton bollworm P450 monooxygenase gene by plant-mediated RNAi impairs larval tolerance of gossypol. *Nat Biotechnol* **25:**1307–1313.

Margolin E, **Chapman R**, **Meyers AE**, **van Diepen MT**, **Ximba P**, **Hermanus T**, **Crowther C**, **Weber B**, **Morris L**, **Williamson A-L**, **Rybicki EP.** 2019. Production and immunogenicity of soluble plant-produced HIV-1 subtype C envelope gp140 immunogens. *Front Plant Sci* **10:**1378.

Marillonnet S, **Thoeringer C**, **Kandzia R**, **Klimyuk V**, **Gleba Y.** 2005. Systemic *Agrobacterium tumefaciens*-mediated transfection of viral replicons for efficient transient expression in plants. *Nat Biotechnol* **23:**718–723.

Mayta ML, **Arce RC**, **Zurbriggen MD**, **Valle EM**, **Hajirezaei M-R**, **Zanor MI**, **Carrillo N.** 2019. Expression of a chloroplast-targeted cyanobacterial flavodoxin in tomato plants increases harvest index by altering plant size and productivity. *Front Plant Sci* **10:**1432.

Mehlo L, **Gahakwa D**, **Nghia PT**, **Loc NT**, **Capell T**, **Gatehouse JA**, **Gatehouse AM**, **Christou P.** 2005. An alternative strategy for sustainable pest resistance in genetically enhanced crops. *Proc Natl Acad Sci USA* **102:**7812–7816.

Molvig L, **Tabe LM**, **Eggum BO**, **Moore AE**, **Craig S**, **Spencer D**, **Higgins TJV.** 1997. Enhanced methionine levels and increased nutritive value of seeds of transgenic lupins (*Lupinus angustifolius* L.) expressing a sunflower seed albumin gene. *Proc Natl Acad Sci USA* **94:**8393–8398.

Mukherjee R, **Mukherjee A**, **Bandyopadhyay S**, **Mukherjee S**, **Sengupta S**, **Ray S**, **Majumder AL.** 2019. Selective manipulation of the inositol metabolic pathway for induction of salt-tolerance in *indica* rice variety. *Sci Rep* **9:**5358.

Murphy DJ. 1996. Engineering oil production in rapeseed and other oil crops. *Trends Biotechnol* **14:**206–213.

Nawrath C, **Poirier Y**, **Somerville C.** 1994. Targeting of the polyhydroxybutyrate biosynthetic pathway to the plastids of *Arabidopsis thaliana* results in high levels of polymer accumulation. *Proc Natl Acad Sci USA* **91:**12760–12764.

Niu Q-W, **Lin S-S**, **Reyes JL**, **Chen K-C**, **Wu H-W**, **Yeh S-D**, **Chua N-H.** 2006. Expression of artificial microRNAs in transgenic *Arabidopsis thaliana* confers virus resistance. *Nat Biotechnol* **24:**1420–1428.

Nochi T, **Takagi H**, **Yuki Y**, **Yang L**, **Masumura T**, **Mejima M**, **Nakanishi U**, **Matsumura A**, **Uozumi A**, **Hiroi T**, **Morita S**, **Tanaka K**, **Takaiwa F**, **Kiyono H.** 2007. Rice-based mucosal vaccine as a global strategy for cold-chain- and needle-free vaccination. *Proc Natl Acad Sci USA* **104:**10986–10991.

Oliva N, **Chadha-Mohanty P**, **Poletti S**, **Abrigo E**, **Atienza G**, **Torrizo L**, **Garcia R**, **Dueñas C Jr**, **Poncio MA**, **Balindong J**, **Manzanilla M**, **Montecillo F**, **Zaidem M**, **Barry G**, **Hervé P**, **Shou H**, **Slamet-Loedin IH.** 2014. Large-scale production and evaluation of marker-free *indica* rice IR64 expressing phytoferritin genes. *Mol Breed* **33:**23–37.

Osusky M, **Zhou G**, **Osuska L**, **Hancock RE**, **Kay WW**, **Misra S.** 2000. Transgenic plants expressing cationic peptide chimeras exhibit broad-spectrum resistance to phytopathogens. *Nat Biotechnol* **18:**1162–1166.

Ozyigit II, **Can H**, **Dogan I.** 2020. Phytoremediation using genetically engineered plants to remove metals: a review. *Environ Chem Lett* **19:**669–698.

Paine JA, **Shipton CA**, **Chaggar S**, **Howells RM**, **Kennedy MJ**, **Vernon G**, **Wright SY**, **Hinchliffe E**, **Adams JL**, **Silverstone AL**, **Drake R.** 2005. Improving the nutritional value of Golden Rice through increased pro-vitamin A content. *Nat Biotechnol* **23:**482–487.

Patankar HV, **Al-Harrasi I**, **Al-Yahyai R**, **Yaish MW.** 2018. Identification of candidate genes involved in the salt tolerance of date palm (*Phoenix dactylifera* L.) based on a yeast functional bioassay. *DNA Cell Biol* **37:**524–534.

Pérez C, **Fernandez LE**, **Sun J**, **Folch JL**, **Gill SS**, **Soberón M**, **Bravo A.** 2005. *Bacillus thuringiensis* subsp. *israelensis* Cyt1Aa synergizes Cry11Aa toxin by functioning as a membrane-bound receptor. *Proc Natl Acad Sci USA* **102:**18303–18308.

Perlak FJ, **Fuchs RL**, **Dean DA**, **McPherson SL**, **Fischhoff DA.** 1991. Modification of the coding sequence enhances plant expression of insect control protein genes. *Proc Natl Acad Sci USA* **88:**3324–3328.

Peschen D, **Li H-P**, **Fischer R**, **Kreuzaler F**, **Liao Y-C.** 2004. Fusion proteins comprising a *Fusarium*-specific antibody linked to antifungal peptides protect plants against a fungal pathogen. *Nat Biotechnol* **22:**732–738.

Prado GS, **Bamogo PKA**, **de Abreu JAC**, **Gillet F-X**, **Dos Santos VO**, **Silva MCM**, **Brizard J-P**, **Bemquerer MP**, **Bangratz M**, **Brugidou C**, **Sérémé D**, **Grossi-de-Sa MF**, **Lacombe S.** 2019. *Nicotiana benthamiana* is a suitable transient system for high-level expression of an active inhibitor of cotton boll weevil α-amylase. *BMC Biotechnol* **19:**15.

Qi B, **Fraser T**, **Mugford S**, **Dobson G**, **Sayanova O**, **Butler J**, **Napier JA**, **Stobart AK**, **Lazarus CM.** 2004. Production of very long chain polyunsaturated omega-3 and omega-6 fatty acids in plants. *Nat Biotechnol* **22:**739–745.

Qin H, **Gu Q**, **Zhang J**, **Sun L**, **Kuppu S**, **Zhang Y**, **Burow M**, **Payton P**, **Blumwald E**, **Zhang H.** 2011. Regulated expression of an isopentenyltransferase gene (IPT) in peanut significantly improves drought tolerance and increases yield under field conditions. *Plant Cell Physiol* **52:**1904–1914.

Quinn JP. 1990. Evolving strategies for the genetic engineering of herbicide resistance in plants. *Biotechnol Adv* **8:**321–333.

Rauf I, **Javaid S**, **Naqvi RZ**, **Mustafa T**, **Amin I**, **Mukhtar Z**, **Jander G**, **Mansoor S.** 2019. *In-planta* expression of insecticidal proteins provides protection against lepidopteran insects. *Sci Rep* **9:**6745.

Rivero RM, **Kojima M**, **Gepstein A**, **Sakakibara H**, **Mittler R**, **Gepstein S**, **Blumwald E.** 2007. Delayed leaf senescence induces extreme drought tolerance in a flowering plant. *Proc Natl Acad Sci USA* **104:**19631–19636.

Romeis J, **Meissle M.** 2020. Stacked Bt proteins pose no new risks to nontarget arthropods. *Trends Biotechnol* **38:**234–236.

Rosa C, **Kuo Y-W**, **Wuriyanghan H**, **Falk BW.** 2018. RNA interference mechanisms and applications in plant pathology. *Annu Rev Phytopathol* **56:**581–610.

Rosales-Mendoza S, **Nieto-Gómez R.** 2018. Green therapeutic biocapsules: using plant cells to orally deliver biopharmaceuticals. *Trends Biotechnol* **36:**1054–1067.

Ryan CA. 1990. Protease inhibitors in plants: genes for improving defenses against insect and pathogens. *Annu Rev Phytopathol* **28:**425–449.

Sahoo A, **Mandal AK**, **Dwivedi K**, **Kumar V.** 2020. A cross talk between the immunization and edible vaccine: current challenges and future prospects. *Life Sci* **261:**118343.

Sayyed AH, **Crickmore N**, **Wright DJ.** 2001. Cyt1Aa from *Bacillus thuringiensis* subsp. *israelensis* is toxic to the diamondback moth, *Plutella xylostella*, and synergizes the activity of Cry1Ac towards a resistant strain. *Appl Environ Microbiol* **67:**5859–5861.

Schellenberger U, **Oral J**, **Rosen BA**, **Wei J-Z**, **Zhu G**, **Xie W**, **McDonald MJ**, **Cerf DC**, **Diehn SH**, **Crane VC**, **Sandahl GA**, **Zhao J-Z**, **Nowatzki TM**, **Sethi A**, **Liu L**, **Pan Z**, **Wang Y**, **Lu AL**, **Wu G**, **Liu L.** 2016. A selective insecticidal protein from Pseudomonas for controlling corn rootworms. *Science* **354:**634–637.

Schnepf HE, **Whiteley HR.** 1981. Cloning and expression of the *Bacillus thuringiensis* crystal protein gene in *Escherichia coli*. *Proc Natl Acad Sci USA* **78:**2893–2897.

Service RF. 2007. A growing threat down on the farm. *Science* **316:**1114–1117.

Shade RE, **Schroeder HE**, **Pueyo JJ**, **Tabe LM**, **Murdock LL**, **Higgins TJV**, **Chrispeels MJ.** 1994. Transgenic pea seeds expressing the α-amylase inhibitor of the common bean are resistant to bruchid beetles. *Bio/Technology* **12:**793–796.

Shen H, **He X**, **Poovaiah CR**, **Wuddineh WA**, **Ma J**, **Mann DGJ**, **Wang H**, **Jackson L**, **Tang Y**, **Stewart CN Jr**, **Chen F**, **Dixon RA.** 2012. Functional characterization of the switchgrass (*Panicum virgatum*) R2R3-MYB transcription factor *PvMYB4* for improvement of lignocellulosic feedstocks. *New Phytol* **193:**121–136.

Shukla AK, **Upadhyay SK**, **Mishra M**, **Saurabh S**, **Singh R**, **Singh H**, **Thakur N**, **Rai P**, **Pandey P**, **Hans AL**, **Srivastava S**, **Rajapure V**, **Yadav SK**, **Singh MK**, **Kumar J**, **Chandrashekar K**, **Verma PC**, **Singh AP**, **Nair KN**, **Bhadauria S**, **Wahajuddin M**, **Singh S**, **Sharma S**, **Omkar**, **Upadhyay RS**, **Ranade SA**, **Tuli R**, **Singh PK.** 2016. Expression of an insecticidal fern protein in cotton protects against whitefly. *Nat Biotechnol* **34:**1046–1051.

Silva KJP, **Brunings A**, **Peres NA**, **Mou Z**, **Folta KM.** 2015. The *Arabidopsis NPR1* gene confers broad-spectrum disease resistance in strawberry. *Transgenic Res* **24:**693–704.

Song J, **Zhang R**, **Yue D**, **Chen X**, **Guo Z**, **Cheng C**, **Hu M**, **Zhang J**, **Zhang K.** 2018. Co-expression of *ApGSMT2g* and *ApDMT2g* in cotton enhances salt tolerance and increases seed cotton yield in saline fields. *Plant Sci* **274:**369–382.

Storozhenko S, **De Brouwer V**, **Volckaert M**, **Navarrete O**, **Blancquaert D**, **Zhang G-F**, **Lambert W**, **Van Der Straeten D.** 2007. Folate fortification of rice by metabolic engineering. *Nat Biotechnol* **25:**1277–1279.

Tabashnik BE, **Dennehy TJ**, **Carrière Y.** 2005. Delayed resistance to transgenic cotton in pink bollworm. *Proc Natl Acad Sci USA* **102:**15389–15393.

Tricoli DM, **Carney KJ**, **Russell PF**, **McMaster JR**, **Groff DW**, **Hadden KC**, **Himmel PT**, **Hubbard JP**, **Boeshore ML**, **Quemada HD.** 1995. Field evaluation of transgenic squash containing single or multiple virus coat protein gene constructs for resistance to cucumber mosaic virus, watermelon mosaic virus 2, and zucchini yellow mosaic virus. *Bio/Technology* **13:**1458–1465.

Töpfer R, **Martini N**, **Schell J.** 1995. Modification of plant lipid synthesis. *Science* **268:**681–686.

Tu J, **Zhang G**, **Datta K**, **Xu C**, **He Y**, **Zhang Q**, **Khush GS**, **Datta SK.** 2000. Field performance of transgenic elite commercial hybrid rice expressing *Bacillus thuringiensis* δ-endotoxin. *Nat Biotechnol* **18:**1101–1104.

Ufaz S, **Galili G.** 2008. Improving the content of essential amino acids in crop plants: goals and opportunities. *Plant Physiol* **147:**954–961.

Umezawa T, **Fujita M**, **Fujita Y**, **Yamaguchi-Shinozaki K**, **Shinozaki K.** 2006. Engineering drought tolerance in plants: discovering and tailoring genes to unlock the future. *Curr Opin Biotechnol* **17:**113–122.

van Zelm E, **Zhang Y**, **Testerink C.** 2020. Salt tolerance mechanisms of plants. *Annu Rev Plant Biol* **71:**403–433.

Vázquez-Núñez E, **Peña-Castro JM**, **Fernández-Loqueño F**, **Cejudo E, de la Rosa-Alvarez MG**, **García-Casteñeda MC.** 2018. A review on genetically modified plants designed to phytoremediate polluted soils: biochemical responses and international regulation. *Pedosph* **28:**697–712.

Verberne MC, **Verpoorte R**, **Bol JF**, **Mercado-Blanco J**, **Linthorst HJM.** 2000. Overproduction of salicylic acid in plants by bacterial transgenes enhances pathogen resistance. *Nat Biotechnol* **18:**779–783.

Vierheilig H, **Alt M**, **Neuhaus J-M**, **Boller T**, **Wiemken A.** 1993. Colonization of transgenic *Nicotiana sylvestris* plants, expressing different forms of *Nicotiana tabacum* chitinase, by the root pathogen *Rhizoctonia solani* and by the mycorrhizal symbiont *Glomus mosseae*. *Mol Plant Microbe Interact* **6:**261–264.

Virdi V, **Coddens A**, **De Buck S**, **Millet S**, **Goddeeris BM**, **Cox E**, **De Greve H**, **Depicker A.** 2013. Orally fed seeds producing designer IgAs protect weaned piglets against enterotoxigenic *Escherichia coli* infection. *Proc Natl Acad Sci USA* **110:**11809–11814.

Waltz E. 2016. Gene-edited CRISPR mushroom escapes US regulation. *Nature* **532:**293.

Wan P, **Xu D**, **Cong S**, **Jiang Y**, **Huang Y**, **Wang J**, **Wu H**, **Wang L**, **Wu K**, **Carrière Y**, **Mathias A**, **Li X**, **Tabashnik BE.** 2017. Hybridizing transgenic Bt cotton with non-Bt cotton counters resistance in pink bollworm. *Proc Natl Acad Sci USA* **114:**5413–5418.

Wang D, **Samsulrizal NH**, **Yan C**, **Allcock NS**, **Craigon J**, **Blanco-Ulate B**, **Ortega-Salazar I**, **Marcus SE**, **Bagheri HM**, **Perez Fons L**, **Fraser PD**, **Foster T**, **Fray R**, **Knox JP**, **Seymour GB.** 2019. Characterization of CRISPR mutants targeting genes modulating pectin degradation in ripening tomato. *Plant Physiol* **179:**544–557.

Wang H, **Sun S**, **Ge W**, **Zhao L**, **Hou B**, **Wang K**, **Lyu Z**, **Chen L**, **Xu S**, **Guo J**, **Li M**, **Su P**, **Li X**, **Wang G**, **Bo C**, **Fang X**, **Zhuang W**, **Cheng X**, **Wu J**, **Dong L**, **Chen W**, **Li W**, **Xiao G**, **Zhao J**, **Hao Y**, **Xu Y**, **Gao Y**, **Liu W**, **Liu Y**, **Yin H**, **Li J**, **Li X**, **Zhao Y**, **Wang X**, **Ni F**, **Ma X**, **Li A**, **Xu SS**, **Bai G**, **Nevo E**, **Gao C**, **Ohm H**, **Kong L.** 2020. Horizontal gene transfer of *Fhb7* from fungus underlies *Fusarium* head blight resistance in wheat. *Science* **368:**eaba5435.

Wen SX, **Teel LD**, **Judge NA**, **O'Brien AD.** 2006. A plant-based oral vaccine to protect against systemic intoxication by Shiga toxin type 2. *Proc Natl Acad Sci USA* **103:**7082–7087.

Whalon ME, **Wingerd BA.** 2003. Bt: mode of action and use. *Arch Insect Biochem Physiol* **54:**200–211.

Wirth MC, **Georghiou GP**, **Federici BA.** 1997. CytA enables CryIV endotoxins of *Bacillus thuringiensis* to overcome high levels of CryIV resistance in the mosquito, *Culex quinquefasciatus*. *Proc Natl Acad Sci USA* **94:**10536–10540.

Wright TR, **Shan G**, **Walsh TA**, **Lira JM**, **Cui C**, **Song P**, **Zhuang M**, **Arnold NL**, **Lin G**, **Yau K**, **Russell SM**, **Cicchillo RM**, **Peterson MA**, **Simpson DM**, **Zhou N**, **Ponsamuel J**, **Zhang Z.** 2010. Robust crop resistance to broadleaf and grass herbicides provided by aryloxyalkanoate dioxygenase transgenes. *Proc Natl Acad Sci USA* **107:**20240–20245.

Wu J, **Lawit SJ**, **Weers B**, **Sun J**, **Mongar N**, **Van Hemert J**, **Melo R**, **Meng X**, **Rupe M**, **Clapp J**, **Haug Collet K**, **Trecker L**, **Roesler K**, **Peddicord L**, **Thomas J**, **Hunt J**, **Zhou W**, **Hou Z**, **Wimmer M**, **Jantes J**, **Mo H**, **Liu L**, **Wang Y**, **Walker C**, **Danilevskaya O**, **Lafitte RH**, **Schussler JR**, **Shen B**, **Habben JE.** 2019. Overexpression of *zmm28* increases maize grain yield in the field. *Proc Natl Acad Sci USA* **116:**23850–23858.

Yamaguchi T, **Blumwald E.** 2005. Developing salt-tolerant crop plants: challenges and opportunities. *Trends Plant Sci* **10:**615–620.

Yang X, **Niu L**, **Zhang W**, **He H**, **Yang J**, **Xing G**, **Guo D**, **Zhao Q**, **Zhong X**, **Li H**, **Li Q**, **Dong Y.** 2019. Increased multiple virus resistance in transgenic soybean overexpressing the double-strand RNA-specific ribonuclease gene PAC1. *Transgenic Res* **28:**129–140.

Ye X, Al-Babili S, Klöti A, Zhang J, Lucca P, Beyer P, Potrykus I. 2000. Engineering the provitamin A (β-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. *Science* **287:**303–305.

Yu J, Langridge WH. 2001. A plant-based multicomponent vaccine protects mice from enteric diseases. *Nat Biotechnol* **19:**548–552.

Yu J, Tu L, Subburaj S, Bae S, Lee G-J. 2020. Simultaneous targeting of duplicated genes in *Petunia* protoplasts for flower color modification via CRISPR-Cas9 ribonucleoproteins. *Plant Cell Rep* **40:**1037–1045.

Yusibov V, Kushnir N, Streatfield SJ. 2016. Antibody production in plants and green algae. *Annu Rev Plant Biol* **67:**669–701.

Zhang B, Oakes AD, Newhouse AE, Baier KM, Maynard CA, Powell WA. 2013. A threshold level of oxalate oxidase transgene expression reduces *Cryphonectria parasitica*-induced necrosis in a transgenic American chestnut (*Castanea dentata*) leaf bioassay. *Transgenic Res* **22:**973–982.

Zhang H-X, Blumwald E. 2001. Transgenic salt-tolerant tomato plants accumulate salt in foliage but not in fruit. *Nat Biotechnol* **19:**765–768.

Zhang J, Khan SA, Hasse C, Ruf S, Heckel DG, Bock R. 2015. Pest control. Full crop protection from an insect pest by expression of long double-stranded RNAs in plastids. *Science* **347:**991–994.

Zhang L, French R, Langenberg WG, Mitra A. 2001. Accumulation of barley stripe mosaic virus is significantly reduced in transgenic wheat plants expressing a bacterial ribonuclease. *Transgenic Res* **10:**13–19.

Zhang W, Lu Y, van der Werf W, Huang J, Wu F, Zhou K, Deng X, Jiang Y, Wu K, Rosegrant MW. 2018. Multidecadal, county-level analysis of the effects of land use, Bt cotton, and weather on cotton pests in China. *Proc Natl Acad Sci USA* **115:**E7700–E7709.

Zhao J-Z, Cao J, Li Y, Collins HL, Roush RT, Earle ED, Shelton AM. 2003. Transgenic plants expressing two *Bacillus thuringiensis* toxins delay insect resistance evolution. *Nat Biotechnol* **21:**1493–1497.

Zhu Q, Maher EA, Masoud S, Dixon RA, Lamb CJ. 1994. Enhanced protection against fungal attack by constitutive co-expression of chitinase and glucanase genes in transgenic tobacco. *Bio/Technology* **12:**807–812.

review questions

1. A local crop is being ravaged by a nonenveloped virus with a single-stranded RNA genome (8,000 nucleotides long). The virus and its RNA can be readily isolated. In addition, you have antibodies against all four of the proteins encoded within the viral genome. Describe a strategy that you could use to protect the crop against this viral infection and prevent subsequent damage.
2. Suggest several different strategies for developing insect-resistant plants.
3. How can protease inhibitors, α-amylase inhibitor, cholesterol oxidase, Vip proteins, and ricin each protect a plant against damage from insect predation?
4. How can RNAi be used to protect plants against damage from insect predation? How can this strategy be made more effective?
5. How can a modified *Bacillus thuringiensis* toxin help to overcome possible insect mutations that make the target insect resistant to the native form of the toxin?
6. Suggest several strategies that may be used to prevent target insects from becoming resistant to *Bacillus thuringiensis*-transformed plants.
7. How would you engineer plants to be lethal to aphids that otherwise damage the plant?
8. Suggest several strategies for simultaneously protecting a plant against damage from several different viruses.
9. How can RNAi be used to protect plants against damage from plant viruses?
10. How can the CRISPR/Cas system be used to protect plants against geminiviruses?
11. What general strategies can be employed in genetically engineering plants to be resistant to herbicides?
12. Suggest two different strategies for engineering plants that are resistant to the herbicide glyphosate. Why is this important?
13. How can crop plants be engineered to be resistant to the herbicide dicamba?
14. How can plants be engineered to resist damage from pathogenic soil fungi?
15. How can single-chain Fv antibodies be engineered to protect plants against fungal pathogens?
16. How can a plant's systematic acquired resistance response be engineered to confer resistance to a broad spectrum of both fungal and bacterial pathogens?
17. How would you protect grape plants from being damaged by the bacterium *Xylella fastidiosa*?
18. Suggest some strategies that could be used to engineer plants that are resistant to growth inhibition by salt and by drought.
19. You have been asked by an avocado grower to find a way to genetically engineer his crop to prevent it from ripening during shipping. What experimental approaches would you consider?
20. How can soybeans be genetically manipulated to increase their lysine content?
21. How can vitamins be overproduced in rice plants?

22. How would you engineer soybean plants to overproduce long-chain omega-3 and omega-6 fatty acids?

23. How can the level of bioavailable iron in plants be increased?

24. How would you engineer rice plants to overproduce tetrahydrofolate? Why would you do this?

25. How would you increase the amount of amylose compared to amylopectin in potato starch? How would you increase the amount of amylopectin compared to amylose? What is the advantage of performing these manipulations?

26. How can plants be genetically modified to lower the amount of gluten that they produce?

27. Why are plants an attractive host system, compared to bacteria and animal cells in culture, for the production of human therapeutic proteins?

28. How does an antigen expressed in a transgenic plant act as an edible vaccine?

29. How would you engineer an edible vaccine directed against *V. cholerae*-caused diarrhea?

30. Briefly describe a vector system that may be used to engineer plants to produce large amounts of full-size IgG molecules.

31. How can plants be engineered to produce a unique antibody molecule to prevent gastrointestinal infection caused by enterotoxigenic *E. coli*?

32. Describe a strategy for developing a plant vaccine against type 2 Shiga toxin.

33. How would you engineer an edible vaccine directed against *V. cholerae*-caused diarrhea?

34. How can a plant's oxygen levels be manipulated? How does manipulation of a plant's oxygen levels affect the yield of that plant?

35. How can the lignin content of trees be decreased? What is the benefit of this type of genetic manipulation?

Transgenic Animal Methodologies
DNA Microinjection Method
Retroviral Vector Method
Engineered Embryonic Stem Cell Method
Somatic Cell Nuclear Transfer for Transgenic Livestock
Genome Editing with the CRISPR-Cas System
Conditional Gene Modification with the Cre–*loxP* Recombination System
Control of Transgene Expression with the Tetracycline-Inducible System
Gene Knockdown by RNA Interference
Transgenic Animal Models of Human Diseases
Mouse Models of Alzheimer's Disease
Mouse Model of Duchenne Muscular Dystrophy
Rabbit Models of Cardiovascular Disease
Zebrafish Melanoma Model
Nonhuman Primate Models of Neurodevelopmental Disorders
Animal Bioreactors for Production of Recombinant Therapeutic Proteins
Production of Recombinant Antithrombin in Goat Milk
Production of a Human Protease Inhibitor in Rabbits
Production of Therapeutic Proteins in Chicken Eggs
Production of Donor Organs in Pigs
Enhancing Production Traits of Food Animals
Disease-Resistant Livestock
Improving Milk Quality
Increasing Muscle Mass in Cattle
Enhancing Growth of Salmon
Gene Drives To Eradicate Vector-Transmitted Diseases
Malaria Vector Population Suppression
Dengue Fever Virus-Resistant Mosquitoes
Reversal Drives
SUMMARY
REFERENCES
REVIEW QUESTIONS

12 Transgenic Animals

USING TRADITIONAL BREEDING METHODS, many generations of selective matings are required to improve livestock and other domesticated animals genetically for traits such as milk yield, wool characteristics, rate of weight gain, and egg-laying frequency. At each successive generation, animals with superior performance characteristics are used as breeding stock. Eventually, high-production animals are developed as more or less pure breeding lines. This combination of mating and selection has been exceptionally successful, and today, almost all aspects of the biological basis of livestock production can be attributed to this process. However, it is time-consuming and costly, and introduction of new traits into an established genetic line requires a new program of multigenerational crosses with rigorous selection procedures to ensure that a new breeding line retains its original attributes and that the new trait has been established.

Early in the 1980s, progress in the development of techniques to transfer genes into the chromosomal DNA of animal cells (that is, to transfect animal cells) led to the birth of the first transgenic animal, a mouse. Since then, transgenic animals have been produced for many different species of mammals, birds, amphibians, fish, and invertebrates. Conceptually, the strategy is simple. A target gene is introduced into the nucleus of an early-stage embryo, where it is integrated into a chromosome. The embryos are implanted into a receptive female when successful completion of embryonic development is not possible outside a female, as is the case for mammals. Some of the offspring derived from the embryos will carry the target gene in all of their cells. Animals with the target gene integrated

Molecular Biotechnology: Principles and Applications of Recombinant DNA, Sixth Edition. Bernard R. Glick and Cheryl L. Patten.
© 2022 American Society for Microbiology.

Genetic Transformation of Mouse Embryos by Microinjection of Purified DNA

Somatic Expression of Herpes Thymidine Kinase in Mice Following Injection of a Fusion Gene into Eggs

milestone

Gordon et al. (*Proc. Natl. Acad. Sci. USA* **77:**7380–7384, 1980) were the first to show the feasibility of DNA transfer by microinjection into the pronucleus of mouse eggs. In their study, the procedure was tested by microinjecting several hundred eggs with a vector–gene construct that consisted of pBR322 carrying both the herpes simplex virus (HSV) thymidine kinase gene and a piece of the simian virus 40 genome. Of the 78 offspring that were obtained from the surrogate mothers, 2 retained some plasmid DNA. The authors concluded, "These data demonstrate that it is possible to use a recombinant plasmid as a vector to transfer foreign genes directly into mouse embryos, and that these embryos can maintain the foreign genes throughout development." Unfortunately, the plasmid DNA was not intact, and the HSV thymidine kinase sequence did not become a transgene.

On the other hand, Brinster et al. (*Cell* **27:**223–231, 1981), who microinjected the pronuclei of a number of mouse eggs with a plasmid carrying the HSV thymidine kinase gene under the control of the promoter of the metallothionein-I gene, found that one of their transgenic mice expressed HSV thymidine kinase at a high level in its liver and kidney in comparison to three other transgenic mice that produced low levels of this enzyme. In addition, eight other transgenic animals carried the HSV thymidine kinase sequence but did not produce any active HSV thymidine kinase. Southern blot analysis showed that all of the transgenic mice contained multiple copies of the microinjected DNA.

These two studies laid the foundation for transgenesis of mice. Despite the technical complexity and relative inefficiency of the microinjection strategy, it has been exceptionally successful. Currently, scores of strains of mice with either foreign genes (transgenic mice) or endogenous genes that have been disrupted by the insertion of foreign DNA (knockout mice) are being used for studying gene regulation, mammalian development, viral pathogenesis, cancer, toxicology, and the mutagenicity of various agents, among other things. In addition, transgenic and knockout mice are of considerable biomedical importance as model systems for human diseases.

in their germ line cells are bred to establish new genetic lines. Studies with these transgenic animals have contributed to an understanding of fundamental biological processes such as gene function and regulation, molecular genetics of development, and animal physiology. They are also useful model systems for studying human diseases.

The genetic improvement of animals of commercial value by the introduction of transgenes encoding desirable traits is now being realized. In 2015, the U.S. FDA approved the first genetically engineered food animal for human consumption, Atlantic salmon carrying a transgene encoding a growth hormone from Chinook salmon that grows faster than its nontransgenic counterpart. Transgenic livestock are also being used to produce pharmaceutically important proteins in their mammary glands. The term "pharming" was coined to convey the idea that milk from transgenic farm ("pharm") animals can be a source of authentic human protein drugs or pharmaceuticals. Milk is a renewable, secreted body fluid that is produced in substantial quantities and can be collected frequently without harm to the animal. A novel drug protein that is confined to the mammary gland and secreted into milk should have little to no side effects on the normal physiological processes of the transgenic animal and should undergo posttranslational modifications that at least closely match those in humans. Purification of a protein from milk, which contains only a small number of different proteins (Table 12.1), is relatively straightforward. In 2009, the use of the first human therapeutic protein, antithrombin, produced in goat's milk, was approved by the U.S. FDA. Since then, pharmaceutical proteins

Table 12.1 Protein composition of milk from cattle and sheep

Concentration (g/liter) in milk from:		
Protein	**Cattle**	**Sheep**
Caseins		
αs_1-Casein	10.0	12.0
αs_2-Casein	3.4	3.8
κ-Casein	3.9	4.6
β-Casein	10.0	16.0
Major whey proteins		
α-Lactalbumin	1.0	0.8
β-Lactalbumin	3.0	2.8
Other proteins		
Serum albumin	0.4	Unknown
Lysozyme	Trace	Unknown
Lactoferrin	0.1	Unknown
Immunoglobulins	0.7	Unknown

produced in rabbit's milk and in chicken eggs have been approved, and many more are in development.

Transgenic Animal Methodologies

Transgenic technology was developed and perfected in the laboratory mouse. Since the 1980s, hundreds of different genes have been introduced into various mouse strains. Today, many different animals have been genetically modified, although not all transgenesis techniques are successful in all animals. In general, DNA can be introduced into an animal by (i) microinjection into the enlarged sperm nucleus (male pronucleus) of a fertilized egg, (ii) retroviral vectors that infect the cells of an early-stage embryo, (iii) *in vitro* transfection of embryonic stem cells which are then inserted into an early-stage developing embryo, or (iv) transfer of a genetically manipulated somatic cell nucleus into an enucleated egg. For many animals, the embryos are then implanted into a receptive female for development.

DNA Microinjection Method

Microinjection of DNA is the preferred method for producing transgenic mice, although it is inefficient for livestock transgenesis. Female egg donors are stimulated to superovulate to increase the number of fertilized eggs that are available to be microinjected (Fig. 12.1). Female mice are given an initial injection of pregnant mare serum gonadotropin and another injection, about 48 hours later, of human chorionic gonadotropin. A superovulated mouse produces about 35 eggs instead of the normal 5 to 10. The superovulated females are mated so that eggs become fertilized and the fertilized eggs are flushed from their oviducts. Microinjection of the fertilized eggs usually occurs immediately after their collection. The microinjected transgene construct is generally a gene, complementary DNA (cDNA), or part of a gene, is usually in a linear form, and is free of vector DNA sequences.

In mammals, after entry of the sperm into the egg, the sperm nucleus (male pronucleus) and female nucleus are separate entities. The female nucleus completes its meiotic division to become a female pronucleus, and then fusion of the two nuclei (karyogamy) occurs. Prior to nuclear fusion, the male pronucleus, which tends to be larger than the female pronucleus, can be located by using a dissecting microscope. The egg can then be maneuvered and held in place by micromanipulation while the DNA is microinjected into the male pronucleus. Some of the injected DNA molecules will integrate into a chromosome.

After microinjection, 25 to 40 eggs are implanted microsurgically into a foster mother that has been made pseudopregnant by being mated to a vasectomized male. In mice, copulation is the only known way to prepare the uterus for implantation. In this case, because the male mate lacks sperm, none of the eggs of the foster mother are fertilized. The foster mother will deliver pups from the microinjected eggs about 3 weeks after implantation.

For identification of transgenic animals, DNA from a small piece of tail or ear tissue is assayed by PCR for the presence of the transgene. A transgenic animal can be mated to another animal to determine if the transgene is in the germ line of the founder animal. Subsequently, progeny can be bred with each other to form pure (homozygous) transgenic lines.

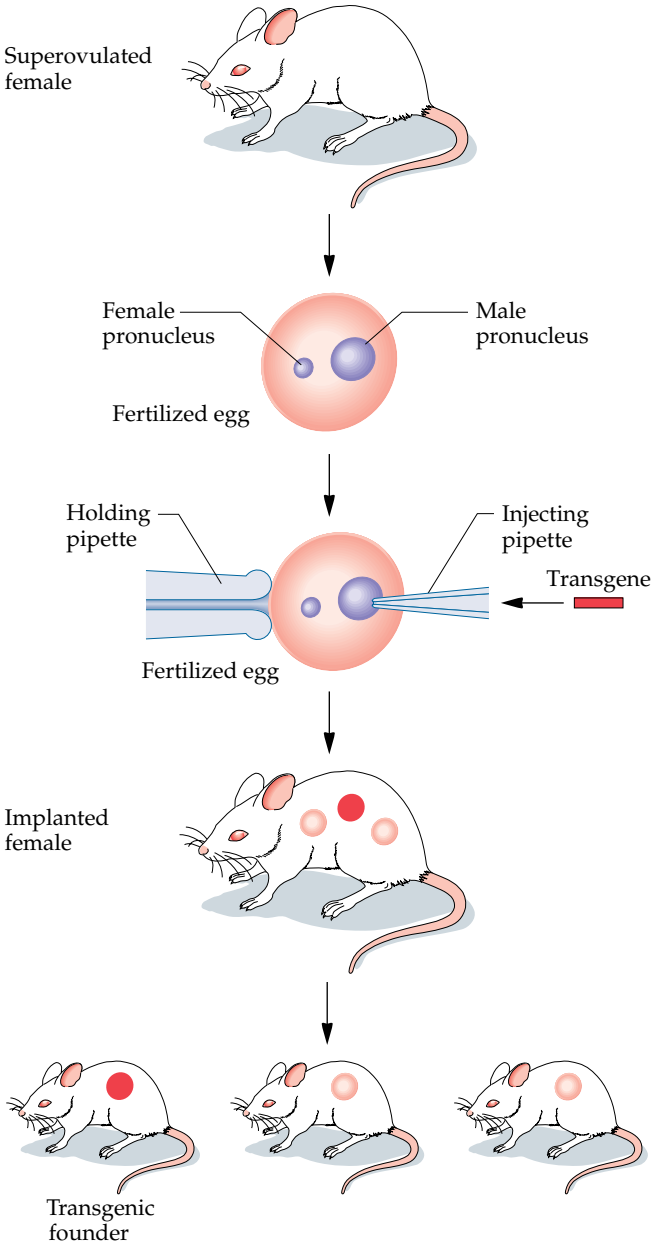


Figure 12.1 Establishing transgenic mice by DNA microinjection. Eggs are obtained from donor females that have been induced to superovulate and then are mated with males. Purified transgene DNA is microinjected into the male pronucleus of a fertilized egg. Implanted females (foster mothers) give birth to transgenic pups, some of which carry the transgene in their germ line (transgenic founder) and can be used to establish transgenic lines.

The procedure, although apparently simple, requires the coordination of a number of experimental steps. Even a highly trained practitioner can expect, at best, only 5% of the microinjected eggs to develop into live transgenic animals (Fig. 12.2), and therefore, large numbers of microinjected fertilized eggs must be used. Not all of the transgenic pups will have the

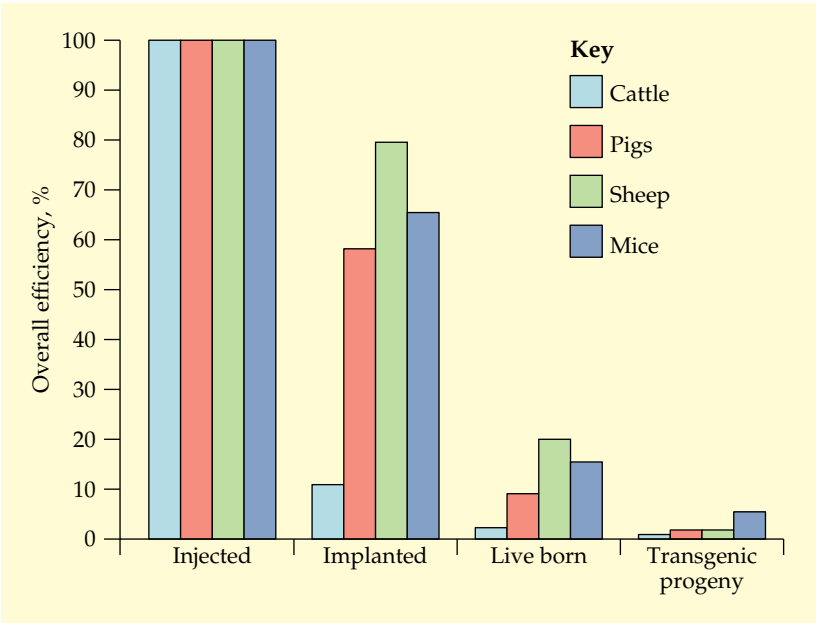


Figure 12.2 Overall efficiency of the transgenesis process after DNA microinjection. All the fertilized eggs (100%) of cattle, pigs, sheep, and mice are microinjected with a transgene, but the success of implantation and giving birth to offspring is much lower, and only 5% or fewer of the treated eggs become transgenic progeny.

appropriate characteristic. With this method, the injected DNA integrates at a random site in the genome, and often multiple copies of the injected DNA are incorporated at one site. In some individuals, the copy number may be excessive and may lead to overexpression, which disrupts the normal physiology of the animal. In others, the transgene may be poorly expressed because it has been integrated into a transcriptionally inactive region of the genome, it is negatively regulated by sequences surrounding the site of insertion, or important gene-specific regulatory sequences that lie either upstream or downstream of the gene were not retained as part of the insert.

High-capacity vectors have been developed that enable inclusion of large DNA sequences that regulate or prevent the surrounding DNA from decreasing transgene expression (see chapter 3). They are also useful for introducing large multigene sequences. These vectors are derived from bacterial and yeast artificial chromosomes (BACs and YACs, respectively) that can carry DNA inserts ranging in size from 100 to more than 1,000 kb. A number of transgenic mice have been produced by microinjection of the pronucleus of the fertilized egg with artificial chromosomes that carry either an array of related genes or a single large gene. The production of mice that synthesize a nearly complete repertoire of human antibodies is a noteworthy example of YAC transgenesis (chapter 5, Fig. 5.32). Notwithstanding the overall inefficiency, it has become routine to use DNA microinjection of the male pronucleus to create lines of mice carrying functional transgenes to study developmental processes, to use as models for human disorders, and for the production of human therapeutic agents.

Retroviral Vector Method

Retroviral vectors exploit the capacity of retroviruses to deliver and integrate a transgene into the genome of a recipient cell. Retroviruses have RNA genomes that are a template for reverse transcriptase to synthesize a linear double-stranded DNA copy that can be inserted into a chromosome of an infected host cell. To create transgenic animals, an engineered retroviral vector carrying the transgene is injected into early-stage embryos, and then infected embryos are transferred into recipient females for development (Fig. 12.3). This method yields high rates of transgenesis in mice; often

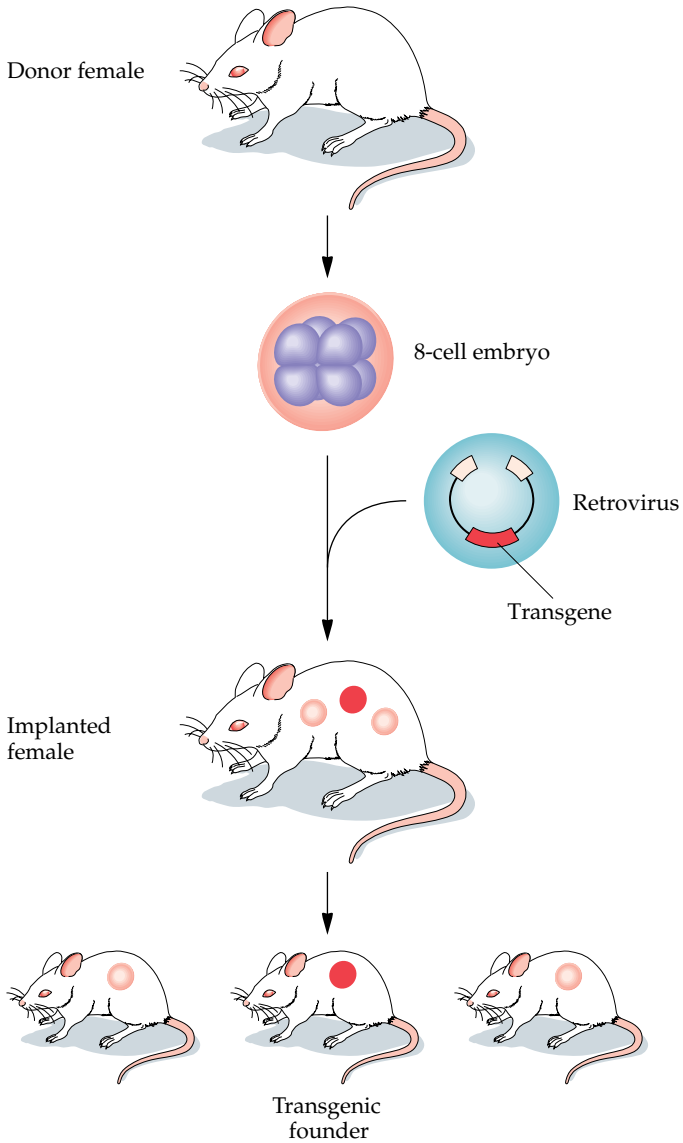


Figure 12.3 Establishing transgenic mice with retroviral vectors. Cleavage-stage embryos, usually at the eight-cell stage, are infected with a defective retrovirus carrying a transgene. Implanted females (foster mothers) give birth to transgenic pups. Matings are carried out to determine which pups have the transgene in their germ line cells. Transgenic lines can be established from these founder transgenic animals.

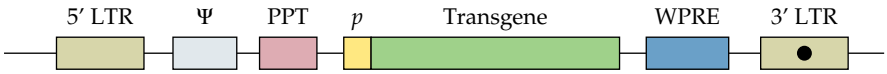


Figure 12.4 A retroviral transfer vector for introducing transgenes into animal cells. The transgene with an appropriate promoter sequence (*p*) is inserted into the retroviral (lentiviral) vector. The transfer vector is introduced into a packaging cell line that produces the viral proteins required for the production of the viral RNA, including the transgene sequence, and for packaging the RNA into viral particles that will be used to deliver the transgene into animal cells (Fig. 12.5). Long terminal repeats (LTR) at the 5′ and 3′ ends of the vector are required for production of lentiviral RNA, and the packaging signal (Ψ) is required for packaging the RNA into viral particles. Following infection of animal cells with lentivirus, lentiviral RNA is reverse transcribed, and the transgene is integrated into the animal cell genome via sequences in the LTRs. A polypurine tract sequence (PPT) and a woodchuck posttranscriptional regulatory element (WPRE) enhance transduction of host cells and increase transgene expression in the animal cells. A regulatory element within the 3′ LTR is deleted (indicated by a black dot) to prevent production of vector RNA from a promoter contained within the LTR following introduction into host cells. Expression of the transgene is not affected by the deletion because it is expressed from its own promoter.

more than 80% of the pups are positive for the transgene. However, the integration site in the host genome is random, and transgenes introduced on some retroviral vectors are silenced in the embryos.

A retroviral vector system is composed of a transfer vector that carries the transgene and a packaging cell line that provides viral proteins for production of large quantities of viral particles that deliver the transgene to host cells. The transfer vector is constructed from the retroviral genome from which all of the viral genes encoding proteins have been removed. Only the viral signals required for encapsidation of vector RNA into viral particles and subsequent integration of the transgene DNA (with appropriate transcription and translation signals) into the host genome are retained on the transfer vector (Fig. 12.4). The transfer vector is introduced into the packaging cell line that produces the viral proteins required for production of transfer vector RNA, including the transgene sequence and the capsids into which the vector RNA is packaged (Fig. 12.5). Following infection of animal cells with the viral particles, the transfer vector RNA is reverse transcribed, and the transgene is integrated into the animal cell genome via the long terminal repeat (LTR) sequences. A deletion in the 3′ LTR prevents the virus from reproducing following introduction into host cells.

Vectors based on lentiviruses, a type of retrovirus, have been used successfully to introduce a variety of transgenes into embryonic cells or early-stage embryos of mice, pigs, cattle, and birds, and these genes are highly expressed either ubiquitously or in a tissue-specific manner, depending on the promoter used. Lentiviruses are also often used as a delivery vehicle to introduce a construct to decrease gene expression in mice and other animals by RNA interference (see below).

Engineered Embryonic Stem Cell Method

Cells from the early, blastocyst stage of a developing embryo can proliferate in cell culture and still retain the capability of differentiating into all other cell types—including germ line cells—after they are reintroduced

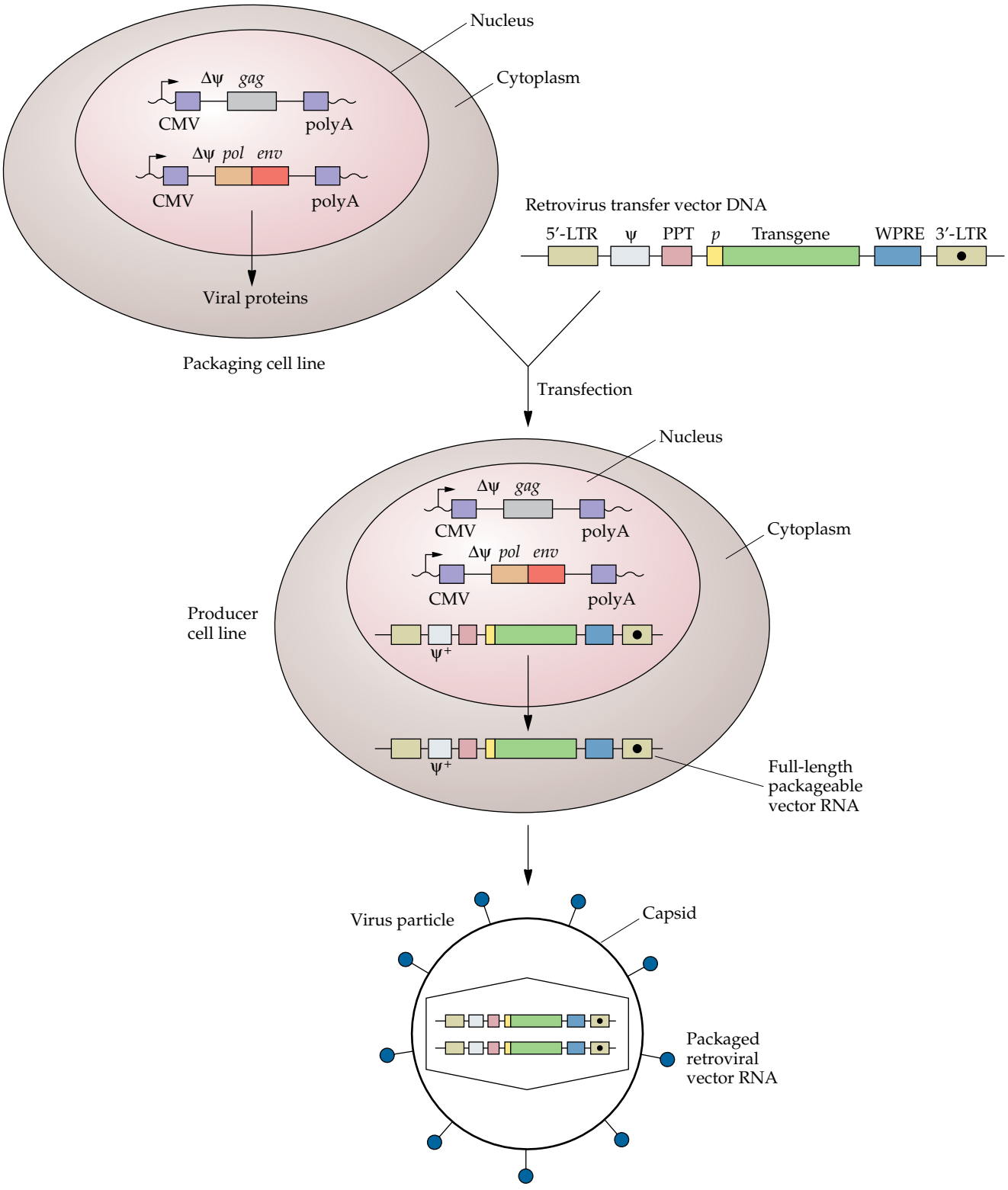


Figure 12.5 Production of packaged retroviral vector RNA. The packaging cell line carries retroviral genes encoding enzymes required for viral genome replication (*pol*), viral capsid and other structural proteins (*gag*), and viral surface glycoproteins (*env*), often on separate plasmids and under the control of a strong constitutive promoter such as the human cytomegalovirus (CMV) promoter. The viral segments lack the encapsidation sequence (ΔΨ) that is required for packaging a retroviral genome into a viral capsid. The packaging cell line synthesizes viral proteins, but in the absence of an encapsidation sequence, empty viral capsids are produced. The packaging cell line is transfected with a retroviral transfer vector carrying a transgene from which vector RNA is produced. The vector RNAs have an encapsidation signal (Ψ^+) and are packaged into viral capsids. The released viral particles are replication defective because they do not have a *pol* gene.

into another blastocyst embryo. Such cells are called pluripotent embryonic stem cells. Cultured embryonic stem cells from some animals, such as mice, can be readily engineered genetically without altering their pluripotency. With this system, for example, a functional transgene that is flanked by specific sequences can be targeted for integration into the genome of embryonic stem cells by homologous recombination. The genetically engineered cells are selected, grown, and used to generate transgenic animals (Fig. 12.6). In this way, the randomness of integration that is inherent in the DNA microinjection and retroviral vector systems is avoided.

The target site for integration of the transgene should be located in a section of genomic DNA that encodes no essential products, so that after integration of the transgene, there is no interference with any developmental or cellular functions. Moreover, it is essential that the transgene be integrated into a part of the genome that does not prevent it from being transcribed, for example, in euchromatin rather than heterochromatin (see chapter 3).

After transfection of embryonic stem cells *in vitro* with a transgene (usually by electroporation) that is designed to integrate within a specific chromosomal location, some cells will have the DNA integrated at nontarget (spurious) sites, whereas in other cells, integration will occur at the target (correct) site. In most of the embryonic stem cells, the transgene will not be integrated at all. To enrich for the cells with the DNA integrated at the target site, a procedure called positive–negative selection is implemented. This strategy uses positive selection for cells that have the introduced DNA integrated anywhere in their genomes and negative selection against the DNA sequence that is integrated at spurious sites.

A targeting DNA vector for the positive–negative selection procedure usually contains (i) two blocks of DNA sequences (HB1 and HB2) that are homologous to separate regions of the target site, (ii) the transgene, which will confer a new function in the recipient, (iii) a DNA sequence that codes for resistance to the compound G-418 (Neo^r), and (iv) two genes for thymidine kinase (*tk1* and *tk2*) from herpes simplex viruses 1 and 2 (HSV-*tk1* and HSV-*tk2*) (Fig. 12.7A). The arrangement of these sequences is key to the positive–negative selection procedure. Between the two blocks of DNA that are homologous to the target site are the genes for the transgene and G-418 resistance (Neo^r gene). Outside each of the homologous blocks are the genes for HSV-*tk1* and HSV-*tk2*. If integration occurs at a spurious site, i.e., not at HB1 and HB2, either one or both of the HSV-*tk* genes have a high probability of being integrated along with the other sequences (Fig. 12.7A). Alternatively, if the integration event is due to homologous recombination by a double crossover at the target site, the HSV-*tk* genes will be excluded, and only the transgene and the Neo^r gene will be incorporated into the genome (Fig. 12.7B). When transfected cells are grown in the presence of G-418, all the cells that lack the Neo^r gene will be killed. Therefore, only cells with integrated DNA will survive; i.e., these cells are positively selected. If the compound ganciclovir is added at the same time as G-418, the cells that express thymidine kinase will be killed because thymidine kinase converts ganciclovir to toxic compounds that kill cells; i.e., these cells are negatively selected. The cells most likely to survive this dual-selection scheme will be those that have DNA integrated at the target site. Although not foolproof, the positive–negative selection method

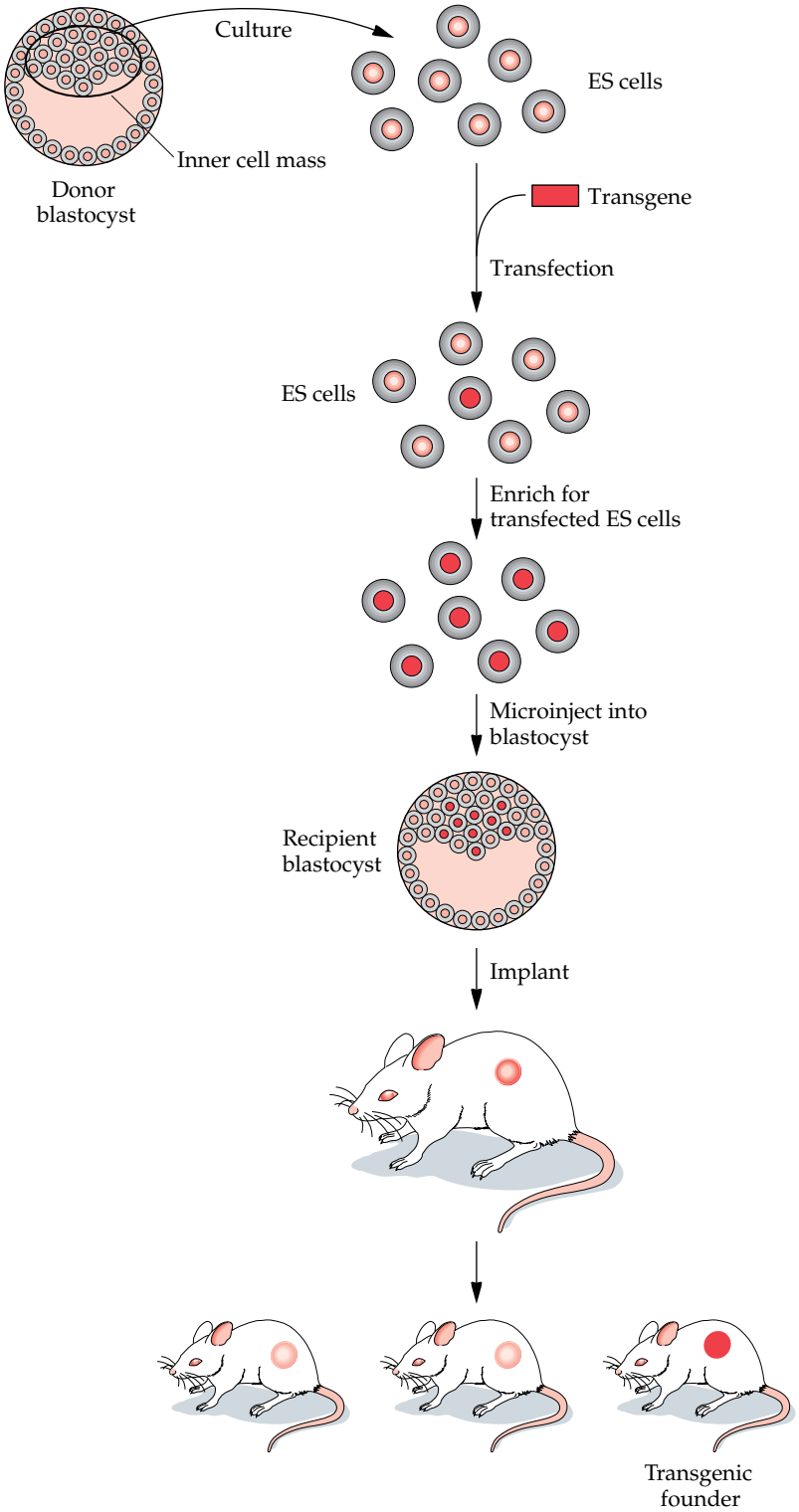


Figure 12.6 Establishing transgenic mice with genetically engineered embryonic stem cells. An embryonic stem (ES) cell culture is initiated from the inner cell mass of a mouse blastocyst. The embryonic stem cells are transfected with a transgene. After growth, the transfected cells carrying the transgene integrated at a suitable site in a chromosome are identified by either the positive–negative selection procedure (Fig. 12.7) or PCR analysis (Fig. 12.8). Populations of transfected cells can be cultured and inserted into blastocysts, which are then implanted into foster mothers. Transgenic lines can be established by crosses from founder mice that carry the transgene in their germ lines.

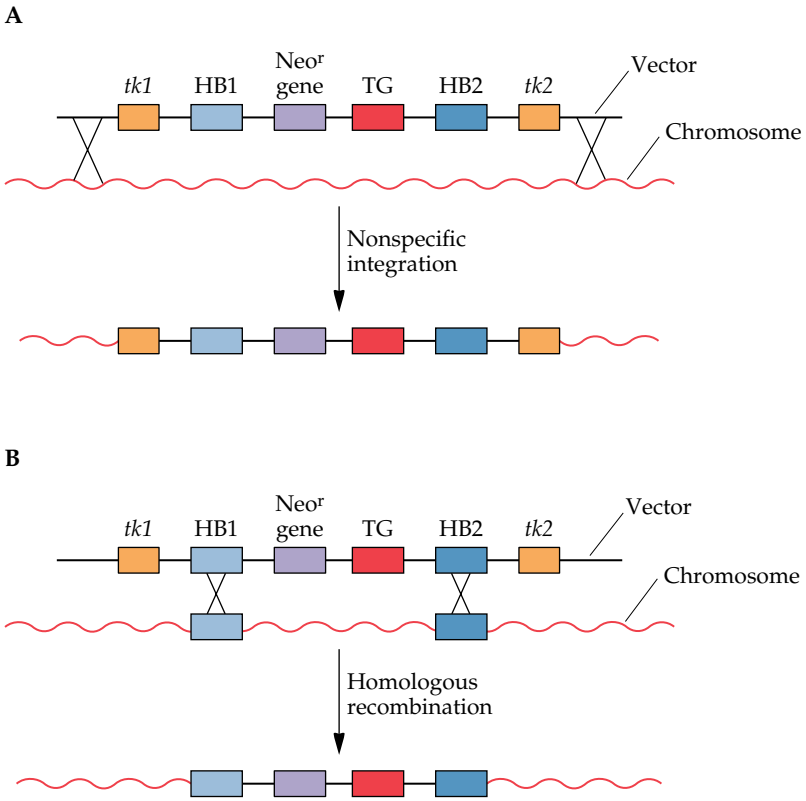


Figure 12.7 Positive–negative selection for chromosomal integration of a transgene. **(A)** Result of nonspecific integration. Both genes for thymidine kinase (*tk1* and *tk2*), the two DNA sequences that are homologous to a specific chromosomal region in the recipient cells (HB1 and HB2), a gene (Neo^r) that confers resistance to the cytotoxic compound G-418, and the transgene (TG) are incorporated into the chromosome. After transfection, cells are selected for resistance to both G-418 and the compound ganciclovir, which becomes cytotoxic to cells that synthesize thymidine kinase. Nonhomologous integrations that insert one and/or the other of the thymidine kinase genes may occur. After treatment with G-418 and ganciclovir, all the cells with nonspecific integration of the transgene DNA construct that includes at least one of the thymidine kinase genes are killed. **(B)** Result of homologous recombination. The product of a double crossover between homologous blocks (HB1 and HB2) of DNA on the vector and in the chromosome does not contain either of the two thymidine kinase genes (*tk1* and *tk2*). After treatment with G-418 and ganciclovir, only cells that have undergone homologous recombination survive.

enriches an embryonic stem cell population for cells that carry a transgene at a specific chromosomal location.

A more direct way to detect embryonic stem cells that carry a transgene at a targeted chromosomal site is to use PCR. The targeting vector contains two DNA sequences that are homologous to the target site, with one on either side of the transgene for site-specific chromosomal integration of the transgene by homologous recombination (Fig. 12.8). After transfection of embryonic stem cells, cells are pooled, and each cell pool is screened by PCR. One of the PCR primers (P1) anneals to a sequence within the introduced DNA, and the other primer (P2) anneals to a sequence adjacent to one of the homologous DNA sequences in the chromosome. If site-specific

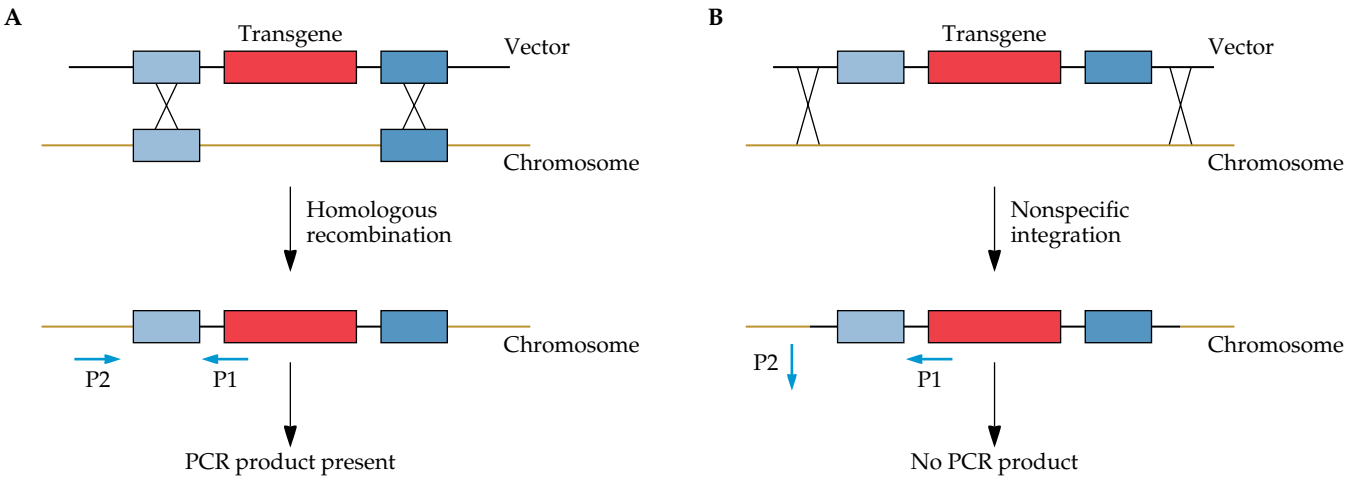


Figure 12.8 Determination of transgene integration at a targeted site in transfected cells by PCR. **(A)** Recombination (×) between homologous DNA sequences (blue) in the transgene construct on the vector and in the chromosome results in integration of the transgene into a specific site in the chromosome of the recipient cell. Primer P1 anneals to a sequence within the transgene construct that does not occur in the chromosome of the recipient cell. Primer P2 anneals to a sequence adjacent to the expected integration site in the recipient chromosome. Amplification by PCR generates a DNA fragment of a specific size that can be visualized by gel electrophoresis. **(B)** If nonspecific integration of the vector DNA occurs, primer P2 is not able to anneal to a chromosomal site that is a predetermined distance from the annealing site for primer P1, so a DNA fragment of a specific size is not amplified.

integration occurred, the PCR amplifies a DNA fragment of the predicted size (Fig. 12.8A). However, if integration occurred at a random site, the predicted PCR product will not be synthesized (Fig. 12.8B). In this way, pools of embryonic stem cells with some cells containing the transgene at the targeted site can be identified. By subculturing from these pools, cell lines carrying the site-specific integration can be established.

Embryonic stem cells carrying an integrated transgene are cultured and inserted into blastocyst-stage embryos, and these embryos are then implanted into pseudopregnant foster mothers. Transgenic lines are established by mating the progeny that carry the transgene in their germ lines. Then, if required, animals that carry a transgene in their germ lines are crossed to produce animals that are homozygous for the transgene.

Not only can a foreign gene be inserted into a specific chromosome site by homologous recombination in embryonic stem cells to provide a new function, but also, a specific gene can be targeted for disruption by the incorporation of a DNA sequence, usually a selectable marker gene, into its coding region (Fig. 12.9). One of the aims of targeted gene disruption (knockout) is to determine the developmental and physiological consequences of inactivating a particular gene. In addition, a transgenic line with a specific disabled gene can be used as a model system to study the molecular pathology of a human disease. For example, inactivation of the mouse rhodopsin gene by targeted gene disruption leads to deterioration of the rod cells of the retina in transgenic mice that closely resembles the human disease retinitis pigmentosa. Thus, the progress of retinal degeneration and the effects of potential therapeutic agents that either delay or

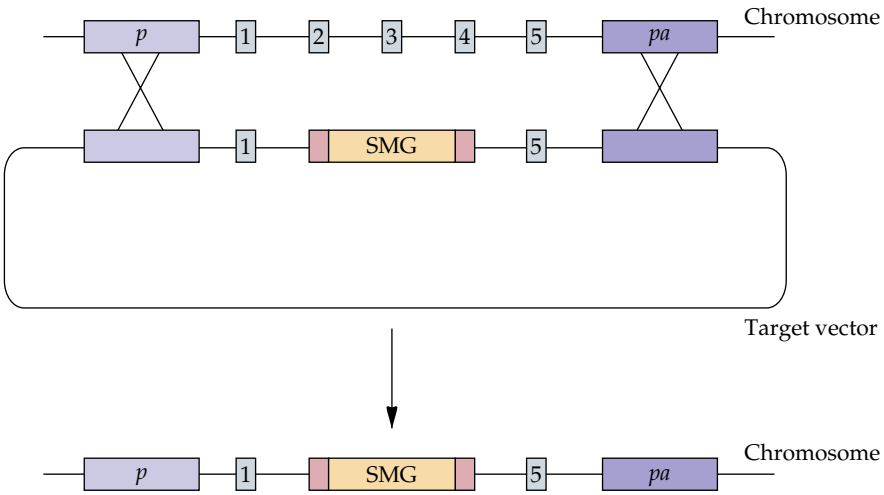


Figure 12.9 Gene disruption by targeted homologous recombination. The target vector carries a selectable marker gene (SMG) with flanking DNA sequences that are homologous to regions of the targeted gene. In this example, the targeted gene has five exons (1 to 5). Homologous recombination (×) disrupts (i.e., knocks out) the targeted gene. *p*, promoter; *pa*, polyadenylation signal.

block the genetically induced retinopathy are studied using the rhodopsin knockout mouse. Hundreds of different types of knockout mice have been created as animal models for the study of various human abnormalities.

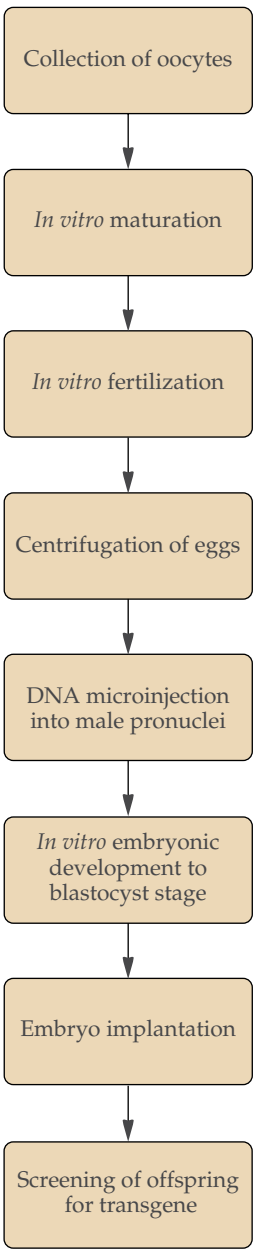
Somatic Cell Nuclear Transfer for Transgenic Livestock

There are several reasons for producing transgenic livestock. First, although mice have proven to be useful in biomedical research as models of human diseases and for testing of disease treatments, the anatomy, lifespan, and some aspects of the physiology of a mouse are different from those of humans. On the other hand, livestock, especially pigs, are often better animals in which to model disease processes, gene regulation, and immune system development. Second, many livestock animals produce large amounts of milk and therefore can be used as a means to produce and secrete large amounts of recombinant proteins and other molecules of pharmaceutical importance. Third, genetic engineering can be employed to rapidly and specifically improve livestock traits such as growth rate, disease resistance, and milk quality.

Although pronuclear DNA microinjection was the primary method for livestock transgenesis for many years (Fig. 12.10), it is labor-intensive and inefficient. The yolk in fertilized eggs prevents the male pronuclei from being readily seen under a dissecting microscope, and the low probability of integration after DNA microinjection contributes to poor yields of transgenic animals. In addition, transgene integration occurs at random sites, expression is often constrained because of the chromatin structure at the site of integration, unstable tandem arrays are formed, and the establishment of a transgenic line requires a number of generations. For these reasons, transgenic livestock are now more commonly generated by transferring nuclei from genetically engineered somatic cells into oocytes.

In a highly publicized case in 1997, a sheep named Dolly was cloned by transfer of the nucleus from a mammary (udder) cell of an adult sheep into

Figure 12.10 Steps in the development of transgenic livestock.



an enucleated sheep egg cell (Fig. 12.11). This was the first demonstration of pluripotency (totipotency) of a nucleus of a differentiated adult cell. Since the cloning of Dolly, somatic cell nuclei have been used to clone a variety of livestock, including cattle, goats, sheep, and pigs, as well as other mammals, amphibians, fish, and insects. For nuclear transfer, embryonic, fetal, or adult donor cells from a variety of cell types (e.g., mammary epithelial, ovarian, fibroblast, and lymphocyte) are isolated, cultured, and genetically modified using the methods described above (Fig. 12.11). Although not always feasible with adult cells, prolonged culture is preferred because experimenters have additional time to carry out successive genetic alterations, such as inactivating both alleles of a locus or creating multiple gene changes. After establishing a cell line with a specific genetic modification(s), individual donor cells are fused to an enucleated oocyte with short-duration electric pulses. For example, two 2.5-kilovolt-per-centimeter (kV/cm) pulses for 10 microseconds each are used to fuse adult cattle fibroblast cells with enucleated oocytes. The pulses simultaneously induce cell fusion and oocyte activation. Each fused cell is cultured to the blastocyst stage before being transferred into the uterus of a pseudopregnant female. At birth, genotype analysis is used to confirm the presence of the transgene.

With somatic cell nuclear transfer, there is substantial loss of individuals before and after birth, and some of the cloned animals display abnormalities. Abnormalities such as increased birth weight are more prevalent in some livestock animals, for example, in cloned calves and lambs, than in others. One reason for this is the failure of the somatic cell donor genome to undergo epigenetic reprogramming; the pattern of DNA methylation and histone modification of the original donor cell is inappropriately maintained in the cells of the recipient animal. Because the epigenetic state of DNA controls gene expression, this could seriously impair cellular function. However, some animals that are produced in this way are healthy, develop normally, and are able to reproduce.

Nuclear transfer has some important advantages over pronuclear DNA microinjection and therefore is the preferred method for generating transgenic livestock. With nuclear transfer, site-specific genetic changes are possible, all offspring are transgenic, and small herds can be produced within a short period of time.

Genome Editing with the CRISPR-Cas System

Transgenesis by the DNA microinjection and retroviral vector methods results in integration of the transgene at a random site in an animal's genome. A transgene may be targeted to a specific site in the genome of an embryonic stem cell by homologous recombination and site-specific integration confirmed prior to implantation; however, generation of a transgenic animal from isolated embryonic stem cells is time-consuming and costly (about 1 year and up to $100,000 for a transgenic mouse), and germ line transmission to the next generation is not always achieved. Moreover, transgenesis using embryonic stem cells has failed for some animals, such as cattle and pigs, because the pluripotent state of these cells is unstable *in vitro*. Recently, it has become possible to modify the genomes of single-cell embryos directly using the clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated endonuclease (Cas) system (see chapter 2). Although it is still being optimized,

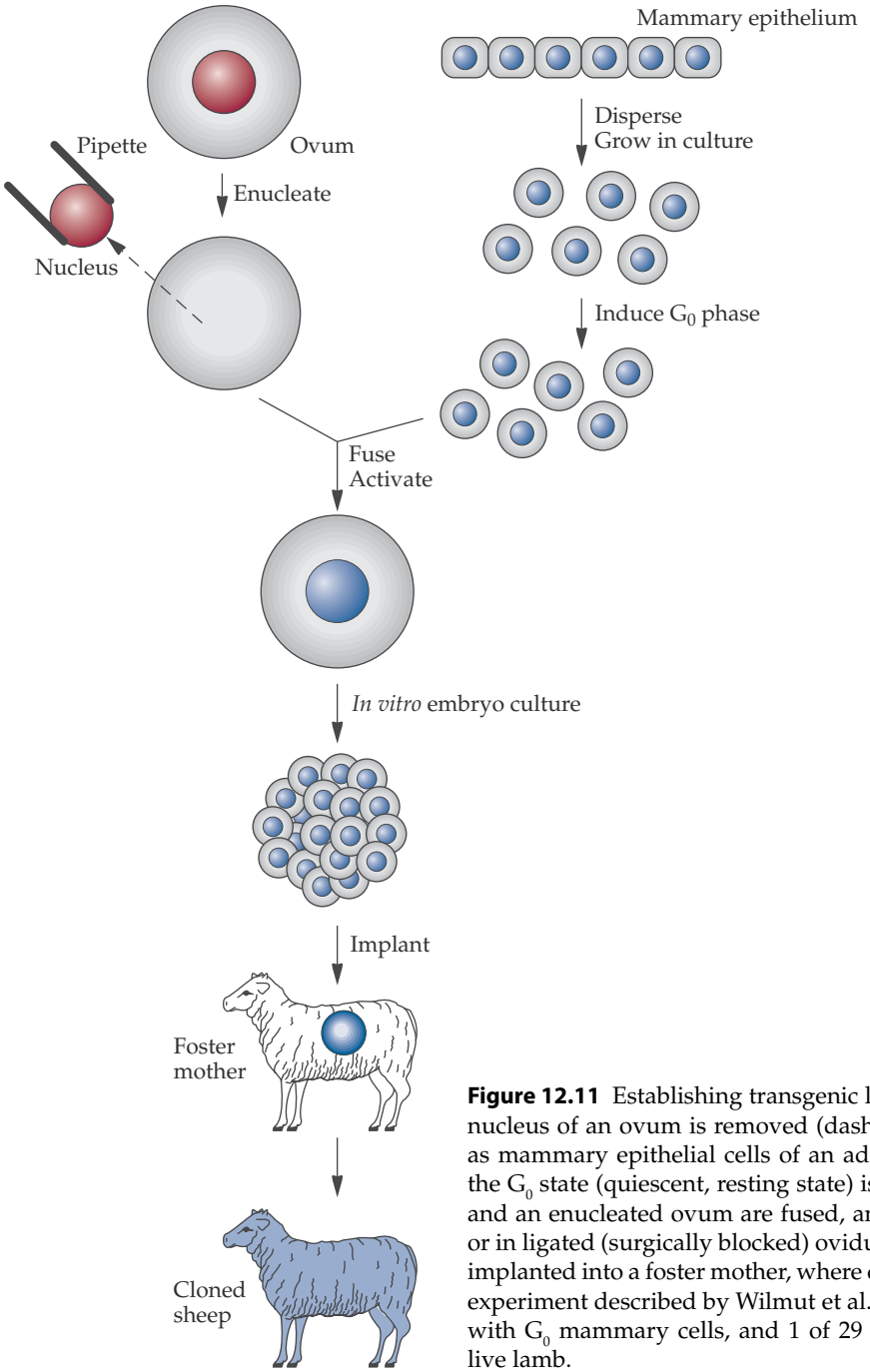


Figure 12.11 Establishing transgenic livestock by somatic cell nuclear transfer. The nucleus of an ovum is removed (dashed arrow) with a pipette. Somatic cells such as mammary epithelial cells of an adult female animal are grown in culture, and the G_0 state (quiescent, resting state) is induced by inhibiting cell growth. A G_0 cell and an enucleated ovum are fused, and the renucleated ovum is grown in culture or in ligated (surgically blocked) oviducts until an early embryonic stage before it is implanted into a foster mother, where development proceeds to term. In the original experiment described by Wilmut et al. (1997), 277 enucleated sheep ova were fused with G_0 mammary cells, and 1 of 29 transferred early-stage embryos produced a live lamb.

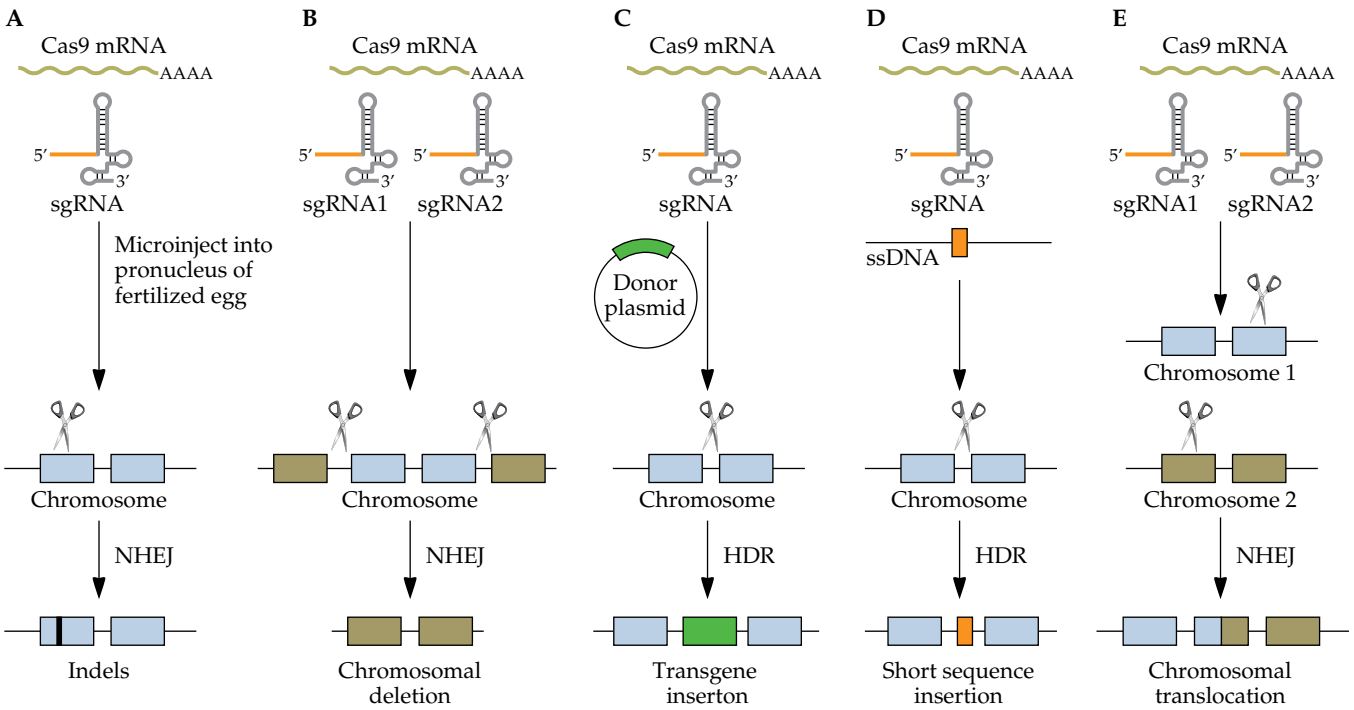
the CRISPR-Cas system has been used to generate insertion and deletion mutations and to introduce transgenes at targeted sites in the genomes of many animals.

To disrupt a gene, Cas9 endonuclease (either the protein or Cas9 mRNA which is translated to produce the protein in the host cell) and single-guide RNA (sgRNA) are microinjected into the pronucleus of a fertilized egg. The sgRNA includes a short (20-nucleotide) sequence that directs Cas9 to cleave the animal's DNA at a complementary sequence in the target gene that is

immediately adjacent to a 3-bp (NGG) protospacer-adjacent motif (PAM) (chapter 2, Fig. 2.18). The ends of the double-stranded break in the genomic DNA, generated by Cas9 cleavage, may be re-ligated by nonhomologous end joining. This is an error-prone repair process that often results in insertion and/or deletion of several nucleotides (indels) that may disrupt the reading frame of a targeted protein-coding gene (Fig. 12.12A). Large segments of the mouse genome (up to 200 kb) have also been deleted using two sgRNAs, each with complementarity to a sequence that delineates the region to be excised (Fig. 12.12B).

To generate a transgenic animal, a transgene is introduced on a donor plasmid, along with Cas9 mRNA and sgRNA, into the pronucleus of a fertilized egg by microinjection. The transgene is flanked by sequences that are homologous to the target site for integration into a chromosome. A double-stranded break generated by Cas9 at the targeted site specified by the sgRNA may be repaired by homologous recombination between the transgene flanking sequences and the sequences at the cut ends of the chromosome, which results in incorporation of the transgene

Figure 12.12 Editing animal genomes using the CRISPR-Cas system. The Cas9 endonuclease, expressed from microinjected Cas9 mRNA, is guided to target sites in chromosomal DNA by single-guide RNA (sgRNA), where it generates double-stranded breaks. Error-prone repair of the cleaved DNA by nonhomologous end-joining (NHEJ) may result in insertion and/or deletion of nucleotides (indels) **(A)** or, if the DNA is cleaved at two different sites in the same chromosome, a large deletion **(B)**. The cleaved DNA may also be repaired by homologous recombination (homology-directed repair, HDR) with donor plasmid DNA, resulting in insertion of a transgene **(C)**, or a with a single-stranded oligonucleotide, resulting in insertion of a short tag sequence **(D)**. Chromosomal translocations may result from joining of cut ends from two different chromosomes **(E)**. Data from Williams et al., 2016 Feb 1, *Cold Spring Harb. Protoc.* 2016:pdb.top087536.



(Fig. 12.12C). This method has also been used to insert short sequences such as recombinase recognition sites or affinity tags (see chapter 3) into a genome. For small insertions, the donor DNA is usually a single-stranded oligonucleotide containing the targeted modification flanked by short (50 to 60 nucleotides) sequences for homologous recombination (Fig. 12.12D).

An example of the application of the CRISPR-Cas technology in mice is the construction of mouse models of human cancers. Genomic studies have revealed the immense complexity of cancers, with cancerous cells carrying hundreds of point mutations, insertions, deletions, and inversions that are not present in noncancerous cells, and with the nature of the mutations varying among cancer types. To generate mouse models in which to study the significance of these mutations, and to validate the involvement of specific cancer genes, would be an enormous undertaking using traditional methods that rely on homologous recombination of donor DNA in embryonic stem cells. In contrast, short sgRNAs that target many different genomic sites can readily be produced by cloning synthetic double-stranded oligonucleotides carrying the 20-nucleotide target-specific sequence into an expression vector encoding the remainder of the sgRNA sequence (Fig. 12.13). Microinjection of the sgRNA as a ribonucleoprotein complex with Cas9 into one-cell embryos results in rapid generation of many different mouse models with cancer mutations. For example, by using two sgRNAs that guide Cas9 to cleave two different chromosomes, chromosomal translocations found in some cancer cells have been generated in mice (Fig. 12.12E). This approach is much less costly and labor-intensive than using the Cre-*loxP* system (described below) to mediate recombination between two chromosomes in embryonic stem cells.

One of the limitations of CRISPR-Cas technology is the potential for mutagenesis at unintended sites. Cleavage at "off-target" sites may be reduced by ensuring that the 20-nucleotide targeting site of the sgRNA has at least three mismatched bases to any other sequence in the genome or by using sgRNA with shorter (as few as 17 nucleotides) target sequences (presumably because they are more sensitive to mismatches). Cas endonucleases with enhanced specificity for cleavage sites have also been engineered.

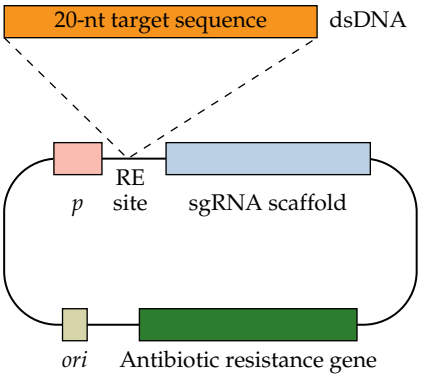


Figure 12.13 An example of a cloning vector for production of single-guide RNA (sgRNA). Double-stranded synthetic DNA carrying the 20-nucleotide target sequence is ligated with a cloning vector that has been linearized using a restriction enzyme (RE). The vector carries the coding sequence for the region of sgRNA (scaffold) that binds to the Cas9 endonuclease. The full-length sgRNA may be expressed from a strong promoter (*p*) such as the T7 promoter using an *in vitro* transcription system, which facilitates purification of the sgRNA for subsequent microinjection into animal cells. An origin of replication (*ori*) and antibiotic resistance gene are included for propagation and selection of the vector in a bacterial host cell.

Conditional Gene Modification with the Cre–*loxP* Recombination System

Transgenic animals usually carry a transgene in all of their cells. However, it may be useful to selectively express an introduced gene in a specific cell type or during a specific developmental stage of a transgenic animal. This is particularly useful when the product encoded by the transgene is lethal in the embryo or in juveniles. The Cre–*loxP* recombination system, which is derived from genetic elements of bacteriophage P1, may be used for this purpose.

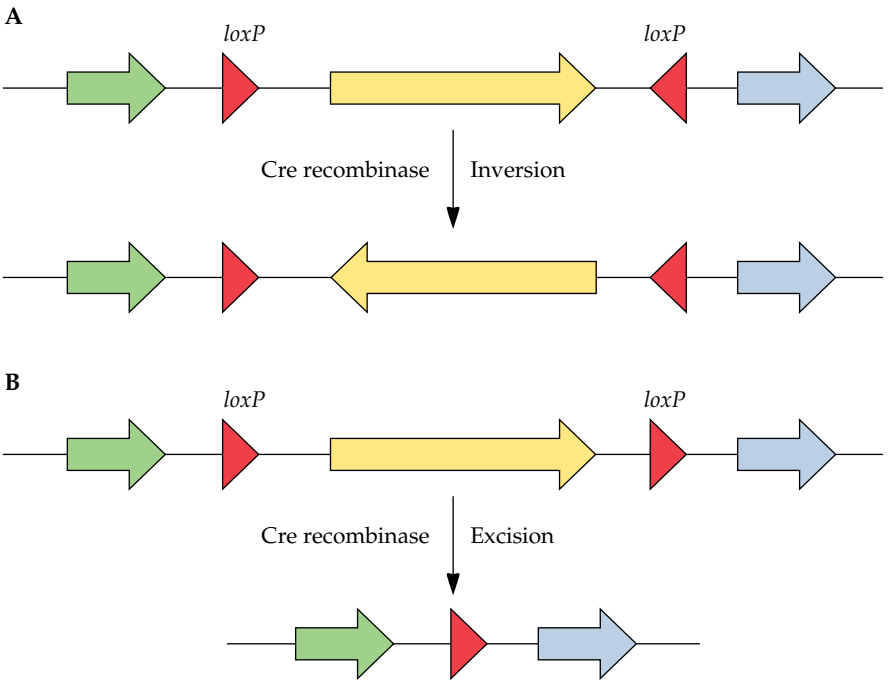
Bacteriophage P1 is one of several tailed viruses that infects the bacterium *Escherichia coli*. It has a double-stranded DNA genome that is about 100 kb long. After injection through the tail into *E. coli*, the linear P1 genome forms a circle. The circularized P1 DNA is a template for replication and, depending on which set of genes is activated, the circular form is either maintained as a plasmid in the host cell or serves as a template for the production of viral genomes that will be packed into new virions during the lytic cycle.

Figure 12.14 A *loxP* site consists of two 13-bp inverted repeat sequences flanking an 8-bp spacer sequence.

On rare occasions, a circularized P1 genome integrates into the *E. coli* chromosome. Circularization and integration of the P1 genome are mediated by the product of the *cre* gene (circularization recombination gene), Cre recombinase, which specifically cleaves and recombines the DNA of *loxP* (locus of crossing over [x] in P1) sites.

The Cre–*loxP* recombination system may be used to excise, invert, or translocate sequences at specific sites in an animal genome. A *loxP* site consists of two 13-base pair (bp) inverted repeats that are separated from each other by an 8-bp spacer sequence (Fig. 12.14). A Cre dimer binds at each *loxP* site (one Cre monomer binds each inverted repeat), and formation of a tetrameric Cre complex brings two *loxP* sites together. Cre cleaves within the spacer regions between the inverted repeat sequences, and the DNA strands are joined to form recombined DNA molecules. The directionality of a *loxP* site is determined by the asymmetry of the spacer sequence, and the outcome of the recombination event depends on the relative orientation of the two *loxP* sites (Fig. 12.15). If the *loxP* sites are oriented in opposite directions, then the exchange inverts the DNA between the two

Figure 12.15 The outcome of Cre recombinase-mediated recombination depends on the relative orientation of the *loxP* sites. **(A)** Recombination between *loxP* sites in the opposite orientation results in inversion of the DNA sequence between the two *loxP* sites. **(B)** Recombination between *loxP* sites in the same orientation results in deletion of the DNA sequence between the two *loxP* sites.



sites (Fig. 12.15A). If the *loxP* sites are in the same orientation, then the intervening sequence is excised (Fig. 12.15B). The Cre–*loxP* recombination system can function when the *loxP* sites are widely separated. For example, the *loxP* sites that are essential for circularization of a P1 genome are about 100 kilobases (kb) apart, at each end of the linear genome. The specificity of the Cre–*loxP* system is absolute because Cre acts exclusively on *loxP* sites.

As a first step in producing cell-specific gene modifications in a mouse, the *cre* gene is placed under the control of a cell-specific promoter. A mouse with the *cre* gene construct is created, and the tissue specificity of Cre expression is confirmed. Numerous transgenic mouse lines that express Cre from tissue-specific or inducible promoters have been established. To control expression of a transgene, a transcription terminator sequence flanked by *loxP* sites oriented in the same direction is inserted between the promoter and the transgene such that it prevents transcription (Fig. 12.16). The transcription terminator sequence with flanking *loxP* sites may be included with the transgene on a vector and integrated into a targeted chromosome site of embryonic stem cells by homologous recombination. The embryonic stem cells carrying the integrated construct are selected, cultured, and used to establish a transgenic mouse line. A transgenic mouse with a tissue-specific *cre* gene is mated with a transgenic mouse with the integrated *loxP*-flanked transcription terminator and transgene sequences to generate double transgenic mice. In cells of these mice that express Cre, the transcription terminator is excised and the transgene is expressed (Fig. 12.16).

The Cre–*loxP* system may also be used to inactivate an endogenous gene in specific tissues. For this, a DNA sequence that is targeted for excision is flanked by *loxP* sites that are oriented in the same direction (Fig. 12.17). The sequence to be excised may be an entire gene targeted for inactivation or an essential exon, and the *loxP* sites must be inserted in the flanking regions (e.g., introns) such that they do not disrupt expression of the gene. The *loxP* sites are inserted on a cloned DNA fragment, and then the construct is integrated in the chromosome of an embryonic stem cell by homologous recombination replacing the original sequence. In transgenic mice carrying both the *loxP*-flanked target sequence and the *cre* gene, the DNA between the two *loxP* sites is deleted only in cells in which Cre is expressed (Fig. 12.17). The Cre–*loxP* technology has been used extensively to study the biological consequences of tissue-specific gene inactivation with the goal of establishing models for human diseases. For example, selective removal of the kinesin-II gene, which is expressed exclusively in retinal photoreceptor cells, leads to an accumulation of the proteins opsin and arrestin and eventually to cell death. This result mimics aspects of inherited retinitis pigmentosa in humans and is used for detailed studies of the pathophysiological effects on the retina.

Control of Transgene Expression with the Tetracycline-Inducible System

Various protocols have been devised for turning on and off the expression of a transgene in a specific cell type at will. Of these methods, the tetracycline-inducible system has been used extensively. In one form of the tetracycline-inducible system, the addition of doxycycline, an analogue of the antibiotic tetracycline, turns off the expression of a transgene. Without doxycycline, the transgene is continuously expressed in a specific

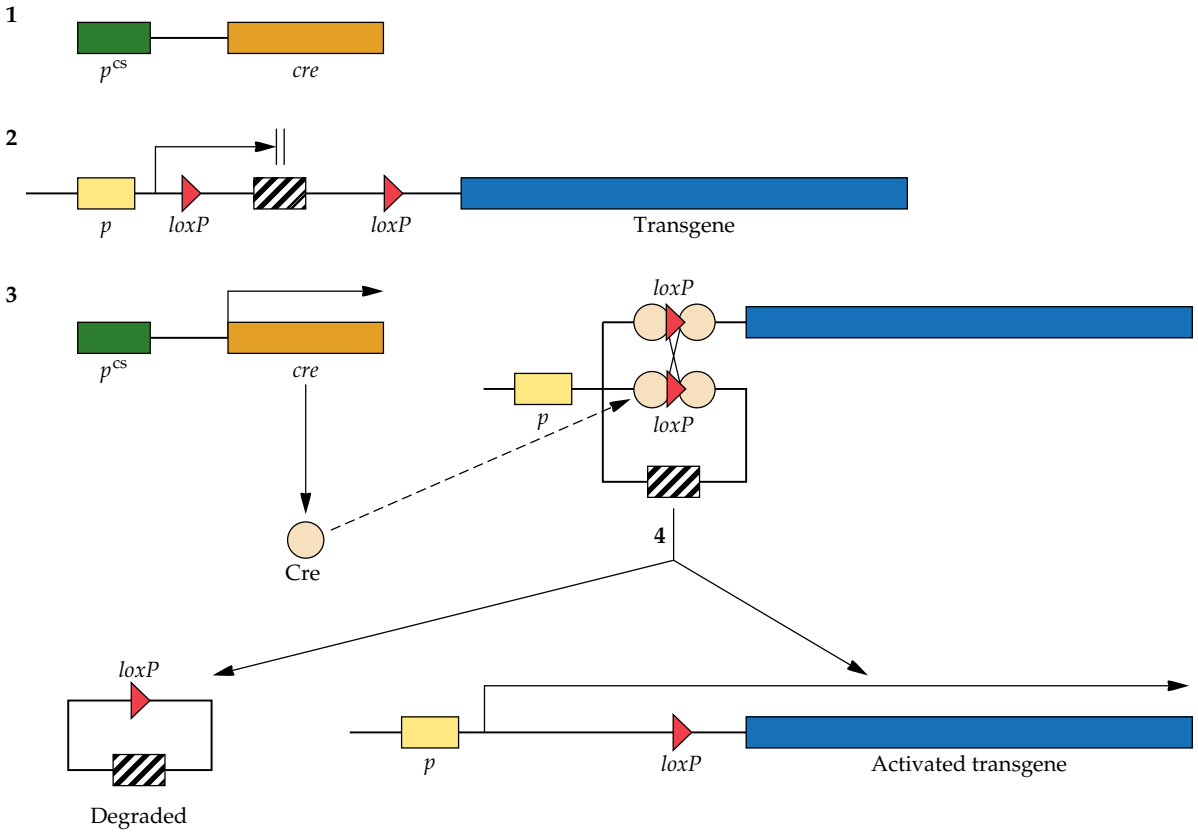


Figure 12.16 Cre–*loxP* recombination system for activating a transgene in a specified cell type. **(1)** The *cre* gene is placed under the control of a cell-specific promoter (p^{cs}) and established as a transgene in a line of mice. **(2)** A piece of DNA with *loxP* sites that flank a transcription termination sequence (hatched box) is inserted between a promoter (*p*) and a transgene. **(3)** The transgene construct with the *loxP* sites is introduced into a chromosomal site of embryonic stem cells by homologous recombination, and a transgenic mouse line is established with these cells. The two transgenic lines are crossed. In cells where both constructs are present, the Cre recombinase (beige circle) is synthesized, and two Cre molecules bind to each *loxP* site (dotted arrow). **(4)** The *loxP* sites undergo recombination (×), leading to the excision of a *loxP* site and the transcription termination sequence on a DNA fragment that is circularized and eventually degraded, and the formation of a transcriptionally active transgene that is retained in the chromosome. Right-angled arrows denote transcription, and parallel vertical bars indicate termination of transcription.

cell type. This "tet-off" system depends on the production of a chimeric (hybrid) protein composed of the tetracycline repressor and a transcription factor that activates the transcription process (Fig. 12.18A). The hybrid protein is called tetracycline-controlled transactivator (tetracycline transactivator, or tTA). The gene (*tTA*) that encodes tTA is under the control of a cell-specific promoter. The promoter that drives the transgene consists of a set of tetracycline operator (*tetO*) sequences upstream from a strong eukaryotic promoter. In the absence of doxycycline, the tTA protein binds to the *tetO* region and activates the transcription of the transgene. However, when doxycycline is present, it binds to the tTA protein, and the doxycycline–tTA complex cannot bind to the *tetO*–promoter sequence, and transcription of

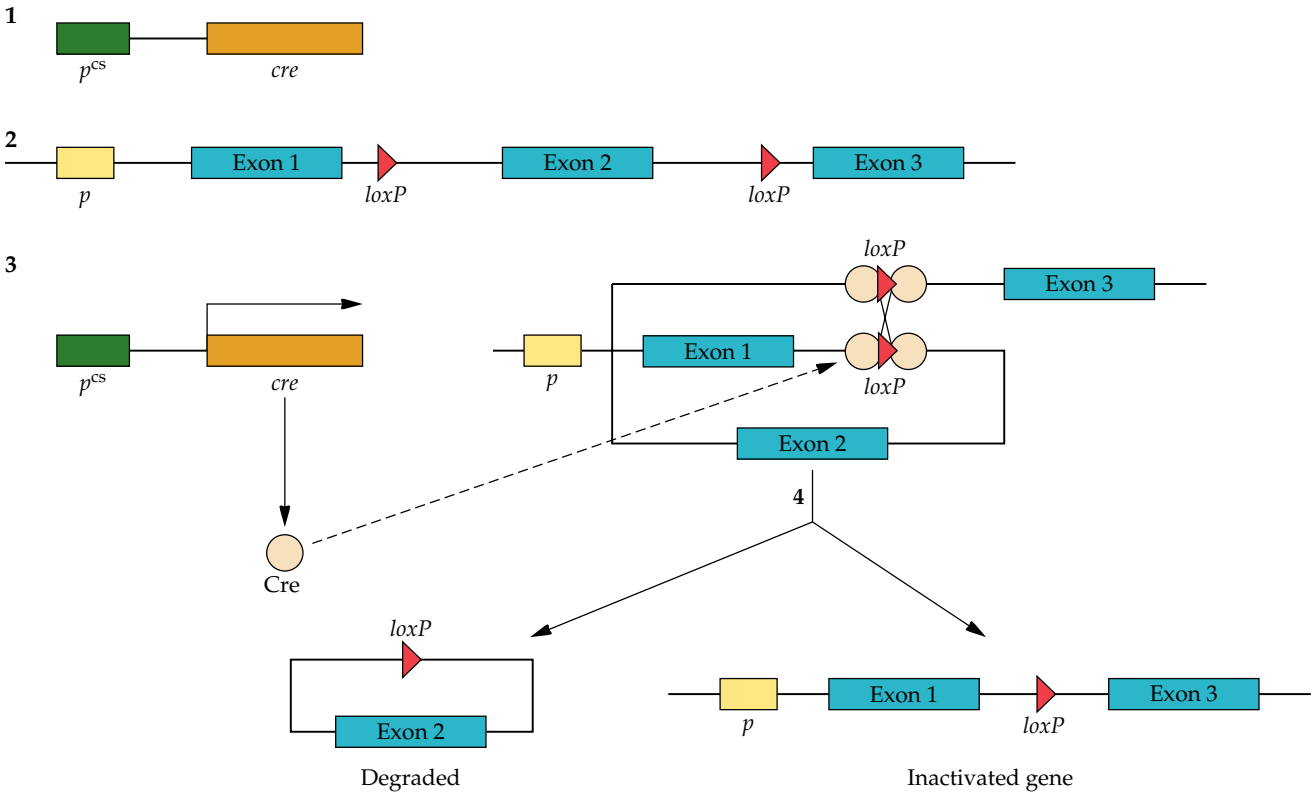


Figure 12.17 Cre–*loxP* recombination system for inactivating a gene in a specified cell type. **(1)** The *cre* gene is placed under the control of a cell-specific promoter (p^{cs}) and established as a transgene in a line of mice. **(2)** A *loxP* site is cloned on either side of an exon. The construct with the *loxP* sites is introduced into a chromosomal site of embryonic stem cells by homologous recombination, and a transgenic mouse line is established with these cells. The two transgenic lines are crossed. **(3)** In cells where both constructs are present, the Cre recombinase (beige circle) is synthesized, and two Cre molecules bind to each *loxP* site (dotted arrow). **(4)** The *loxP* sites undergo recombination (×), leading to the excision of a *loxP* site and an exon (exon 2) on a DNA fragment that is circularized and eventually degraded, and the formation of an inactivated gene that is retained in the chromosome. A right-angled arrow denotes transcription.

the transgene does not occur. Thus, the presence or absence of doxycycline, which is administered by adding it to the animals' drinking water, acts as a switch in cells in which the promoter of the *tTA* gene is active.

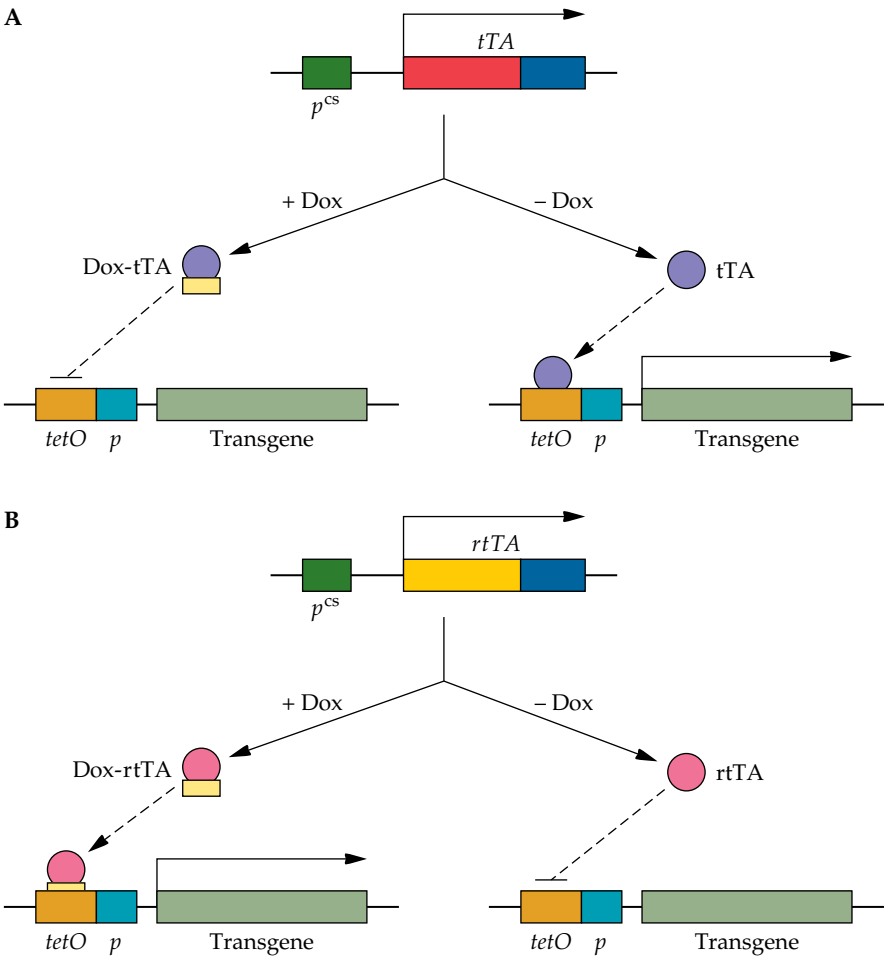
A reverse system called "tet-on," in which doxycycline must be present for transcription of the transgene, has also been devised (Fig. 12.18B). In this form, the nucleotide sequence for the tetracycline repressor carries mutations that prevent the combined repressor protein and transactivator from binding to the *tetO*–promoter sequence. This tetracycline repressor/transactivator protein is designated rtTA (reverse tetracycline-controlled transactivator). Doxycycline binds to rtTA and changes its configuration, which allows the complex to attach to the *tetO*–promoter sequence and initiate transcription of the transgene.

The tet-off and tet-on systems have innumerable uses. For example, the biological consequences of the production of a defective protein or overexpression of a normal protein can be examined in detail, cell-specific disease

conditions can be simulated, and gene-based treatments for diseases that affect a particular cell type can be tested. However, tetracyclines, including doxycycline, inhibit translation of genes in mitochondria, which reflects their evolutionary relationship with prokaryotes, and the potential impact of this on an experimental system must be considered.

A fascinating example of the utility of the tetracycline-regulatory system is the development of a mouse model for Huntington's disease, a disease

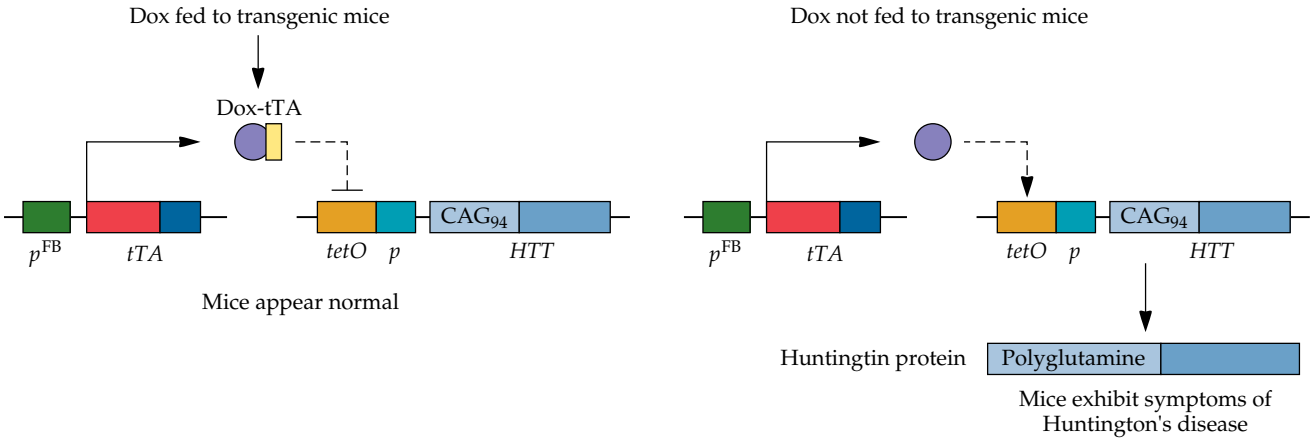
Figure 12.18 Tetracycline-regulated transgene expression. **(A)** Tet-off system. The *tTA* gene is driven by a cell-specific promoter (p^{cs}) and encodes a tetracycline repressor (red) and a transcription activator (blue). The product of the *tTA* gene is a hybrid protein (tTA [purple circle]) that binds to a tetracycline operator (*tetO*)–eukaryotic promoter (*p*) sequence in the absence of doxycycline (–Dox) and activates the transcription of the transgene. When doxycycline is present (+Dox), it (yellow rectangle) binds to tTA, preventing tTA from binding to the *tetO*–promoter region, and the transgene is not transcribed. **(B)** Tet-on system. The *rtTA* gene is driven by a cell-specific promoter (p^{cs}) and encodes a mutated tetracycline repressor (dark yellow) and a transcription activator (blue). The product of the *rtTA* gene is a hybrid protein (rtTA [pink circle]) that does not bind to a tetracycline operator (*tetO*)–eukaryotic promoter (*p*) sequence, and in the absence of doxycycline (–Dox), the transgene is not transcribed. When doxycycline is present (+Dox), it (yellow rectangle) binds to rtTA, and the Dox–rtTA complex binds to the *tetO*–promoter region and initiates the transcription of the transgene.



that normally occurs only in humans. This incurable, fatal neurological disorder affects about 1 in 10,000 people worldwide. The symptoms in most cases become evident when the patient is about 45 years old. Initially, muscle coordination is impaired. The disorder is progressive and, eventually, both voluntary and involuntary movements become uncontrolled, speech is slurred, and severe psychiatric conditions appear. In the late stage of Huntington's disease, the patients are mute, cognitively nonfunctional, and immobilized. The neurological damage is confined to specific regions of the brain. At the genetic level, the alteration that is responsible for Huntington's disease is the addition of CAG units to an existing sequential array of these trinucleotides in exon 1 of the *HTT* gene that encodes huntingtin protein. A CAG trinucleotide is the codon for glutamine, and during translation a contiguous set of these codons produces a string of glutamines (polyglutamine) in the huntingtin protein. This causes the huntingtin protein to misfold and form aggregates that disrupt the function of neurons in the brain. Symptoms of Huntington's disease occur when the polyglutamine-coding segment has 38 or more CAG codons (see chapter 6).

To create a mouse model for Huntington's disease, the tet-off system was used with a variant of the human *HTT* gene that consists only of exon 1 with 94 CAG repeats as the transgene (Fig. 12.19). The *tTA* gene was placed under the control of a promoter that is active in the cells of the forebrain. Loss of embryos was avoided during pregnancy by adding doxycycline to the drinking water, which turned off the expression of the mutant *HTT* gene. At birth, doxycycline was not supplied to the transgenic mice, which allowed continuous expression of the mutant *HTT* gene and the production of a protein with a long polyglutamine sequence. A neurological condition that is similar to Huntington's disease in humans developed over time in these mice. Interestingly, the features of the disease disappeared when the expression of the mutant *HTT* gene was prevented by the addition of doxycycline. Thus, at least in this model, continuous expression of a mutant *HTT* gene is required for establishment of the disease, and brain cells can recover when this synthesis ceases.

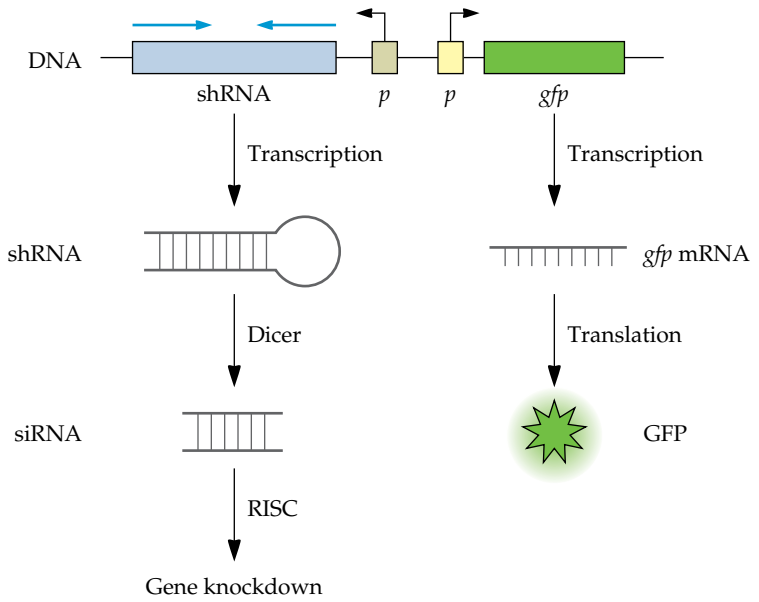
Figure 12.19 A transgenic mouse model of Huntington's disease carrying a mutant form of the *HTT* gene encoding huntingtin protein expressed under the control of the tet-off system. CAG_{94} indicates a sequence of 94 CAG repeats in exon 1 of the mutant *HTT* transgene that encode polyglutamine. p^{FB}, forebrain specific promoter; tTA, tetracycline transactivator; *tetO*, tetracycline operator.



Gene Knockdown by RNA Interference

There are two main approaches to abolish expression of an endogenous gene in animal cells. One approach disrupts a target gene (knockout method) in all cells of an animal using the CRISPR-Cas system in single-cell embryos or homologous recombination in embryonic stem cells, or in specific cells using the Cre–*loxP* recombination system, as described above. The other decreases expression of a target gene (knockdown method) by preventing mRNA translation using RNA interference (RNAi). This latter method exploits a natural mechanism for regulation of gene expression by endogenous RNA molecules in animals and plants and for protecting cells against exogenous RNA molecules from invading viruses (see chapter 6). In RNAi, double-stranded RNA is recognized by a ribonuclease called Dicer that cleaves the RNA into smaller double-stranded RNA molecules known as small interfering RNA (siRNA) (Fig. 6.14). A large nuclease complex called RISC (RNA-induced silencing complex) separates the strands of siRNA, and the single-stranded RNA products, together with RISC, bind to complementary sequences on mRNA molecules. The nuclease component of RISC then degrades the mRNA, which prevents the encoded protein from being synthesized. Short endogenous RNAs (microRNA, abbreviated miRNA) transcribed from regions of the genomes of animals (and plants) are also recognized by Dicer and RISC and block translation of a target mRNA.

Figure 12.20 RNA interference to knock down expression of a target gene in transgenic animals. A transgenic construct encoding a small hairpin RNA (shRNA) to target specific mRNA for degradation and a green fluorescent protein marker gene (*gfp*), each under the control of their own promoters (*p*; direction of transcription is indicated by black arrows), is introduced into early-stage embryos to generate transgenic animals that are identified by production of green fluorescent protein (GFP). Transcription of the transgene encoding short inverted repeats (blue arrows) separated by a spacer sequence yields an shRNA that is processed by the host cell Dicer and RISC proteins to reduce expression of a target gene.



To create transgenic animals with reduced expression of a target gene, a small region of the target sequence is cloned into a vector as an inverted repeat separated by a short spacer region. The RNA transcript that is produced from this sequence will form a short (19 to 21 bp) hairpin RNA (shRNA) due to intramolecular base pairing (Fig. 12.20). Shorter sequences are used to avoid a general downregulation of translation that is often elicited with longer sequences (thought to occur as part of the viral defense response). The construct is introduced into early-stage embryos by one of the methods described above to create a transgenic animal. The shRNA is transcribed and processed by the animal's Dicer and RISC proteins into siRNA to reduce expression of a target gene. This technique has been effectively applied to a variety of animals, including mice, rats, cows, pigs, goats, and frogs.

In cases where low levels of expression of a gene might impair growth and development of the animal, the timing or cell-specificity of shRNA expression can be controlled. For conditional knockdown, a blocking sequence is inserted into the vector between the sequence encoding the shRNA and the promoter that controls its expression. The blocking sequence carries a termination signal for RNA polymerase and can also contain a marker gene, for example, encoding green fluorescent protein, under the control of its own promoter. The blocking sequence is flanked by two *loxP* sites in the same orientation and therefore is excised following induction of Cre recombinase. Excision of the blocking sequence restores expression of the shRNA (Fig. 12.21) and downregulation of the target gene

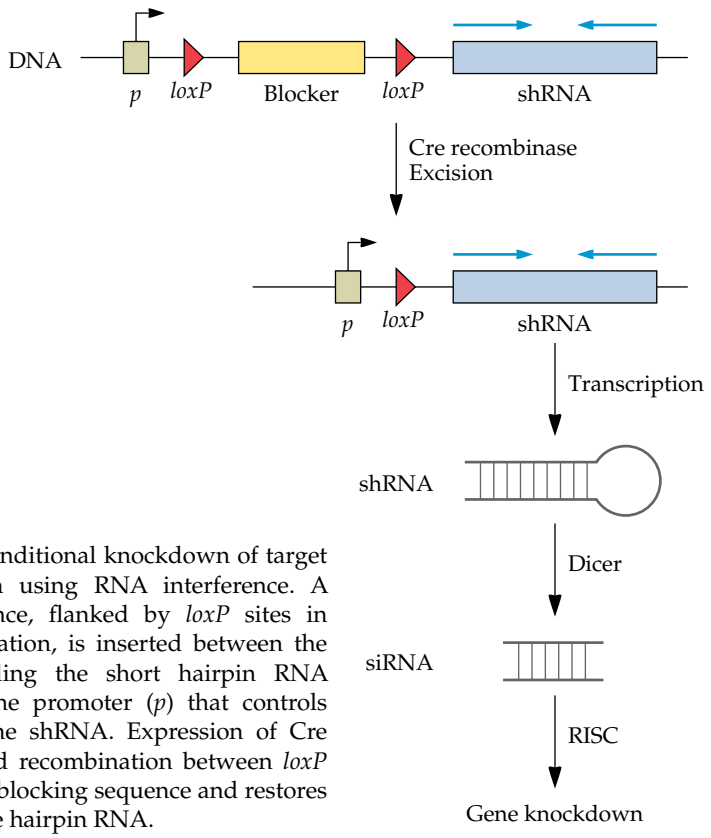


Figure 12.21 Conditional knockdown of target gene expression using RNA interference. A blocking sequence, flanked by *loxP* sites in the same orientation, is inserted between the sequence encoding the short hairpin RNA (shRNA) and the promoter (*p*) that controls expression of the shRNA. Expression of Cre recombinase and recombination between *loxP* sites excises the blocking sequence and restores expression of the hairpin RNA.

in the animal's genome. In addition to conditional transgenic expression of shRNA, gene knockdown by RNAi can occur by introducing exogenous siRNA into specific animal cells (see chapter 6).

Transgenic Animal Models of Human Diseases

Transgenic animals are important models for studying the biological basis of human diseases and for developing and testing potential therapeutic strategies. Mice are the most commonly employed animal models. They are genetically and physiologically similar to humans, sharing about 99% of their genes, and are susceptible to many of the same genetic diseases. Moreover, as described above, mice are amenable to genetic manipulation and can be engineered to accurately simulate the onset and progression of a human disease, and to investigate links between a disease and specific genes. Transgenic mouse models have been developed for many human genetic diseases, including neurodegenerative diseases such as Alzheimer's disease, amyotrophic lateral sclerosis, and Huntington's disease, tumorigenesis, arthritis, muscular dystrophy, endocrine disorders such as diabetes, and coronary disease. Researchers have also developed other animal disease models that more accurately reflect the pathology of some human diseases, for example, rabbit models of human cardiovascular disease and nonhuman primate models of neurological disorders, or that are easier to manipulate and less costly to study, such as zebrafish.

Mouse Models of Alzheimer's Disease

Alzheimer's disease is a degenerative brain disorder that is characterized by the progressive loss of abstract thinking and memory and is accompanied by personality change, language disturbances, and a slowing of physical capabilities. Clinical diagnosis of Alzheimer's disease is poor, although 1% of the population between 60 and 65 years old and 30% of the population over 80 years old may develop it. Neurofibrillary tangles of hyperphosphorylated Tau protein accumulate within the cell body of the neurons, dense extracellular aggregates called amyloid plaques develop at the ends of inflamed nerves (neuritis), and brain cells (neurons) are lost in the neocortex and hippocampus of the brain in patients with Alzheimer's disease (chapter 4, Fig. 4.15). There is currently no effective treatment for this disease, although more than 1,500 clinical trials have tested potential therapeutic drugs and antibodies.

The amyloid plaques associated with Alzheimer's disease comprise aggregates of a peptide called amyloid-β (Aβ). The Aβ peptides range in length from 39 to 43 amino acids; the Aβ40 and Aβ42 forms (40 and 42 amino acids long, respectively) are the main variants found in amyloid plaques. The Aβ peptides are derived from internal proteolytic cleavage of the amyloid-β precursor protein (APP), a transmembrane protein produced in large quantities in neurons (Fig. 12.22A). There are two pathways for cleavage of APP. One (nonamyloidogenic), at the cell surface by the sequential activity of proteases α- and γ-secretase, does not generate Aβ (Fig. 12.22B), and the other (amyloidogenic), in an endosome by proteases BACE1 (β-site APP cleaving enzyme 1, also known as β-secretase)

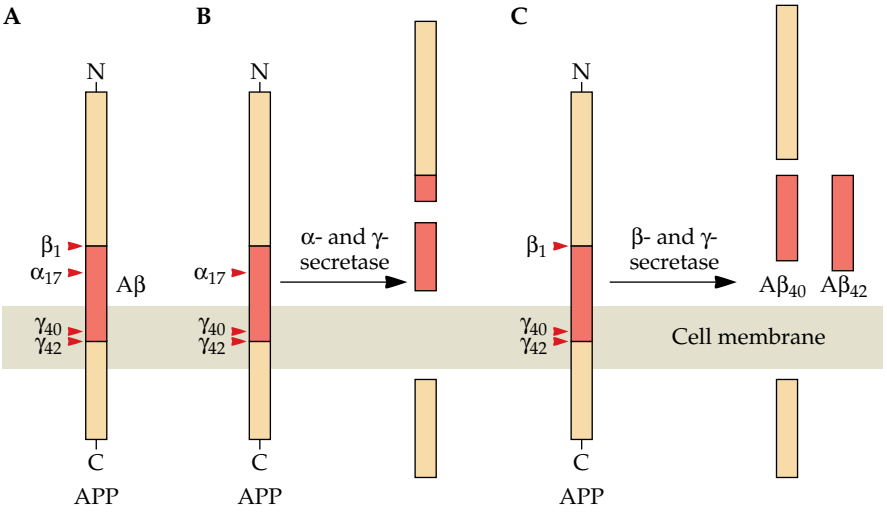


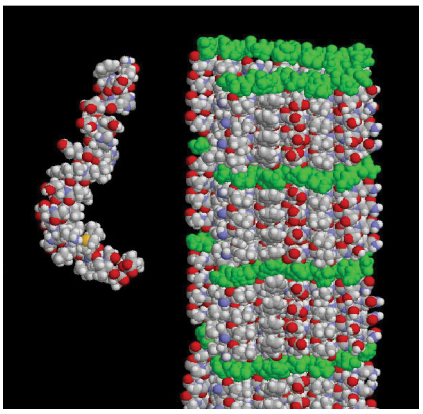
Figure 12.22 The amyloid precursor protein (APP) **(A)** may be cleaved sequentially by α- and γ-secretase in the nonamyloidogenic pathway **(B)** or by β- and γ-secretase that produces the amyloidogenic Aβ40 and Aβ42 peptides **(C)**. Small arrows indicate cleavage sites within the Aβ peptide region for the corresponding secretases.

and γ-secretase, produces Aβ40 and Aβ42 (Fig. 12.22C). Abnormal folding of these variants causes them to spontaneously aggregate and form amyloid fibrils (plaques) when they are released from the cell (Fig. 12.23). Aβ processing and deposition are major targets of the efforts to develop a treatment for Alzheimer's disease.

Although most cases are not genetically inherited, a small number of families with a high incidence of Alzheimer's disease have mutations in the *APP* gene, a finding that implicated this gene in the disorder. It is not possible to study the onset and pathogenesis of Alzheimer's disease in human subjects, and therefore an animal model that mimics Alzheimer's disease is an invaluable research tool. Mouse models for Alzheimer's disease were created that express transgenes containing mutations in the *APP* gene that occur in some families with a high incidence of early onset (before 50 years of age) of Alzheimer's disease. In one set of these families, site 717 of APP contains phenylalanine instead of valine (V717F). In another group of families with Alzheimer's disease, sites 670 and 671 of APP contain asparagine and leucine instead of lysine and methionine (K670N/M671L).

A transgene with the V717F mutation was constructed from cDNA derived from the human APP gene. Three modified introns were introduced into the APP cDNA because experiments have shown that transgenes with introns have an increased rate of transcription in comparison to constructs without them (Fig. 12.24). Expression of the transgene is controlled by the promoter for the gene encoding platelet-derived (PD) growth factor β, which is expressed in brain tissue. The complete construct is called the PDAPP minigene. Aging transgenic mice (more than 6 months old) with about 40 copies of the PDAPP minigene display amyloid plaques composed mainly of Aβ42, neuronal cell death, and memory defects. Transgenesis with an APP transgene carrying the K670N/M671L

Figure 12.23 Aβ peptides attain an extended configuration when they are part of APP in the cell membrane (left). Cleavage of APP by the amyloidogenic pathway releases abnormally folded Aβ peptides that aggregate (right) and form amyloid plaques that are hallmarks of Alzheimer's disease. Protein Data Bank ID: 2BEG; Luhrs et al., *Proc. Natl. Acad. Sci. USA* **102:**17342, 2005.



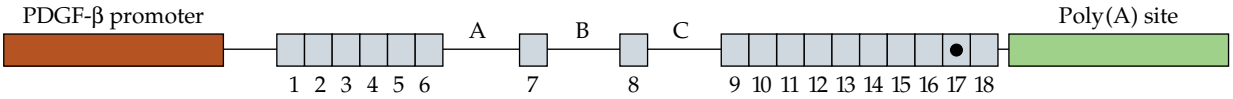
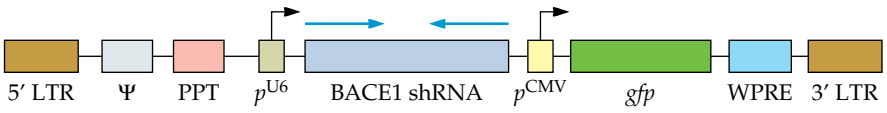


Figure 12.24 DNA construct for modeling Alzheimer's disease in transgenic mice. The exons of the APP cDNA are marked by numbers (1 to 18), and the introduced introns are indicated by letters (A to C). The regulatory sequences are the platelet-derived growth factor β (PDGF-β) promoter and simian virus 40 poly(A) sequence. The dot marks the exon with the V717F mutation. The construct is called the PDAPP minigene.

mutations driven by a brain-specific promoter produced mice with Alzheimer's disease-like features, including an excess of Aβ42. Interestingly, neither aging PDAPP minigene mice nor APP double mutant (K670N/M671L) transgenic mice have neurofibrillary tangles. Possibly, these structures are a response to the overproduction of Aβ42 in humans. Several other Aβ-overproducing transgenic mouse lines have been developed that carry human APP transgenes with single or multiple mutations, as well as mutant forms of presenilin, a subunit of γ-secretase that processes APP.

The formation of amyloid plaques in humans has also been shown to be associated with increased production of the protease BACE1 that cleaves APP to produce Aβ. Transgenic mice that produce APP but also carry a knockout mutation in BACE1 do not develop Aβ amyloid plaques. However, BACE1 knockout mice exhibit behavioral defects, which indicates that some BACE1 is required for normal development and/or normal adult brain activity. RNA interference to reduce, but not abolish, production of BACE1 may therefore represent an attractive treatment to reduce or delay Alzheimer's disease. This strategy was tested in mouse models that carry transgenes encoding mutant forms of APP that are genetically linked to familial Alzheimer's disease in humans. Short hairpin RNAs that target *BACE1* mRNA were encoded on a lentiviral vector that was injected into the hippocampus (the region of the brain where amyloid plaques associated with Alzheimer's disease are typically observed) of transgenic mice (Fig. 12.25). The mice injected with the RNAi construct showed a 38% reduction in the Aβ deposits and plaque formation within one month of injection.

Figure 12.25 Lentiviral vector to knock down BACE1 expression in mouse models of Alzheimer's disease. A sequence encoding the inverted repeats (blue arrows) of an shRNA targeting *BACE1* mRNA was inserted into the lentiviral vector downstream of the mouse RNA polymerase III promoter (p^{U6}). The gene for green fluorescent protein (*gfp*), under the control of the cytomegalovirus promoter (p^{CMV}), is used as a reporter gene to identify infected mouse cells. LTR, long terminal repeats at the 5′ and 3′ ends of the vector; Ψ, viral packaging signal; PPT, polypurine tract sequence; WPRE, woodchuck posttranscriptional regulatory element for enhancing transduction and expression of transgenes in host cells.



Overproduction of Aβ peptides is only one pathological hallmark of Alzheimer's disease. Mouse models have also been created that produce human tau protein to study the formation of neurofibrillary tangles in neurons, variants of human apolipoprotein E that influence Aβ clearance in the brain, variants of the membrane receptor TREM2 (triggering receptor expressed on myeloid cells 2) that increase the risk of developing late-onset Alzheimer's disease, and combinations of these transgenes with APP variants. While there is much to learn about Alzheimer's disease, the availability of mouse models has helped to elucidate the molecular basis, and reveal some potential targets for treatment, of a disorder that affects about 35 million people worldwide.

Mouse Model of Duchenne Muscular Dystrophy

Since the development of the CRISPR-Cas system, there has been considerable excitement about its potential to edit disease-causing mutations in the human genome. Duchenne muscular dystrophy is a genetic disease that progressively destroys skeletal, respiratory, and cardiac muscle tissues and shortens life expectancy to just over 20 years. Mutations in the dystrophin gene abolish production of the dystrophin protein that attaches the extracellular matrix to the actin cytoskeleton. In 83% of individuals with this disease, the reading frame of the dystrophin gene is disrupted by nucleotide deletions in one or more exons. The nature of the mutations makes them candidates for corrective gene therapies that replace a defective exon with a functional copy or that excise a defective exon to correct the reading frame.

The potential of the CRISPR-Cas system to correct a defective exon in the dystrophin gene was tested in a mouse model of Duchenne muscular dystrophy. These mice, known as *mdx* mice, carry a point (nonsense) mutation that results in a premature stop codon in exon 23 of the dystrophin gene (Fig. 12.26A). An sgRNA was designed to target the stop codon for cleavage by Cas9 (Fig. 12.26B). Donor DNA consisted of a single-stranded oligonucleotide to replace the stop codon with a codon for an amino acid (glutamine) by homologous recombination with sequences flanking the target site. One-cell embryos were coinjected with the sgRNA, Cas9 mRNA, and donor DNA, and then the embryos were implanted into pseudopregnant mice. In 7 of 11 transgenic mice, the stop codon in exon 23 was replaced with the codon for glutamine, while in 4 transgenic mice, the reading frame of the dystrophin gene was restored by in-frame deletions in the stop codon as a consequence of nonhomologous end joining. In most of the transgenic mice fewer than 100% of cells carried the restored exon. This mosaicism is due to differences in the timing of transgene integration into the genome during the development of the embryo; fewer cells will carry the transgene if integration occurs after DNA replication. Transgenic mice with at least 41% corrected alleles or 83% of alleles with in-frame deletions expressed dystrophin in skeletal muscles and exhibited normal muscle function. There was no evidence of CRISPR-Cas-mediated mutations at potential off-target sites in the genome.

Although these results demonstrate that it is possible to correct the dystrophin gene in early-stage embryos of a mouse model of Duchenne muscular dystrophy, corrective editing of human embryos is currently

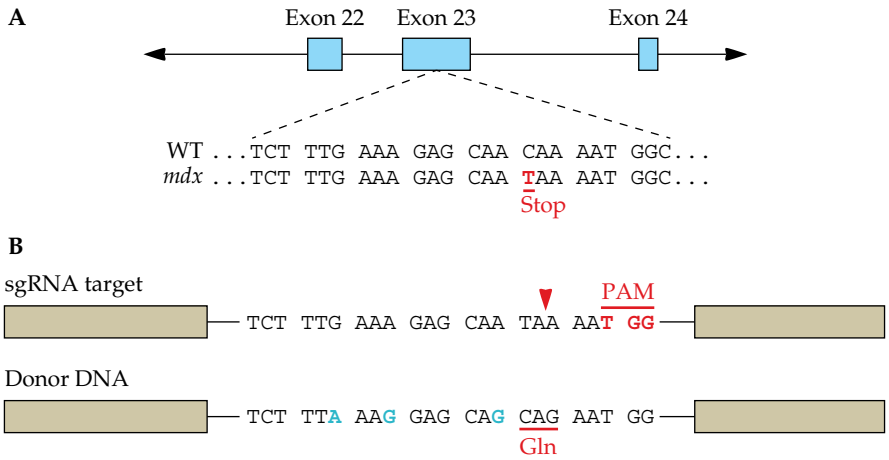


Figure 12.26 Correction of a deleterious mutation in the dystrophin gene in a mouse model of Duchenne muscular dystrophy using the CRISPR-Cas system. **(A)** The *mdx* mouse carries a point mutation in exon 23 of the dystrophin gene that creates a premature stop codon, which results in production of a truncated dystrophin protein. **(B)** To correct the mutation, an sgRNA was designed to target Cas9 to a cleavage site (red arrow) within the stop codon, which is upstream of a PAM. The stop codon was replaced with the sequence encoding glutamine (Gln) following repair of the cleaved DNA by recombination between homologous sequences (brown boxes) flanking the target site in the genome and on an introduced donor oligonucleotide. Silent mutations (shown in blue) were introduced via the donor DNA to facilitate identification of mice with edited genomes. Adapted from Long et al., *Science* **345:**1184–1188, 2014, with permission.

not feasible due to safety and ethical concerns. However, it may be possible to restore dystrophin production and improve muscle function in individuals with the disease by editing the dystrophin gene specifically in muscle cells using the CRISPR-Cas system. This possibility was tested in *mdx* mice. Because mature, nondividing cells do not normally produce the proteins that are required for sequence insertion via homologous recombination, the defective exon with the premature stop codon was targeted for deletion rather than replacement. This was expected to restore the reading frame of the dystrophin gene and production of a partially functional dystrophin protein that would lead to a milder disease phenotype. Adeno-associated virus vectors (see chapter 6) were used to deliver Cas9 cDNA and two sgRNAs targeting sequences in introns 22 and 23 that flank exon 23 (Fig. 12.27). The vectors were coinjected into muscles of the *mdx* mice. Eight weeks after injection, PCR and sequence analysis revealed the targeted 1,171-bp deletion encompassing exon 23 of the dystrophin gene in some of the muscle cells. Exon 23 was deleted from 59% of dystrophin mRNA in muscle tissues, and production of dystrophin increased to levels that are considered to be sufficient to increase muscle function in individuals with Duchenne muscular dystrophy; this was borne out in an examination of muscle function in the treated mice. Although some safety and technical issues remain to be addressed before it is applied in humans, the CRISPR-Cas system shows enormous potential as a therapy for Duchenne muscular dystrophy and other genetic diseases caused by loss of specific proteins.

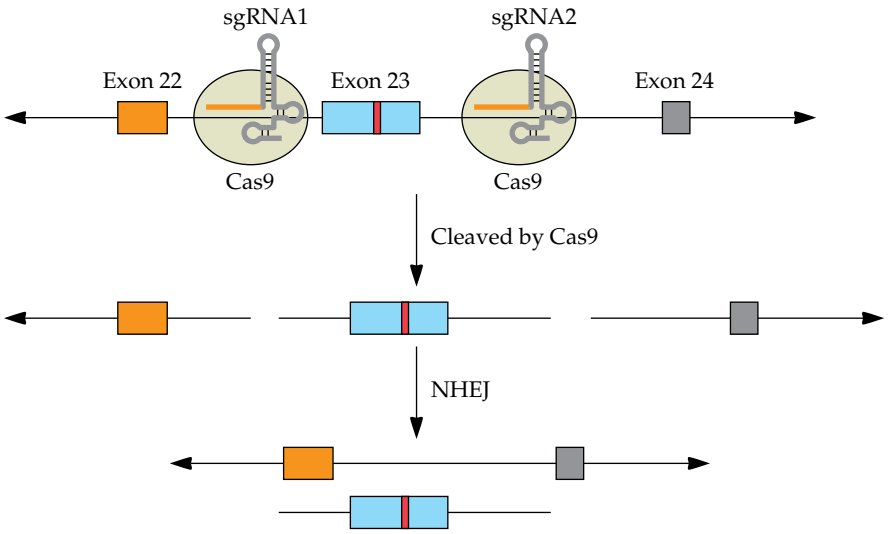


Figure 12.27 Excision of a defective exon in the dystrophin gene in muscle cells of *mdx* mice using the CRISPR-Cas system. Two sgRNAs were designed to target Cas9 (beige oval) to specific sequences in the introns flanking exon 23, which carries a premature stop codon (red line). The sgRNAs and Cas9 cDNA were introduced into skeletal and cardiac muscle cells using a viral vector. Expression of Cas9 resulted in double-stranded breaks at the target sites that were repaired by nonhomologous end-joining (NHEJ). Removal of exon 23 restored the reading frame of the dystrophin gene and production of modified dystrophin and improved muscle function. Data from Nelson et al., *Science* **351:**403–407, 2016.

Rabbit Models of Cardiovascular Disease

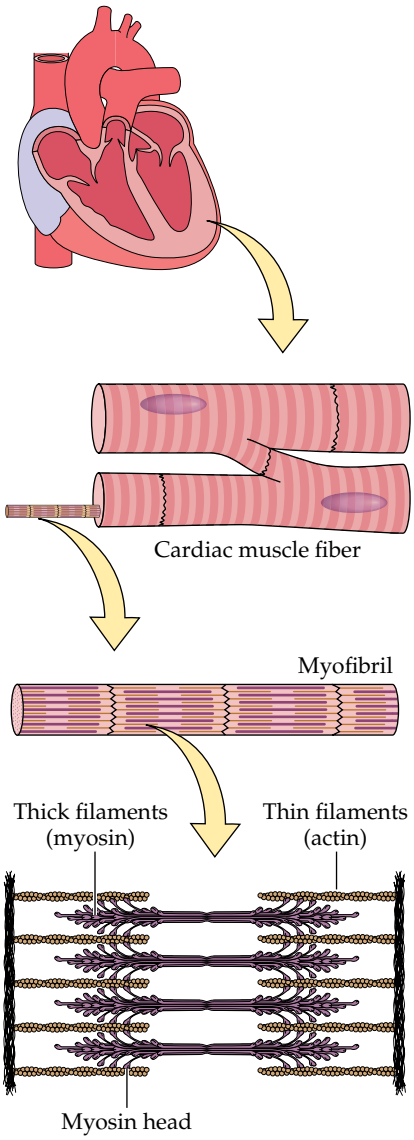
Despite the predominance of transgenic mouse models for studying human diseases, they do not accurately recapitulate some diseases, particularly disorders of lipid metabolism and cardiovascular disease. Rabbits are well suited as models for these diseases. Compared to mice, their lipid metabolic system is more similar to that of humans, especially lipoprotein profiles and cholesterol processing. For this reason, they are useful models of hypercholesterolemia and atherosclerosis. The protein composition of heart muscles is also similar in rabbits and humans. Moreover, rabbits have been used as laboratory animals for nearly 200 years, particularly for drug safety testing and vaccine development. They are tame and therefore easy to handle, have high reproductive rates, and have large ears, which facilitates intravenous injection and blood sampling.

Transgenic rabbits are generated primarily by pronuclear injection, although the efficiency is low (0.5% of injected and implanted embryos), and multiple copies (2 to 100) of the transgene are often introduced. Embryonic stem cell lines have not been established for introducing transgenes into rabbits by homologous recombination; however, there has been some success using lentiviral vectors and somatic cell nuclear transfer and, more recently, genome editing using the CRISPR-Cas system, which tends to integrate a single copy of a transgene at the targeted site.

A transgenic rabbit model was created to study a devastating form of familial hypertrophic cardiomyopathy that is due to a single point mutation in the human β-myosin heavy chain gene resulting in a change from

arginine (R) to glutamine (Q) at position 403 in the amino acid sequence of the protein. The disease is often asymptomatic until cardiac arrest following ventricular arrhythmia causes sudden death, including in young athletes. A mouse model carrying the R403Q mutation has been available for 25 years; however, the activity of mutant cardiac myosin from the engineered mice contrasts with that of mutant human cardiac myosin.

Figure 12.28 Structure of a myofibril, the contractile unit of cardiac muscle. Located in multinucleate muscle cells (muscle fibers), myofibrils are composed of parallel thick filaments of myosin that interact with actin in the thin filaments. During a muscle contraction, ATP binds to the globular head region of myosin and is hydrolyzed to ADP, causing myosin to form a cross-bridge with actin. Release of ADP from myosin pulls actin along the myosin filament, generating a contractile force. In a transgenic rabbit model of familial hypertrophic cardiomyopathy, a mutation (R403Q) in the head region of human β-myosin weakened its interaction with actin and reduced power output from cardiac muscle.



Muscle contraction is mediated by the interaction between myosin in thick filaments and actin in thin filaments that make up the contractile myofibrils (Fig. 12.28). It has been proposed that the R403Q mutation in β-myosin heavy chain, which is present at the myosin-actin interface, weakens its interaction with actin. Whereas the ATPase activity and actin movement of myosin from humans with the R403Q mutation decrease, they increased in myosin from the mouse model. This may be because the mutation was introduced into the mouse gene encoding α-myosin heavy chain, the major muscle protein in mice, whereas in humans the predominant protein is β-myosin heavy chain, which is 93% identical to the amino acid sequence of the α-myosin heavy chain. The β-myosin heavy chain has only a very low level of expression in cardiac tissue of mice, and therefore mice with the R403Q mutation in this gene do not exhibit a phenotype. Similar to humans, the major muscle protein in rabbits is β-myosin heavy chain, and therefore rabbits are more suitable disease models for cardiac arrythmias and other cardiopathies.

To create a rabbit model of hypertrophic cardiomyopathy, the R403Q mutation was introduced into cDNA encoding a rabbit β-myosin heavy chain by overlap extension PCR (see chapter 3). The mutant gene was inserted into a vector downstream of the rabbit β-myosin heavy chain promoter and introduced into fertilized oocytes by pronuclear injection. Of the total β-myosin heavy chain mRNA produced by the transgenic rabbits, 33% carried the R403Q mutation (levels >40% are lethal), and heterodimers of mutant and wild-type β-myosin heavy chain were isolated from cardiac muscle. In contrast to previous reports from transgenic mice, the movement of actin filaments by myosin purified from the heart muscle of transgenic rabbits was significantly lower than that from nontransgenic rabbits, suggesting an impairment in actin-myosin interaction which is necessary for muscle contraction. The force produced by microfibrils isolated from transgenic rabbits was lower than that from nontransgenic rabbits, and force development during contraction and relaxation after contraction were slower. The results from the transgenic rabbits indicate that the R403Q mutation in the β-myosin heavy chain reduces power output from heart muscle. Further study of this and other transgenic rabbit models of the more than 300 mutations that occur in human β-myosin heavy chain will contribute to our understanding of human heart disease.

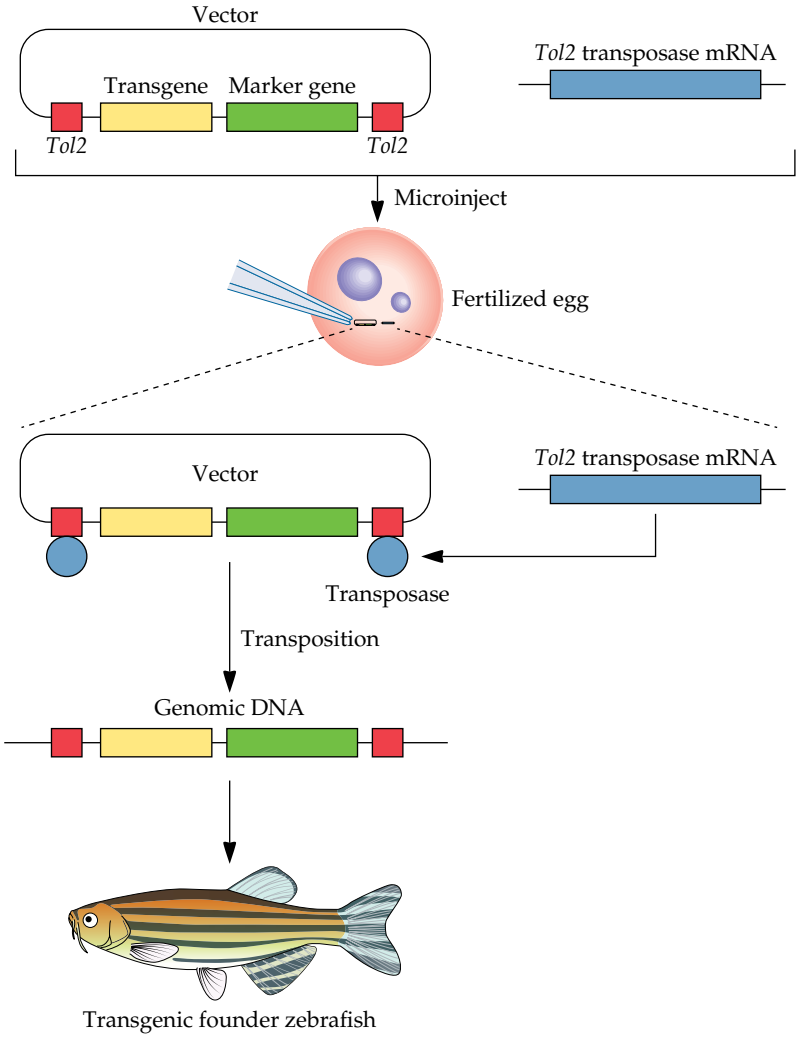
Zebrafish Melanoma Model

Zebrafish (*Danio rerio*) are small tropical freshwater fish that are rapidly gaining popularity as animal models for human diseases. At first glance, they may not seem anything like humans; however, at the genetic and physiological levels they are surprisingly similar. About 70% of human genes, and 82% of those known to be associated with human diseases, are found in zebrafish. It is relatively easy to create transgenic zebrafish by microinjection of DNA into fertilized eggs because each female fish produces hundreds of eggs per mating, the eggs are externally fertilized, and the embryos are readily amenable to genetic manipulation. Zebrafish are especially useful for studying normal and aberrant developmental processes because their embryos mature quickly (24 hours), development is *ex vivo* (outside an organism), the fish have a short generation time (3 months), and embryos and larvae are transparent, enabling

expression patterns of proteins or single cells that are genetically labeled with a fluorescent protein to be visualized during development in live animals.

To create transgenic zebrafish, fertilized eggs are collected and microinjected with a transgene construct. For integration into the zebrafish genome, the transgene is flanked by recognition sites for *Tol2* transposase and cloned into a plasmid, which is then injected into the embryos together with mRNA encoding *Tol2* transposase (Fig. 12.29). In the embryo, the *Tol2* mRNA is translated, and the transposase mediates integration of the

Figure 12.29 Zebrafish transgenesis. A vector carrying the transgene and a selectable marker gene inserted between two inverted repeat sequences (*Tol2*) and *Tol2* transposase mRNA are microinjected into fertilized zebrafish eggs. In the embryo, the injected mRNA is translated to produce *Tol2* transposase that binds to the *Tol2* sequences on the vector. *Tol2* transposase catalyzes the excision of the transgene and marker gene from the vector and its stable integration (transposition) into the zebrafish genome. The embryos develop *ex vivo* to produce transgenic founder fish.



construct at a random site in the zebrafish genome; the mRNA is degraded, which reduces the opportunity for movement of the transgene after it has been integrated. As in other animals, the transgene may integrate into a heterochromatin region, and therefore a marker gene is often included on the injected plasmid between the transposase recognition sites to identify embryos that express the transgene. The embryos are raised to produce transgenic founder fish which are outcrossed with nontransgenic zebrafish to screen for germline transmission. Offspring with stable transgene integration are selected to establish a transgenic zebrafish line. To reduce the possibility of the transgene integrating into heterochromatin, it may be targeted to a specific site in the zebrafish genome using the CRISPR-Cas genome-editing system.

Transgenic zebrafish are an excellent model in which to visualize cancer development and progression in a live animal. Melanoma is a type of skin cancer that develops from transformed melanocytes, which are pigment-producing cells of the skin and other tissues. Many transformed melanocytes carry a mutation in the *BRAF* gene at codon 600 that replaces a glutamic acid (E) with valine (V) in the encoded protein B-Raf. B-Raf is a serine/threonine protein kinase that regulates a signal transduction pathway controlling cell division, differentiation, and apoptosis in response to growth factors (chapter 4, Fig. 4.27A). The $BRAF^{V600E}$ mutation may cause constitutive kinase activity of B-Raf and uncontrolled proliferation of melanocytes. However, not all cells with the $BRAF^{V600E}$ mutation will develop into a tumor. The early molecular processes that initiate tumorigenesis are not well understood, in part because it is difficult to predict which of the cells carrying the mutation will be transformed. Transgenic zebrafish carrying the human $BRAF^{V600E}$ gene are an excellent model in which to study the development of melanoma.

A zebrafish model of melanoma carrying the human $BRAF^{V600E}$ transgene (cDNA derived from mRNA extracted from human melanoma cells) under the control of a melanocyte-specific promoter was generated by DNA microinjection of single-cell embryos using the *Tol2* transposase system described above. The zebrafish also carries a knockout mutation in the endogenous *p53* gene because a deficiency in p53 tumor suppressor was previously shown to be required for melanoma progression. The transgenic zebrafish develop invasive melanoma after several months. However, despite all melanocytes carrying the $BRAF^{V600E}$ oncogene, only rare single cells are activated to form a tumor. To visualize activated melanocytes that initiate melanoma, green fluorescent protein was expressed as a fusion protein with the zebrafish protein crestin from the *crestin* gene promoter and enhancer. The *crestin* promoter is normally active only in embryonic neural crest cells (progenitor cells from which melanocytes and some other cell types develop). However, in the zebrafish melanoma model, the *crestin* promoter was specifically reactivated in melanoma tumors of adult fish as determined by expression of green fluorescent protein. This indicates that melanoma is initiated in melanocytes that revert to an embryonic neural crest-like state. Expression of green fluorescent protein enabled tracking of tumor development from single melanocytes in live zebrafish (Fig. 12.30). Transgenic zebrafish have been used as models for many human cancers, such as pancreatic, hepatocellular, and thyroid cancer, as well as blood diseases, collagenopathies, diabetes, and more.

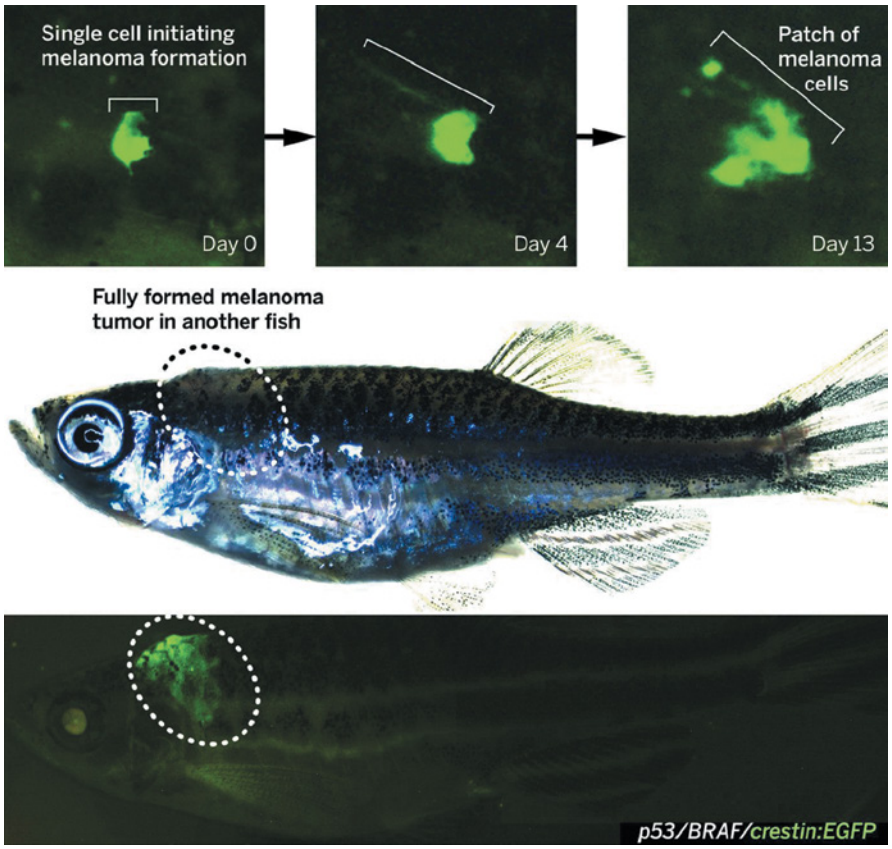


Figure 12.30 Development of melanoma from a single melanocyte in a transgenic zebrafish model. (Top) A single melanocyte expressed green fluorescent protein (GFP) from the *crestin* promoter in a zebrafish carrying the human $BRAF^{V600E}$ transgene and a knockout mutation in the endogenous tumor suppressor *p53* gene. (Bottom) The melanocyte proliferated to form a melanoma tumor which was visualized as green fluorescence. Reprinted from Kaufman et al., *Science* **351:**464, 2016, with permission.

Nonhuman Primate Models of Neurodevelopmental Disorders

While mouse models have contributed greatly to our understanding of human diseases, the short lifespan of the mouse may reduce its utility for the study of slow and progressive diseases such as Huntington's disease that require observation over a longer period of time (Fig. 12.19). Transgenic primate models such as the rhesus macaque may represent more accurate models for the study of human neurodegenerative diseases. Transgenic macaques were produced that expressed the mutant human *HTT* gene encoding the huntingtin protein with an expanded polyglutamine sequence found in individuals with Huntington's disease. Rhesus macaque oocytes were microinjected with a lentivirus vector containing exon 1 of the human *HTT* gene with 84 CAG repeats under the control of the human polyubiquitin C promoter. The oocytes were fertilized *in vitro* by intracytoplasmic injection of sperm and then cultured to the 4- or 8-cell stage before implantation of embryos in a surrogate female. Of 30 transplanted embryos, 5 live monkeys were delivered, and 2 of these died within 1 day of birth. Although the contribution of the *HTT* gene to their deaths cannot be established, postmortem analysis confirmed the presence of multiple

copies of the *HTT* gene and expression of huntingtin with a polyglutamine sequence. One of the surviving monkeys carries a single copy of the *HTT* transgene with 29 CAG repeats and at 6 months old showed features of Huntington's disease that are found in humans with the disease, including involuntary, jerky body movements (chorea) and muscle contractions (dystonia). Over a 4-year period, the transgenic macaque exhibited abundant neuronal aggregates of huntingtin protein, neuronal loss, a lower volume of specific regions of the brain (hippocampus and striatum), and a decline in motor and cognitive functions consistent with the progression of Huntington's disease in humans that were not observed in nontransgenic control macaques, which carry an *HTT* gene with 10 to 11 CAGs. The macaque will be used to establish a transgenic line for long-term studies of the pathology of Huntington's disease and to assess treatment and diagnostic strategies. Therapeutic strategies for Huntington's disease will probably be directed toward either blocking the expression of the mutant *HTT* gene in those people who test positive for more than 38 CAG repeats or clearing the abnormal protein from the neurons. In either case, the treatment must precisely target the mutant gene or protein, leaving the normal counterparts unaffected.

In addition to models of Huntington's disease, transgenic nonhuman primate models have been created for other human neurodevelopmental disorders, including Parkinson's disease and autism. Nonhuman primates are well suited as models for these diseases because brain anatomy, organization, and development, and motor, perceptual, and cognitive functions are more similar to those of humans than are those of rodents. They are also better models of aging, which is important for the study of age-related neurodegeneration. Moreover, the transgenic primates are a source of cells from which to generate induced pluripotent stem cells for the study of pathogenesis during development. For example, a neural progenitor cell line was established from the macaque Huntington's disease model and used to test drug and RNAi treatments for the disease. While there are many advantages to using transgenic nonhuman primate disease models, they are costly and time-consuming to produce, in large part because the genetic modification techniques are inefficient and the number of embryo transfers that result in pregnancy is low, with only a small number of these resulting in live births. Because they are closely related to humans and have sophisticated cognitive abilities and behaviors, the creation and use of nonhuman primates with devastating neurological diseases is a sensitive issue. However, many believe that it is justified if it will reasonably lead to alleviation of human suffering from these diseases.

Animal Bioreactors for Production of Recombinant Therapeutic Proteins

Mammary glands of transgenic animals are attractive bioreactors for the production of human therapeutic proteins such as hormones, growth factors, and enzymes. A transgene encoding the target protein is placed under the control of a mammary gland-specific promoter and introduced into one-cell embryos by one of the methods described previously in this chapter. The large quantities of recombinant proteins produced by the mammary glands of the transgenic animals are usually appropriately glycosylated, have biological activities identical to those from human

Table 12.2 Some recombinant therapeutic proteins expressed in the mammary glands of transgenic animals that are in clinical trials or have been approved for use[a]

Protein	Transgenic animal	Application
Antithromobin	Goat	Prevent excessive blood clotting
C1 esterase inhibitor	Rabbit	Attenuate dysfunction of transplanted organs
α-Fetoprotein	Goat	Autoimmune disease
Butyrylcholinesterase	Goat	Organophosphate poisoning, Alzheimer's disease
Lactoferrin	Cow	Nutraceutical to prevent infectious disease
Growth hormone	Cow	Dwarfism
Factor VIIa	Rabbit	Hemophilia
Factor IX	Pig	Hemophilia
Fibrinogen	Rabbit, cow	Tissue sealant
α-1 Antitrypsin	Goat	Hereditary antitrypsin deficiency
Anti-CD20 monoclonal antibody	Goat	Cancer immunotherapy
Anti-CD137 monoclonal antibody	Goat	Cancer immunotherapy
Malaria merozoite surface protein	Goat	Vaccine for prevention of malaria
Rotavirus capsid proteins VP2/VP6	Rabbit	Vaccine for prevention of viral gastroenteritis

[a]Data from Bertolini et al., *Transgenic Res*. **25:**329–343, 2016.

sources, and can be readily purified from milk. Restricting production to the mammary glands avoids potential harm to the animal from the biologically active proteins. The transgenic animals are raised in a physically contained environment so they cannot transmit the transgene to other animals or enter the human or animal food supply. More than 100 different human proteins have been produced in the milk of transgenic sheep, goats, pigs, and rabbits (Table 12.2).

A major advantage of pharmaceutical protein production in mammary glands is that milk can be easily collected in large volumes. This contributes to reducing the cost of recombinant protein production to a fraction of that of production in cultured mammalian cells. Dairy cows each annually produce approximately 10,000 liters of milk. If 1 gram of recombinant protein is produced per liter of milk and it could be purified with 50% efficiency, the yield from 20 transgenic cows would be about 100 kg per annum. This is the annual global requirement for human protein C, which is used for the prevention of blood clots. One transgenic cow would be more than sufficient for the production of the annual world supply of factor IX (plasma thromboplastin component), which is used by hemophiliacs to facilitate blood clotting.

Production of Recombinant Antithrombin in Goat Milk

Although the quantity of milk produced by either a sheep or a goat is much smaller than that produced by a cow, lactation in sheep and goats yields hundreds of liters of milk per year (Table 12.3). Goats can also be raised to produce milk more rapidly than cows. In 2009, the U.S. Food and Drug Administration approved the use of the human protein antithrombin produced in goat's milk for individuals with a hereditary deficiency in the production of this protein and who are undergoing surgery

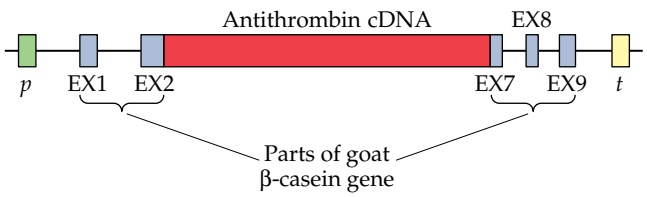
Table 12.3 Milk production and estimated recombinant protein yields from animals used for the expression of transgenes in mammary glands

Organism	Annual milk yield (liters)	Estimated recombinant protein per female (kg/yr)
Rabbit	5	0.02
Pig	300	1.5
Sheep	500	2.5
Goat	900	4
Cow	10,000	60

or giving birth. Antithrombin is a protease inhibitor that acts as an anti-clotting factor by inhibiting the activity of thrombin and other coagulation proteases and thereby prevents the excessive formation of blood clots and promotes the clearing of clotting factors. It also has anti-inflammatory activity. Approximately 1 in 5,000 Americans are unable to produce this protein naturally, which puts them at risk for heart attacks and strokes. While antithrombin can be extracted from the plasma of donated blood, the supply is not sufficient to meet the needs of all potential patients. Blood extraction is also less efficient, more costly, and has a higher risk of contamination with human pathogens compared to milk extraction.

To produce antithrombin in goats, human antithrombin cDNA was fused to the promoter of the goat β-casein gene (Fig. 12.31), and the construct was introduced into embryos by microinjection. Importantly, the recombinant antithrombin that was secreted into the milk of the transgenic goats was structurally similar to the human protein except for some differences in the oligosaccharides at three of the four *N*-linked glycosylation sites. These differences do not affect the thrombin inhibition activity of the protein. Transgenic goats' milk is a significant source of human antithrombin, with yields of 2 to 10 grams per liter of milk. In cell cultures, yields on the order of 0.2 to 1 gram per liter of culture medium have been attained. It has been estimated that 75 transgenic goats would be required to meet the annual worldwide demand for antithrombin. Several other human therapeutic proteins produced in transgenic goats are in clinical trials, including other blood proteinase inhibitors, such as antitrypsin, alpha-fetoprotein for the treatment of autoimmune disease, butyrylcholinesterase for protection against organophosphate nerve toxins, a malaria vaccine, and monoclonal antibodies for cancer immunotherapy.

Figure 12.31 Transgene construct for expression of human antithrombin in goats' milk. The full-length cDNA for human antithrombin was cloned between exon 2 (EX2) and exon 7 (EX7) of the goat β-casein gene. The promoter (*p*) and transcription termination (*t*) sequences and exons 1, 8, and 9 (EX1, EX8, and EX9) of the β-casein gene are retained.



Production of a Human Protease Inhibitor in Rabbits

Other mammals are also being developed for the commercial production of pharmaceutical proteins in milk. Rabbits produce high levels of proteins in their milk, are easy to handle and house because of their small size, reproduce rapidly, and do not transmit serious infectious diseases to humans. In 2014, the U.S. FDA approved the use of recombinant human C1 esterase inhibitor protein produced in the mammary glands of transgenic rabbits for individuals suffering from acute attacks of hereditary angioedema. Hereditary angioedema is a rare disorder, affecting 1 in 50,000 to 100,000 individuals globally, caused by a genetic deficiency in C1 esterase inhibitor, a plasma glycoprotein that normally inhibits the activity of several proteases, including those that regulate blood coagulation, inflammation, and activity of complement proteins of the innate immune system. A deficiency causes recurrent attacks of swelling due to enhanced vascular permeability in the skin, larynx, and gastrointestinal system that can be life-threatening. For several decades, acute attacks have been treated with C1 esterase inhibitor that was derived from human plasma; however, there are concerns that this approach could result in the simultaneous transmission of human pathogens. The availability of recombinant human C1 esterase inhibitor purified from the milk of transgenic rabbits maintained in a specific-pathogen-free environment addresses this concern.

Transgenic rabbits were created by introducing the cDNA encoding the human C1 esterase inhibitor, including the signal sequence for secretion and the mammary gland-specific promoter from the bovine αS_1 casein gene, into a fertilized rabbit oocyte by microinjection (Fig. 12.32). Transgenic founder animals were selected based on production levels of the recombinant protein in milk, the number of copies and the site of integration of the transgene, and the stability of transgene transmission to offspring. Yields of active C1 esterase inhibitor in the milk of transgenic rabbits were 12 g/liter on average. The amino acid sequence of the recombinant protein produced by the rabbits is identical to that of the human protein, but the pattern of glycosylation is different, which may contribute to the observed shorter half-life of the recombinant protein in blood samples of patients (~3 hours compared to the 22 to 56 hours for the human plasma-derived protein). Importantly, purified recombinant C1 esterase inhibitor effectively inhibited target proteases *in vitro* to the same extent as the human plasma-derived inhibitor and relieved the symptoms of acute angioedema attacks in several placebo-controlled, double-blind clinical trials. Other than one incident of a severe allergic reaction in a patient with a preexisting rabbit allergy, there were no serious adverse reactions directly related to the treatment during the trials.

Figure 12.32 Transgene construct for production of human C1 esterase inhibitor in rabbits. The cDNA derived from the human gene *SERPING1* (serpin family G member 1) encoding the serine protease inhibitor (serpin) C1 esterase inhibitor, including the signal sequence (orange), was expressed from the 5′ regulatory elements of the bovine mammary gland-activated *CSN1S1* gene encoding αS_1 casein for expression and secretion of C1 esterase inhibitor in rabbit's milk.

Production of Therapeutic Proteins in Chicken Eggs

Chicken eggs, which are released externally by a hen and have a high protein content, are well suited for the production of recombinant therapeutic proteins. The expression of a transgene in the cells of the reproductive tract of a hen could lead to the accumulation of a transgene-derived protein that becomes encased in the eggshell. The recombinant protein can then be extracted from the sterile eggs. Ovalbumin constitutes more than 50% of the protein of egg white; therefore, expression of a transgene under the control of the ovalbumin promoter and regulatory elements can yield high levels of recombinant protein. Yields of up to 1 gram of recombinant protein have been achieved per egg and, considering that a single hen lays more than 300 eggs per year, the productivity of these animal bioreactors could be substantial.

Several features that are unique to avian reproduction and development make the production of transgenic strains by microinjection of DNA into fertilized eggs extremely inefficient. During fertilization in birds, several sperm penetrate the ovum instead of one, as usually occurs in mammals. As a result, it is impossible to identify the male pronucleus that will fuse with the female pronucleus. DNA injected into the cytoplasm of the fertilized egg does not integrate into genomic DNA. In one method for avian transgenesis, pluripotent cells derived from blastoderm or primordial germ cells that are maintained in culture are transfected with a transgene construct, selected for integration of the transgene by homologous recombination, and then implanted into the subgerminal space of recipient embryos of freshly laid eggs (Fig. 12.33). Another, preferred method for transgenesis of chickens is by injection of a retroviral vector carrying a transgene (Fig. 12.4) into the subgerminal space below the embryo in a newly laid (stage X) egg. A small hole is drilled in the shell of a surface-disinfected egg through which to inject the viral particles, the hole is then sealed, and the egg is incubated *in vitro* until it hatches. The chicks from this method are usually mosaics, because only a portion of the cells of the embryo from which they develop have integrated the transgene in their genome. Transgenic lines are established from the mosaic animals by several rounds of matings.

Human lysosomal acid lipase is an example of a therapeutic protein produced in transgenic chicken eggs. It was approved by the U.S. FDA in 2015 for enzyme replacement therapy in individuals with a genetic deficiency in this protein. These individuals cannot metabolize lysosomal cholesterol esters and triglycerides derived from low-density lipoproteins, and therefore the lipids accumulate in their liver, spleen, and blood vessels. One severe, rapidly progressing form of the disease (Wolman disease) causes fatalities in infants, and another later-onset form in older children and adults leads to liver cirrhosis and atherosclerosis. In a transgenic chicken line, human *LIPA* cDNA encoding lysosomal acid lipase, introduced into chicken embryos on a retroviral vector, is stably integrated at a single chromosomal site. The protein is expressed from the chicken ovalbumin promoter and enhancer sequences and secreted into egg whites from which it is purified. Treatment with purified recombinant lysosomal acid lipase increased survival of infants with Wolman disease and substantially reduced triglycerides and low-density lipoprotein cholesterol in older children and adults with less severe forms of the disease.

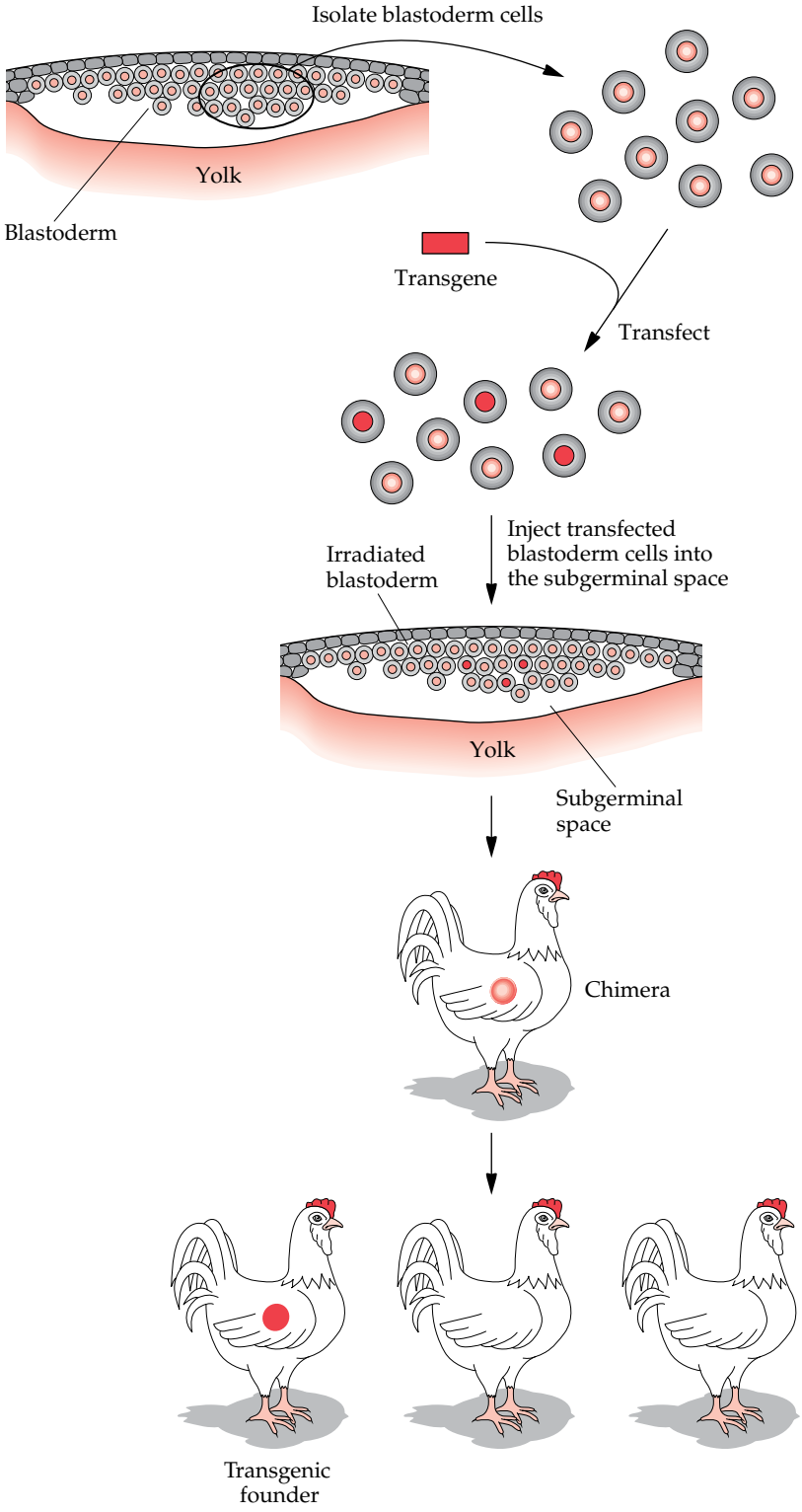


Figure 12.33 Establishing transgenic chickens by transfection of isolated blastoderm cells. Cells from blastoderm donors are removed, transfected with a transgene, and inserted into the subgerminal space of a recipient blastoderm. The resulting chickens are usually chimeras which contain cells derived from two different individuals, in this case, from the donor and recipient blastoderms. (Note that a chimera is not the same as a mosaic in which different cell populations arise from a single zygote.) Recipient embryos may be gamma-irradiated prior to transfection to destroy some but not all of the blastoderm cells, thereby increasing the final ratio of donor to recipient cells in the chimeric chicken. The chimeras that have the transgene in germ line cells are bred to establish transgenic lines.

Transgenic chickens that synthesize monoclonal antibodies, growth hormone, erythropoietin, lysozyme, interferon, and other human proteins have been created. Regulatory approval of therapeutic proteins produced in eggs may be relatively straightforward, as chicken eggs are already used to produce vaccines for injection into humans.

Production of Donor Organs in Pigs

Animals are a potential source of organs for transplantation into humans. Human-to-human organ transplants (allotransplantations, or allografts) of hearts, livers, and kidneys are 75% to 95% effective for the first year, and on average, transplant patients survive for 10 to 15 years. However, throughout the world, the demand for donated organs far exceeds the available supply. In the United States, for example, more than 107,000 organ transplants were required in 2020, but only about 39,000 were performed. With this in mind, animal-to-human transplants (xenotransplantations, or xenografts) have been proposed as a way to alleviate this organ shortage. Swine are considered to be the most likely source of organs for xenotransplantation because their organs are similar in size and physiological functions to humans and they can be raised in large numbers. Because they are produced for food, it might be socially acceptable to use them as organ donors.

A major impediment to organ transplantation between species is hyperacute rejection of the animal organ. Hyperacute rejection occurs when preexisting antibodies of the host (e.g., human) bind to the epitope galactose-α1, 3-galactose (α-Gal) on the surface of the cells of the grafted organ. The carbohydrate α-Gal is present on cell surface glycoproteins and glycolipids of pigs and many other mammals, but not humans and some other primates. The antibodies bound to α-Gal elicit an inflammatory response (complement cascade) that destroys the antibody-coated cells and leads to the loss of the transplanted organ within hours. Under natural conditions, proteins on the surface of the cells lining the blood vessels protect the cells from the inflammatory response. These complement-inhibiting proteins are species specific. Therefore, it was reasoned that if the donor animal carried one or more of the genes for a human complement-inhibiting protein, a transplanted organ might be protected from the initial inflammatory response.

Transgenic pigs with different human complement inhibitor genes were produced by pronuclear microinjection of the transgene into fertilized eggs. Hyperacute rejection did not occur after kidneys from transgenic pigs were transplanted into a primate host, and survival rates were 20 to 90 days depending on the human complement inhibitor expressed. Survival rates also depended on the levels of immunosuppressive drugs that were administered. Survival improved significantly in primate transplantation trials of hearts and kidneys produced in transgenic pigs in which production of the antigenic α-Gal epitope was abolished by deleting the gene encoding α1,3-galactosyltransferase. When these genetic modifications were combined in transgenic pigs that expressed human coagulation regulatory proteins (e.g., thrombomodulin) to reduce immune-activated formation of blood clots, survival of a heart transplanted into a baboon increased by several months to more than 1 year. Knockout pigs lacking other surface antigens have also been created, and human sera showed reduced levels of antibodies binding to cells from these pigs.

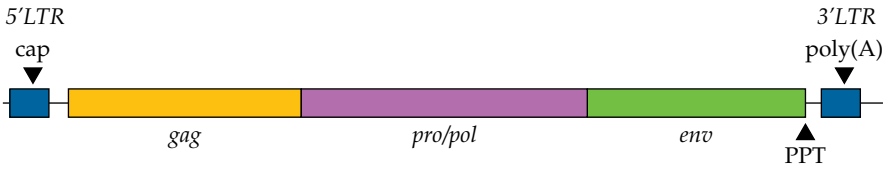


Figure 12.34 Structure of a PERV genome integrated into a pig chromosome. The PERV genome contains three genes encoding polyproteins that are processed by proteases to produce viral proteins. The group-specific antigen gene, *gag*, encodes capsid, nucleic acid-associated, and membrane-associated proteins; the protease/polymerase gene, *pro/pol*, encodes reverse transcriptase, integrase, and a protease; and the envelope protein gene, *env*, encodes a surface glycoprotein and a transmembrane protein. The cap sequence in the 5′ long terminal repeat (LTR) contains the transcription start site; PPT encodes a polypurine tract; and the 3′ LTR contains a polyadenylation [poly(A)] signal. To inactivate the PERV, sgRNA was designed to guide Cas9 endonuclease to cleave within the *pro/pol* gene, which is essential for production and host integration of PERV DNA. Adapted from Denner and Tönjes, *Clin. Rev. Microbiol*. **25:**318–343, 2012.

The potential for latent, genome-integrated pig pathogens such as porcine endogenous retrovirus (PERV) to become activated after xenotransplantation and cause infections in humans must also be addressed. Preliminary information indicates that PERVs can infect and replicate in some human cell lines, which can then transmit the viruses to other human cells *in vitro*. While there is no evidence that PERVs infect or produce adverse symptoms in humans *in vivo*, pigs with all copies of PERV inactivated were created by CRISPR-Cas genome editing. First, all 25 functional copies of PERV in the genome of cultured pig fetal fibroblast cells were targeted for inactivation using two sgRNAs that direct Cas9 to cleave the PERV *pol* gene (Fig. 12.34); this gene encodes a protease, reverse transcriptase, and integrase required for integration of viral genomes into a host chromosome. Loss of *pol* gene expression was confirmed by RNA sequencing, and reverse transcriptase activity of PERVs was not detected, suggesting that, in contrast to parental cells, viral genomes were not produced in the genome-edited pig fibroblast cells. Next, the nuclei of the engineered fibroblast cells were transferred into enucleated ova to generate PERV-inactivated pigs (Fig. 12.11). Prior to transfer into a surrogate sow, the *pol* genes of several embryos were sequenced to confirm the presence of small insertions or deletions that are a consequence of nonhomologous end-joining of Cas9-cleaved DNA. PERV inactivation in the piglets was verified by sequencing *pol* loci in the genome and *pol* mRNA. These pigs can serve as founder animals to generate animals that provide safe organs for xenotransplantation. The potential for pathogen transmission and all other likely complications must be resolved, and prolonged survival of nonhuman primate transplant recipients of a life-supporting organ must be demonstrated before xenografts are considered safe for human clinical trials.

Enhancing Production Traits of Food Animals

Disease-Resistant Livestock

Infectious diseases of domestic animals are currently controlled by vaccination, drugs, physical isolation, and careful monitoring. In practice, the cost of disease prevention can be as much as 20% of the total production value.

The development of genetically engineered animals with inherited resistance to bacterial, viral, and parasitic diseases would potentially decrease the use of drugs to treat these problems, increase productivity, provide safer foods derived from these animals, and increase economic benefits.

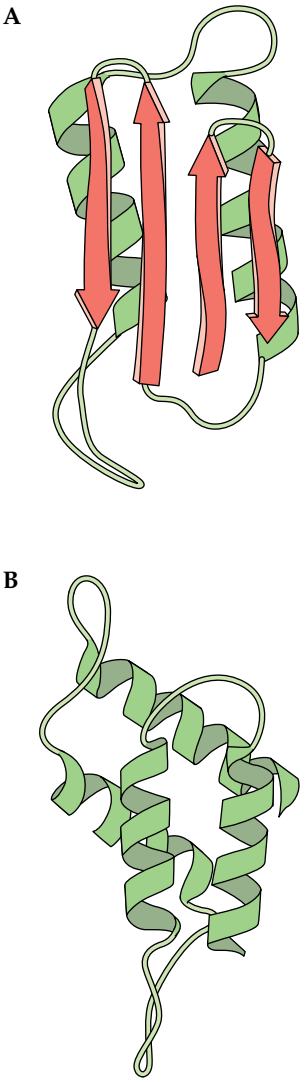
One approach to creating infectious disease-resistant transgenic animals entails introducing the gene for an antibody that binds to a pathogen-specific antigen. The rationale behind this strategy is to provide an inherited biological protection mechanism for the transgenic animal that eliminates the need for immunization by vaccination. Although the animal usually has an intact humoral immune system, expression of a monoclonal antibody against a specific pathogen would provide immediate protection without prior exposure to the pathogen. If the transgene encoding a monoclonal antibody is engineered to be secreted into milk, young suckling animals might acquire passive immunity against a pathogen.

Prion-Resistant Cattle

Host proteins that render an animal susceptible to a specific infectious disease can be eliminated if they are not essential for animal health. Prion diseases are caused by aberrant forms of normal brain proteins (Fig. 12.35). They are infectious in the sense that once they are acquired by a cell, the aberrant proteins induce the normal versions of the brain proteins to misfold. The misfolded proteins aggregate and disrupt brain function. One prion disease that is particularly problematic for the beef industry is bovine spongiform encephalopathy (BSE), caused by a mutant form of the protein known as prion protein BSE (Prp^{BSE}) to distinguish it from the normal protein, or Prp^{C}, a glycoprotein found on the surface of neurons of the central nervous system, where it plays a role in stress protection. Scrapie is a similar prion disease found in sheep. BSE, often referred to as mad cow disease for the neuropathological symptoms in cows, has caused huge economic losses for cattle and dairy farmers and is the motivation for contentious barriers to the trade of livestock between countries. There is no known treatment for the disease, and therefore, many animals have to be destroyed, usually by incineration. Moreover, there is some evidence that prions can be transmitted to humans through consumption of prion-contaminated meat products to cause a variant form of encephalopathy or Creutzfeld-Jakob disease (in humans the prion is referred to as Prp^{CJD}). This insidious disease can be asymptomatic for very long periods of time but can eventually manifest as mental deterioration. Thus, it would be a relief for farmers, and indeed consumers of animal products, if livestock could be protected from prion protein infection.

The potential of engineering resistance by abolishing production of the normal version of the protein Prp^{C} was tested first in mice and then in cattle in which both alleles of the *PRNP* gene encoding Prp^{C} were disrupted by insertion of an antibiotic-resistant gene into the coding sequences (Fig. 12.36). The genetically modified animals were assessed for a variety of morphological and physiological features, in particular, for the presence of features that are used to diagnose spongiform encephalopathy, including mental status, sensory and motor functions, immune function, and brain tissue morphology. The brain tissue of infected animals becomes filled with holes that give the brain a characteristic sponge-like appearance. In all aspects, the engineered cattle were found to be normal and remained normal for almost 2 years. Brain tissue homogenates were collected from wild-type and Prp^{C} knockout cattle and incubated with brain homogenates

Figure 12.35 Prion proteins are aberrant forms **(A)** of a normal brain protein, Prp^{C} **(B)**. Infection with prions induces the normal variants of the brain protein to misfold and aggregate, which disrupts normal brain function. Misfolded proteins can be transmitted to other animals and induce normal Prp^{C} proteins in those animals to misfold.



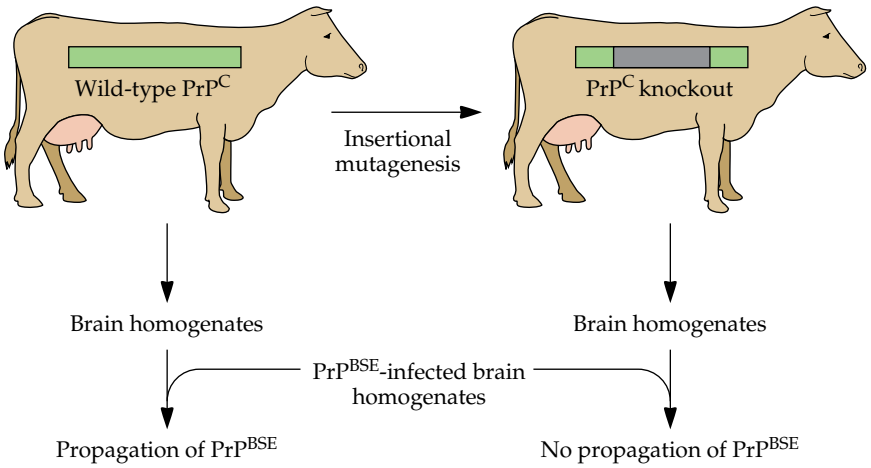


Figure 12.36 Normal brain protein Prp^C found in the brain of wild-type cattle is required for propagation of the aberrant form of the protein known as prion protein BSE (Prp^{BSE}), which causes bovine spongioform encephalopathy (BSE).

of BSE-infected cattle carrying the abnormal version of the prion protein Prp^{BSE} (Fig. 12.36). Propagation of Prp^{BSE} could not be detected in the homogenates of Prp^C knockout animals, while it was readily detected in the wild-type homogenates. These results indicate that loss of function of Prp^C does not cause BSE and that the normal Prp^C is required for propagation of the aberrant form of the prion protein and suggest that the genetically engineered Prp^C knockout cattle could be resistant to BSE infection.

Since this work was reported, other researchers have inactivated the *PRNP* gene in bovine fetal fibroblast cells and early embryos using genome-editing technologies. Successful disruption of the *PRNP* gene by direct injection of a plasmid or RNA encoding *PRNP*-targeting sgRNA and Cas9 endonuclease into the cytoplasm of single-cell bovine embryos demonstrated proof of principle. This method avoids introduction of antibiotic resistance genes and difficulties associated with somatic cell nuclear transfer, which was used to generate the knockout cattle described above. The genome-edited embryos were cultured *in vitro* to the blastocyst stage but were not implanted into a surrogate female for further development.

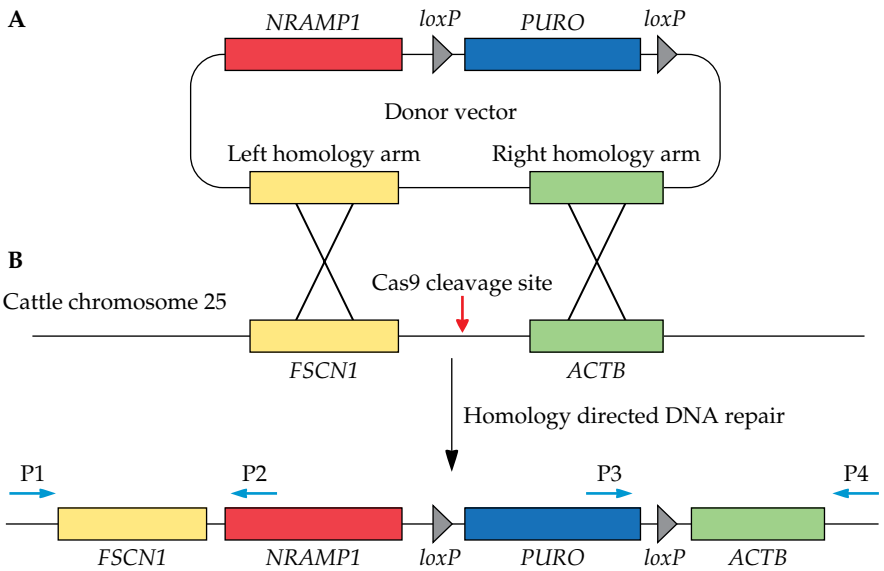
Tuberculosis-Resistant Cattle

Tuberculosis caused by the bacterium *Mycobacterium bovis* is a major problem in cattle, affecting more than 50 million cattle worldwide. *M. bovis* infects macrophages (a phagocytic white blood cell that engulfs bacteria and other foreign particles), and infection typically manifests as enlarged lesions (tubercles), primarily in the lungs and lymph nodes associated with the respiratory tract. Infection can remain latent without symptoms for many years and is reactivated in older or stressed animals. Despite efforts to control bovine tuberculosis by testing and culling infected animals from the herd, the disease persists in many regions, including the United Kingdom, China, Africa, and Latin America. *M. bovis* can also be transmitted to other animals, such as pigs, bison, and deer, and humans by contact with infected animals or consumption of unpasteurized dairy products, although the number of cases is low in humans, in whom most

cases of tuberculosis are caused by *Mycobacterium tuberculosis*. A vaccine to protect cattle against tuberculosis is available but has limited efficacy, and therefore it would be advantageous to engineer cattle to be resistant to *M. bovis* infection.

Genome editing was employed to create tuberculosis-resistant transgenic cattle. The natural resistance-associated macrophage protein-1, encoded by the *NRAMP1* gene, is associated with innate resistance to *M. bovis* and other intracellular pathogens. The natural resistance-associated macrophage protein-1 is expressed only in phagocytes and induces inflammation to limit infection. Overexpression of *NRAMP1* in macrophages, the site of *M. bovis* proliferation, was hypothesized to increase resistance to the bacterium. To this end, bovine *NRAMP1* cDNA, under the control of its natural promoter to restrict expression to phagocytes, was inserted into a donor vector between two homology arms that target the transgene to an intergenic region between two actin genes (Fig. 12.37A). The actin genes are constitutively expressed, and therefore a transgene inserted into this region is unlikely to be silenced by heterochromatin. A marker gene

Figure 12.37 Transgenesis to create tuberculosis-resistant cattle. **(A)** The donor vector carries the bovine *NRAMP1* gene encoding natural resistance-associated macrophage protein-1, which confers resistance to *M. bovis*, under the control of its native promoter, and the *PURO* selectable marker gene encoding resistance to puromycin [promoters and poly(A) signals are present for each gene but not shown]. Left and right homology arms mediate integration of the donor DNA into the cattle genome. **(B)** Cas9, guided by an sgRNA, cleaved at a sequence (red arrow) in the intergenic region between the fascin actin-bundling protein 1 gene (*FSCN1*) and the actin beta gene (*ACTB*) on chromosome 25 of the cattle genome. DNA repair by recombination (×) between homologous sequences resulted in integration of the *NRAMP1* and *PURO* genes into the genome. Primers P1 and P4, which anneal to sequences in chromosome 25 outside the homologous regions, were used in combination with primers P2 and P3, respectively, which anneal to sequences in the transgene construct, in PCRs to confirm integration of the donor DNA at the target site. The *loxP* sites enabled excision of the *PURO* gene by Cre recombinase after its integration. Data from Gao et al., *Genome Biol*. **18:**13, 2017.



encoding resistance to the antibiotic puromycin was included between the homology arms for selection of transfected cells and was flanked by *loxP* sequences in the same orientation for subsequent excision of the marker gene by Cre recombinase (Fig. 12.15B). Single-guide RNAs were designed to target the actin locus using an algorithm to predict and avoid similar, off-target sequences in the region. Bovine fetal fibroblast cells were transfected with the donor vector carrying the transgene and an expression plasmid encoding Cas9 endonuclease and the sgRNA and then selected on puromycin. PCR to amplify chromosomal DNA across the target site revealed that 35% of puromycin-resistant cells had correctly integrated the transgene via homologous recombination (Fig. 12.37B). The nuclei of confirmed transgenic fetal fibroblast cells were transferred to enucleated eggs for generation of transgenic animals by somatic cell nuclear transfer (Fig. 12.11).

Natural resistance-associated macrophage protein-1 was expressed from the transgene in macrophages, and not heart, liver, lungs, kidneys, skin, or muscles, of the *NRAMP1* transgenic cattle, as expected. The growth rate of *M. bovis* in macrophages isolated from transgenic animals was significantly lower than in nontransgenic macrophages. Transgenic cattle challenged with *M. bovis* via the bronchus (lower lungs) were much less susceptible to infection than nontransgenic animals as determined by lower levels of *M. bovis*-specific interferon-γ secretion from lymphocytes in an enzyme-linked immunosorbent assay (ELISA) routinely used for cattle tuberculosis testing.

Mastitis-Resistant Cows

The bacterium *Staphylococcus aureus* is responsible for 25% of the cases of mammary gland infection (mastitis) in cows. These infections are contagious, recur frequently following termination of antibiotic treatment, and readily spread through an entire herd. Milk yields from infected cows are significantly reduced. Currently, outbreaks of mastitis caused by *S. aureus* cannot be effectively controlled, with an annual cost of about $2 billion in the United States. Transgenic cows have been produced that secrete a staphylolytic agent into their milk to prevent the infections. *Staphylococcus simulans* produces lysostaphin, a peptidoglycan hydrolase that specifically attacks the cell wall of *S. aureus*. However, when cultured eukaryotic cells were initially transfected with the bacterial lysostaphin gene, only inactive lysostaphin was produced, because two asparagines are abnormally glycosylated. This problem was overcome by using directed mutagenesis (see chapter 3) to replace the codons for these two asparagines with those for glutamine (Fig. 12.38). The modified lysostaphin was not glycosylated after synthesis in eukaryotic cells and was fully active against *S. aureus*.

The efficacy of this approach was first tested by engineering mice to express the altered lysostaphin gene under the control of the promoter of the sheep β-lactoglobulin gene, whose protein product is secreted into milk. Based on the successful protection of the transgenic mice against large inocula of *S. aureus*, the approach was extended to produce cows that express the lysostaphin transgene. The altered lysostaphin gene under the control of the sheep β-lactoglobulin promoter was introduced into cow fibroblast cells, and the nuclei from these cells were then transferred to enucleated cow oocytes. Blastocysts were implanted into the uterus of cows, and several calves were subsequently born and used to establish transgenic lines. After nine infusions of *S. aureus* into the mammary glands of high-lysostaphin-expressing cows, no infections were observed. In contrast, 71%

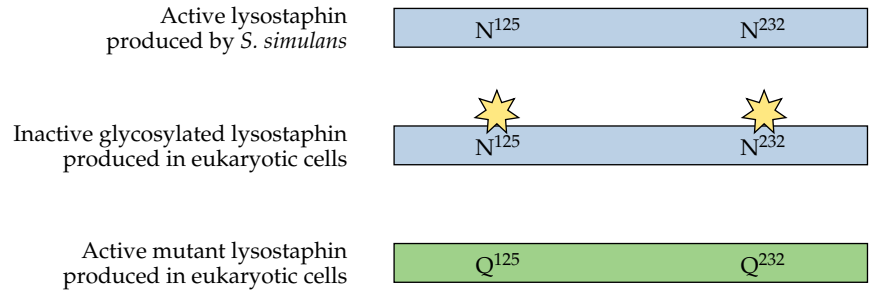


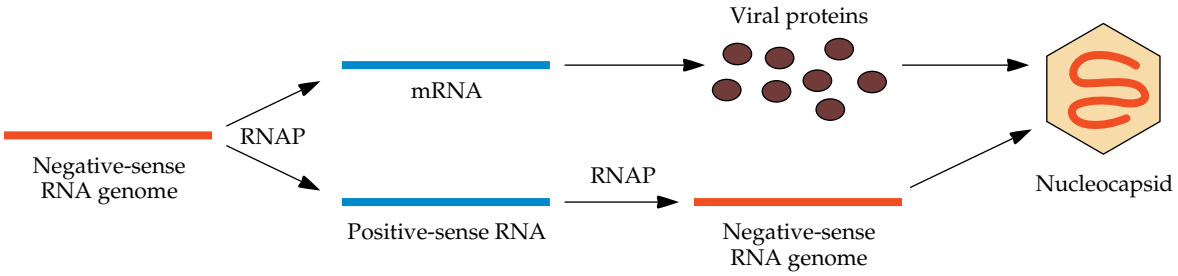
Figure 12.38 Native lysostaphin produced by *S. simulans* is a peptidoglycan hydrolase that cleaves the cell wall of the animal pathogen *S. aureus*. In animal cells transfected with the lysostaphin coding sequence from *S. simulans*, lysostaphin is glycosylated (yellow stars) and inactive against *S. aureus*. Site-directed mutagenesis of the lysostaphin gene to replace the coding sequences for two asparagines in amino acid positions 125 and 232 (N^{125} and N^{232}) with glutamines (Q^{125} and Q^{232}) results in production of nonglycosylated, active lysostaphin in animal cells.

of the infused mammary glands of nontransgenic animals were infected. Even low levels of lysostaphin expression afforded a significant level of protection against the pathogen. While these results show promise for solving an important problem for the beef and dairy industries, the safety of milk containing lysostaphin, such as the potential to alter the normal intestinal microflora of consumers and to select for lysostaphin-resistant strains of pathogenic *S. aureus*, needs to be investigated.

Influenza-Resistant Chickens

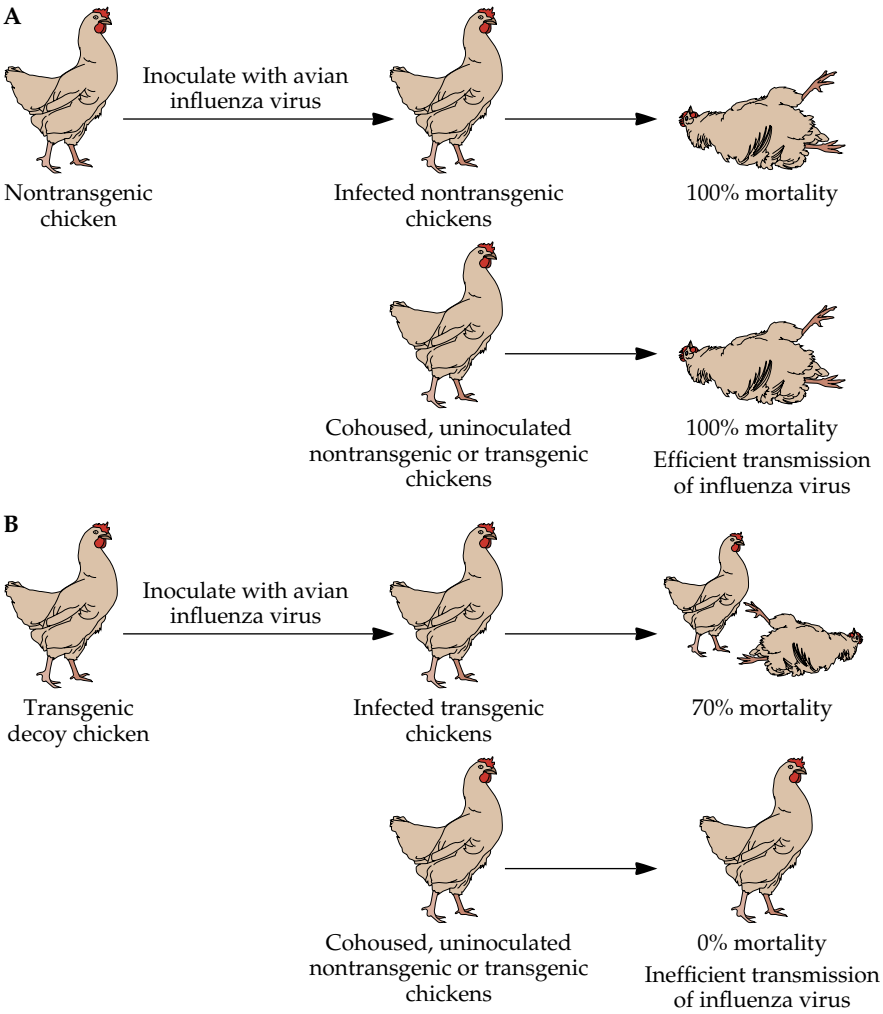
One important application for chicken transgenesis is prevention of the spread of avian influenza virus. Avian influenza virus primarily infects wild birds, but it is highly contagious and can infect domestic birds, including chickens, from which it may be transmitted to humans. Influenza viruses have segmented RNA genomes with eight separate strands that require an RNA-dependent RNA polymerase for genome replication and synthesis of viral proteins (Fig. 12.39). The RNA polymerase binds to a specific sequence found on each of the eight RNA molecules that make up the viral genome and catalyzes RNA synthesis. It was reasoned that if transgenic chickens expressed a small RNA molecule that acts as a decoy by binding to and inhibiting the activity of the RNA polymerase, synthesis of new virions that can spread to other animals would be prevented.

Figure 12.39 Production of mRNA and genomic RNA molecules in negative-sense single-stranded RNA viruses such as influenza virus requires an RNA-dependent RNA polymerase (RNAP).



As proof of principle, chicken embryos were microinjected with a lentiviral vector encoding the RNA polymerase binding site under the control of a chicken promoter. A rooster carrying the transgene in its reproductive cells was crossed with nontransgenic hens to produce heterozygous transgenic offspring that were crossed to produce homozygous transgenic chickens. All nontransgenic chickens directly challenged with avian influenza virus died from the infection within 6 days postinfection (Fig. 12.40). All transgenic and nontransgenic birds cohoused with the infected nontransgenic chickens, but not directly inoculated with the virus, also succumbed to influenza infection, indicating efficient transmission

Figure 12.40 Mortality of nontransgenic and transgenic decoy chickens following infection with avian influenza virus. The transgenic chickens produce a small RNA molecule that inhibits RNA-dependent RNA polymerase required for production of new virions. All of the uninoculated chickens (nontransgenic or transgenic) that were in contact with the infected nontransgenic chickens succumbed to influenza infection, indicating efficient transmission of the virus **(A)**. In contrast, none of the uninoculated chickens that were in contact with the infected transgenic chickens died from influenza infection, indicating inefficient transmission of the virus **(B)**. Data from Lyall et al., *Science* **331:**223, 2011.



of influenza virus. In contrast, 4 of the 10 directly infected transgenic chickens survived for 7 days, and 2 were healthy to the end of the 10-day study. Significantly, none of the uninoculated birds that were cohoused with infected transgenic birds died from influenza infection, and none were positive for the viral antigen. High levels of viral RNA were present in 100% of buccal and 90% of cloacal samples from the nontransgenic chickens, indicating viral shedding, compared to 50% of buccal and 20% of cloacal samples from transgenic birds. These results support the inefficient transmission of the virus from the transgenic chickens. Genetic resistance to influenza is preferable to immunization with vaccines that have to be administered frequently due to the high mutation rate of the influenza virus. Moreover, to overcome the inhibitory effect of the RNA decoy and replicate in the transgenic chickens, the influenza virus would have to mutate both the RNA polymerase gene and its binding site on all eight chromosomes, an extremely rare possibility.

Improving Milk Quality

One of the goals of transgenesis of dairy animals is to enhance the health-promoting qualities of milk for humans and for suckling animals. Overexpression of proteins involved in the production of milk nutrients, for example, can improve the growth, health, and survival of young animals. Specific components of milk can also be altered to benefit human consumers.

The antibacterial and immunomodulatory protein lysozyme is naturally produced in high levels in human breast milk, but much lower levels are present in the milk of ruminants such as cows and goats. Lysozyme cleaves glycosidic bonds in peptidoglycan that leads to the rupturing of bacterial cell walls. The presence of lysozyme in human breast milk contributes to a decrease in gastrointestinal infections in breastfed infants and allows the growth of beneficial bacteria such as *Lactobacillus* and *Bifidobacteria*. Consumption of this protein in animal milk, a major food source for many young children, could help to reduce the high incidence of diarrheal illnesses in children in many regions of the world and in suckling animals. Expression of the gene encoding human lysozyme from the α-casein promoter in the mammary glands of transgenic goats resulted in milk with 68% of the lysozyme levels of human milk. Piglets fed pasteurized milk from the transgenic goats recovered more rapidly from infection with enterotoxigenic *E. coli* than those fed milk from nontransgenic goats. They had a more solid fecal consistency, were less dehydrated and more active, and cleared the *E. coli* infection faster than those fed milk from control goats, which exhibited greater intestinal inflammation and damage. Clinical trials are needed to test the safety and efficacy of the lysozyme-enriched transgenic goat milk as a prophylactic for diarrheal illnesses that kill more than one million children every year.

Many people who are allergic to bovine milk would benefit from the abolition of β-lactoglobulin, a major allergen in milk. The feasibility of this approach was demonstrated in cows using zinc-finger nucleases. Engineered zinc-finger nucleases contain a DNA-binding domain that targets the endonuclease domain to a specific sequence for cleavage (chapter 6, Fig. 6.17). The endonuclease makes a double-stranded break in the target DNA that may be repaired by nonhomologous end-joining, which often results in insertion and/or deletion of nucleotides that disrupt

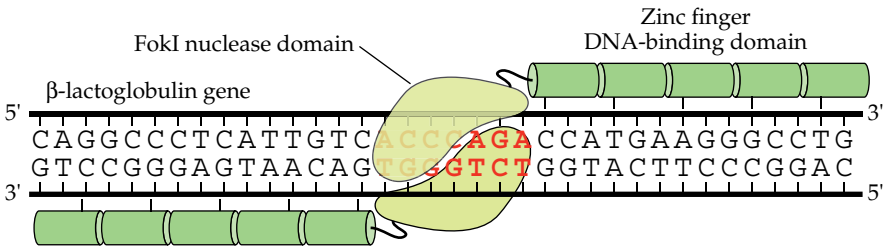


Figure 12.41 Zinc-finger nuclease designed to disrupt the bovine β-lactoglobulin gene. The DNA-binding domain binds to a specific nucleotide sequence in the β-lactoglobulin gene and positions the FokI nuclease domain to cleave a target site (red). Repair of the cut DNA by nonhomologous end-joining introduced frame-shift mutations into the gene. Data from Yu et al., *Cell Res*. **21:**1638, 2011.

the reading frame of protein-coding sequences. To abolish β-lactoglobulin production, bovine fetal fibroblast cells were transfected with mRNA encoding a zinc-finger nuclease designed to bind to and cut within a sequence in the β-lactoglobulin gene (Fig. 12.41). PCR analysis showed that 80% of the mutant fibroblast cells had small (<20 bp) deletions or insertions in the target sequence. Engineered cows were generated by somatic cell nuclear transfer, using the nuclei from fibroblast cells confirmed to carry mutations in both alleles of the β-lactoglobulin gene. Of eight cloned animals born, only one survived. The surviving animal has deletions of 9 and 15 bp in the β-lactoglobulin alleles, which are not likely to disrupt the reading frame but will produce a truncated protein. This preliminary result demonstrates the targeted disruption of a gene encoding a bovine milk protein using a genome editing nuclease; however, it remains to be determined whether the milk from these animals is tolerated by individuals who are allergic to β-lactoglobulin.

Increasing Muscle Mass in Cattle

Improving production traits such as muscle mass in meat animals is often more difficult than engineering animals to alter milk composition because multiple genes are involved in controlling growth, and a detailed understanding of the genetic basis for traits of interest is required. Initially, researchers sought to increase the body mass of livestock by introducing genes encoding growth hormones or insulin-like growth factor. Although these early efforts yielded animals with increased ability to convert feed into body mass, they were hampered by difficulties in controlling the expression of the transgenes. Overproduction of growth hormone was found to adversely affect the animals' health, which manifested as gastric ulceration, kidney dysfunction, lameness, inflammation of the lining of the heart, immobility of the joints, and susceptibility to pneumonia; the reasons for these symptoms are not known. To overcome this problem, inducible promoters were used, such as the metallothionein promoter that can be activated by zinc administered in the animal's diet; however, expression was often poor because the transgene was incorporated into heterochromatin, transcriptionally silent regions of the genome.

Some cattle breeds, such as Belgian blue, have larger and leaner muscle mass than other breeds because of a naturally occurring mutation (11-bp deletion) in the *MSTN* gene encoding myostatin, a growth factor that

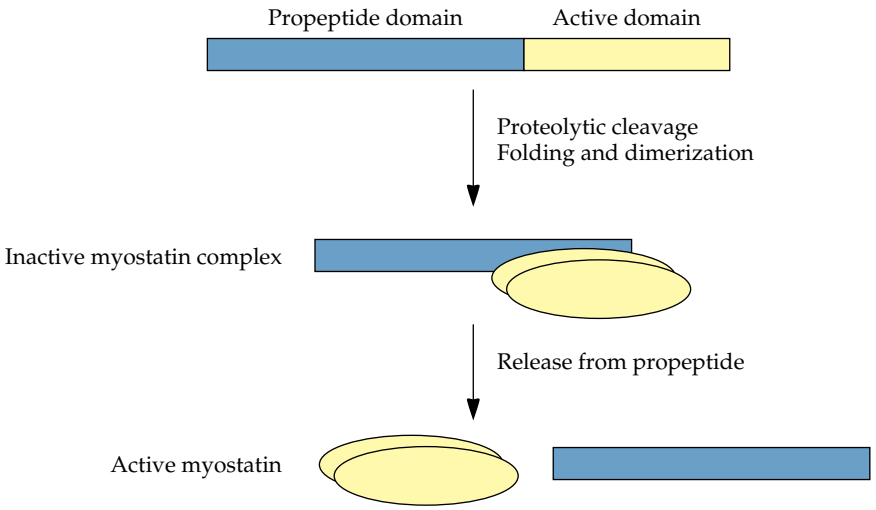


Figure 12.42 Myostatin consists of an N-terminal propeptide domain (latency-associated peptide, LAP) and an active C-terminal domain. Proteolytic cleavage and folding yields an active C-terminal dimer. Following proteolytic cleavage, the C-terminal dimer can form an inactive (latent) complex with the propeptide domain.

inhibits growth of skeletal muscle. Interestingly, humans with rare mutations that disrupt the function of myostatin are also unusually muscular, and dogs of the whippet breed that have enhanced muscle mass, known as "double muscling," and increased racing speed, carry mutations in both copies of the *MSTN* gene. These mutations result in production of a truncated, nonfunctional myostatin protein. Some researchers have reasoned that other cattle breeds can be engineered to increase meat production by disrupting myostatin function. Moreover, if disruption could be limited to male animals, then milk production in females would be unaffected. This approach was initially tested in mice, in which a DNA fragment encoding a myostatin inactivator was targeted to the Y chromosome so that it would be expressed only in male animals. The myostatin inactivator consists of the N-terminal propeptide domain (latency-associated peptide; LAP) of the myostatin protein. Following proteolytic cleavage, the N-terminal propeptide can sequester, and thereby inactivate, the C-terminal portion of the protein, which is the biologically active component (Fig. 12.42). This blocks myostatin activity that would normally lead to inhibition of muscle growth.

A two-step strategy was used to target the gene encoding the myostatin inactivator to the mouse Y chromosome. In the first step, a cassette containing positive (neomycin resistance gene *neo*) and negative (HSV-*tk*) selectable markers flanked by a wild-type and a sequence variant *loxP* site was inserted into a nonessential region of the Y chromosome of mouse embryonic stem cells by homologous recombination (Fig. 12.43A). Transfected cells were selected for resistance to the antibiotic G-418 conferred by the *neo* gene, and the integration site in the Y chromosome was confirmed by PCR. In the second step, a gene encoding the myostatin inactivator, under the control of a strong rat skeletal muscle promoter and enhancer, and also flanked by the same heterologous *loxP* sites, was cloned into a plasmid and introduced by electroporation into the transfected embryonic stem cells carrying the selectable marker cassette (Fig. 12.43B). A second plasmid encoding Cre recombinase was also introduced into the transfected cells. Cre recombinase activity

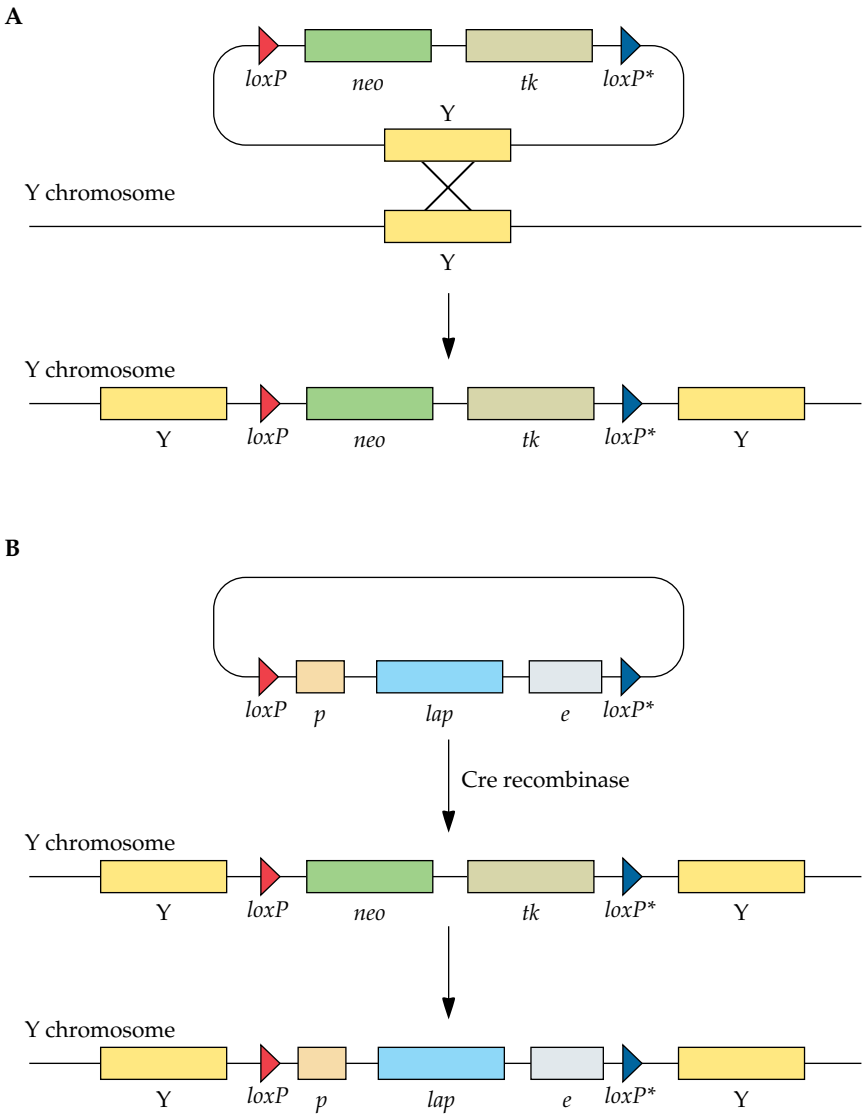


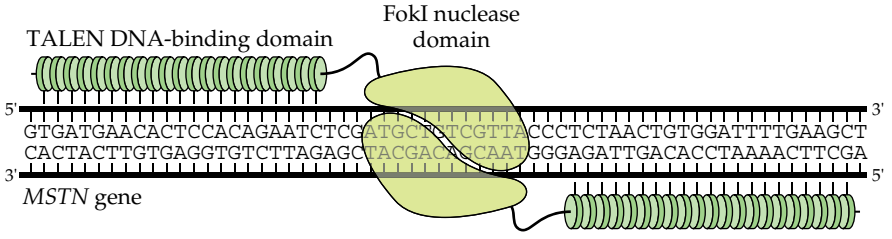
Figure 12.43 Strategy to generate transgenic mice with increased muscle mass. The gene encoding the myostatin inhibitor, which consists of the N-terminal propeptide domain (LAP) of myostatin, was introduced into the Y chromosome of mouse embryonic stem cells using a two-step procedure. **(A)** In the first step, a cassette carrying genes encoding selectable markers (*neo* and *tk*) flanked by a wild-type *loxP* site (red arrow) and a mutant *loxP** site with two nucleotide changes (blue arrow) was introduced into the Y chromosome by recombination (×) between a nonessential sequence in the Y chromosome (Y) and a homologous sequence cloned into the vector. The positive selectable marker (*neo*) confers resistance to the antibiotic G-416, and the negative selectable marker (*tk*) confers sensitivity to ganciclovir. **(B)** In the second step, the selectable markers in the Y chromosome were replaced with a cassette carrying the myostatin inhibitor (*lap*), under the control of the rat skeletal muscle promoter (*p*) and enhancer (*e*), by Cre recombinase-mediated exchange at homologous *loxP* sites. Adapted from Pirottin et al., *Proc. Natl. Acad. Sci. USA* **102:**6413, 2005.

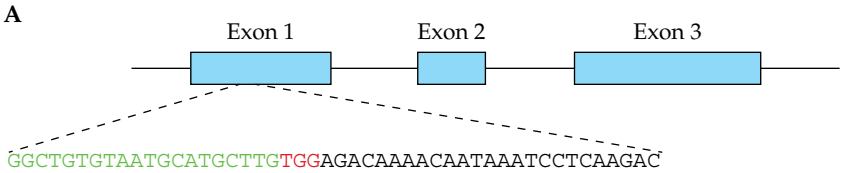
resulted in recombination between homologous *loxP* sites and integration of the myostatin inactivator gene into the Y chromosome (Fig. 12.43B). Exchange of the myostatin inactivator DNA sequence for the selectable

marker gene cassette was initially selected by resistance to ganciclovir, indicating loss of the *tk* gene encoding thymidine kinase that converts ganciclovir into toxic compounds, and then confirmed by PCR. Transfected embryonic stem cells carrying the myostatin inhibitor gene were cultured, inserted into blastocysts, and then implanted into pseudopregnant foster mothers. A single transgenic male founder mouse was obtained and then mated with nontransgenic females. As expected for a gene integrated into the Y chromosome, all male offspring carried the myostatin inhibitor transgene, whereas none of the female offspring were transgenic. In the transgenic males, the myostatin inhibitor was expressed in skeletal muscle tissue and not in heart or liver tissue. Muscle mass was greater by 5% to 20% in transgenic males than in nontransgenic male control mice.

To inactivate myostatin and increase meat yields in livestock animals, the *MSTN* gene was disrupted using genome-editing nucleases. In one case, a transcription activator-like effector nuclease (TALEN) was employed. A TALEN is an engineered protein that is similar to a zinc-finger nuclease (chapter 6) in that it consists of a nonspecific FokI nuclease domain fused to a DNA-binding domain that guides the nuclease to a target sequence. In a TALEN, the DNA-binding domain consists of series of 34-amino-acid modules that each recognize and bind to a single specific nucleotide, compared to three nucleotides for a module of a zinc-finger nuclease. Sequence-specific TALENs are synthesized by combining the coding sequences for these modules and the nuclease domain in an expression vector. The nuclease generates double-stranded breaks in target DNA that may be repaired by nonhomologous end-joining, which tends to introduce missense mutations that disrupt the function of a gene. To inactivate myostatin, a TALEN was constructed to cleave within a region of the bovine *MSTN* gene in which deletions were previously determined to cause the double muscling phenotype in Belgian blue cattle (Fig. 12.44). TALEN mRNA was injected directly into fertilized oocytes of cows and sheep for production of the *MSTN*-targeting nuclease. After 7 days, embryos were analyzed for the desired mutations before transfer to recipient females for further development. Two live births resulted from 20 transferred bovine embryos; however, only one of the offspring carried edited *MSTN* alleles. Three different alleles (with 1- and 3-bp deletions, and a wild-type allele) were detected in the bull, indicating that the TALEN was active for more than one cell division. Despite the mosaicism, the bull exhibited enhanced muscle development. From 26 transferred ovine embryos, 9 surviving offspring resulted. Only one of the lambs carried a single edited *MSTN* allele (the other

Figure 12.44 A TALEN to introduce mutations into the bovine *MSTN* gene encoding myostatin. Abolition of myostatin increases muscle mass and, thereby, meat yields in cattle.





B

WT	CTGTGTAATGCATGCTTGTGGAGACAAAACAATAAATCCTCAAGAC	LCNABLWRQNN
Ins	CTGTGTAATGCATGCATGTAATGCATGTTTTGTGGAGACAAAACAA	LCNACM*
Ins	CTGTGTAATGCATGCATTTGTGGAGACAAAACAATAAATCCTTAAG	LCNACICGDKTINPQD*
Del	CTGTGTAA.....ACAATAAATCCTCAAGAC	LCKQ*
Del	CTGTGTAATGCATG.....GAGACAAAACAATAAATCCTCAAGAC	LCNA..WRQNN
Del+Ins	CTGTGTAATGCATGCTTTTGTGGAGACAAAACAATAAATCCTCAAC	LCNABFVETLQ*
Del+Ins	CTGTGTAATGCATGCT.....GAAACAAAACAATAAATCCTGAA	LCNAC*

Figure 12.45 Disruption of the ovine *MSTN* gene using the CRISPR-Cas system. **(A)** An sgRNA was microinjected into one-cell embryos to guide Cas9 to a target sequence (green) downstream of a PAM (red) within exon 1 of the *MSTN* gene. **(B)** Some of the mutations introduced into exon 1 following cleavage by Cas9 and DNA repair by nonhomologous end-joining. Inserted nucleotides are highlighted in blue, deletions are represented as dots, and introduced stop codons are underlined. The effect on the translated sequence is shown with introduced amino acids highlighted in blue and premature termination of translation indicated by an asterisk. From Crispo et al., *PLoS ONE* **10**(8)**:**e0136690, 2015.

allele was wild type), and interestingly, this was the same 3-bp deletion found in the bull. It is not yet known whether this single codon deletion alters the function of myostatin.

The *MSTN* gene of sheep was also successfully edited using the CRISPR-Cas9 system. Cas9 mRNA and sgRNA targeting a sequence adjacent to a PAM in exon 1 in *MSTN* (Fig. 12.45A) were microinjected into the cytoplasm of one-cell ovine embryos and did not have any adverse effects on early embryo development *in vitro*. From 53 blastocysts transferred to recipient females, 22 lambs were born, although 3 died within the first day after delivery. Sequencing of *MSTN* exon 1 in skin and muscle tissue from the surviving lambs indicated that 10 animals carried mutations in one or both alleles. Most of the mutations were small (up to 27 bp) insertions and/or deletions that disrupted the reading frame of exon 1; however, three were in-frame deletions of 6 bp (Fig. 12.45B). In nine animals, a premature stop codon was introduced into the reading frame of at least one allele. None of the lambs exhibited mosaicism, and off-target mutations were detected in only two animals. Animals with biallelic mutations did not express myostatin in muscle fibers and at 2 months old were 20% to 30% heavier than wild-type lambs. Animals with only one mutant allele produced myostatin at similar levels to wild-type animals and did not exhibit increased body weight. This research demonstrates the feasibility of using the CRISPR-Cas system to efficiently introduce enhanced production traits in livestock animals.

Enhancing Growth of Salmon

As natural fisheries become depleted, production of this worldwide food resource has come to depend more heavily on aquaculture. In this context, enhanced growth rates, tolerance to environmental stress, and resistance

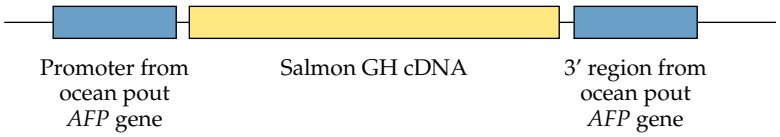


Figure 12.46 "All-fish" construct used to generate Atlantic salmon with enhanced growth rate. Expression of the growth hormone (GH) cDNA from Chinook salmon is controlled by the promoter and transcription terminator-polyadenylation signals from the antifreeze protein (*AFP*) gene from the fish ocean pout.

to diseases are some of the features that may be created by transgenesis. Transgenes have been introduced by microinjection or electroporation of DNA into the fertilized eggs of a number of fish species, including carp, catfish, trout, salmon, Arctic char, tilapia, and zebrafish. The pronuclei of fish are not readily seen under a microscope after fertilization; therefore, linearized transgene DNA is microinjected into the cytoplasm of either fertilized eggs or embryos that have reached the four-cell stage of development. Unlike mammalian embryogenesis, fish egg development is external; hence, there is no need for an implantation procedure. Development of transgenic fish occurs in temperature-regulated holding tanks. The survival of fish embryos after DNA microinjection is high (35% to 80%), and the production of transgenic fish ranges from 10% to 70%. The presence of a transgene is determined by PCR analysis of either nucleated erythrocytes or scale DNA. Founder fish are mated to establish true-breeding transgenic lines.

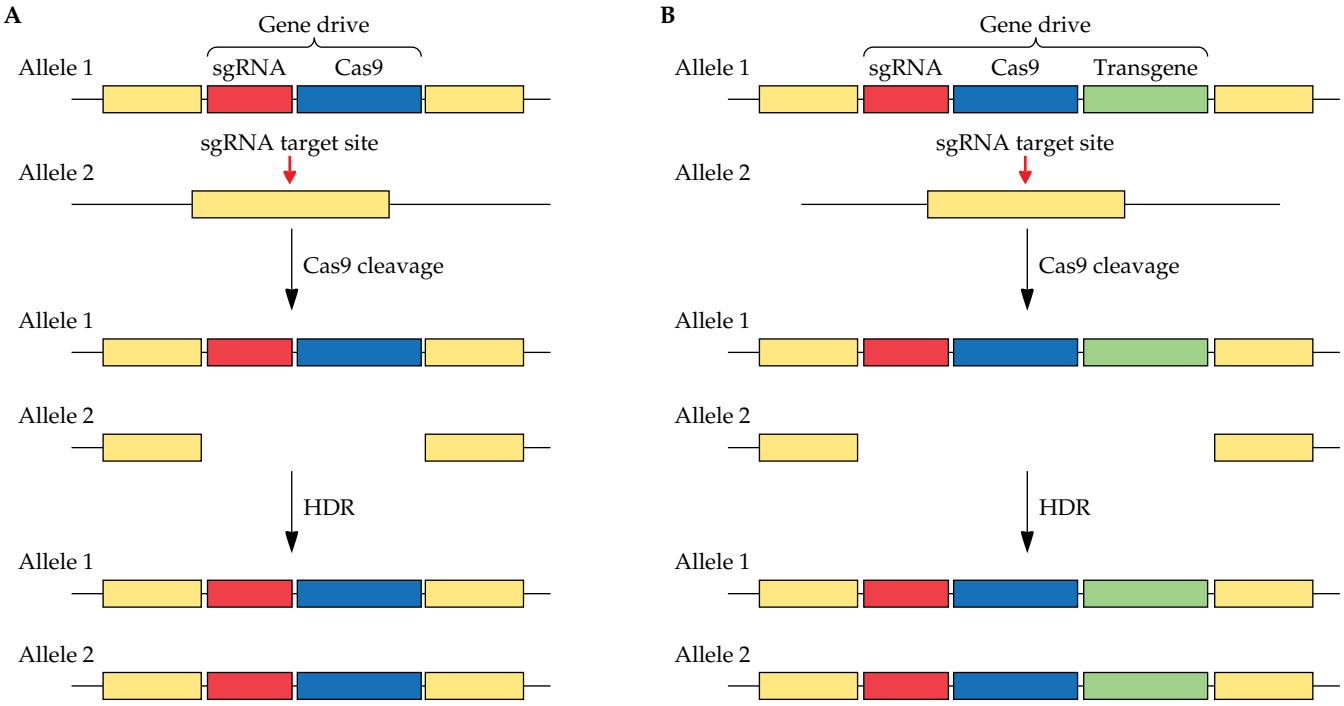
In 2015, the U.S. FDA approved the first genetically modified animal for human consumption, a transgenic Atlantic salmon that overproduces a growth hormone to enhance its growth rate. The transgene is derived from Chinook salmon growth hormone cDNA and is expressed from the promoter and the termination–polyadenylation signals from the antifreeze protein gene of the fish called ocean pout (Fig. 12.46). The "all-fish" construct was assembled to allay concerns about possible biological incompatibilities that might arise from using a growth hormone gene from nonfish sources. The ocean pout expression system was chosen to enhance the transcription of the growth hormone in cold waters. Young transgenic salmon grow more rapidly than nontransgenic salmon, reaching market size in half the usual time and using 25% less feed, which lowers the cost of production. Aquaculture with transgenic fish must be carried out in contained, land-based facilities to prevent the escape of transgenic fish into the wild. In addition, biological containment is achieved by raising only triploid female fish that are unable to breed with natural populations of diploid fish in the event of the accidental release of the transgenic salmon.

Gene Drives To Eradicate Vector-Transmitted Diseases

A gene drive is a selfish genetic element that propagates in a population by biasing inheritance in its favor. Gene drives can be harnessed to spread specific mutations or heterologous genes, and the gene drive itself, through a population. If the mutation or transgene introduced by a gene drive prevents an animal vector from transmitting human pathogens, then vector-borne infectious diseases may be eradicated. A common gene drive consists of genes encoding Cas9 endonuclease and sgRNA (Fig. 12.47A), and

in some cases, a transgene ("cargo" gene) (Fig. 12.47B). To propagate an inactivated gene, a transgenic animal is created with the gene drive inserted into one copy of a gene such that it disrupts the function of the gene. When expressed, the sgRNA directs Cas9 to cleave a target sequence in the other copy of the gene (Fig. 12.47). By homology-directed repair, the entire gene drive is copied to the other chromosome at the cleavage site, and both copies of the gene are nonfunctional. Normally, diploid animals pass one of their two alleles of a gene to their offspring. If one of the alleles carries a mutation, offspring have a 50% chance of inheriting the mutation (Fig. 12.48A). In contrast, an animal engineered to carry a gene drive that is expressed in germ cells and modifies a target sequence in the genome will pass the genetic modification to all of its offspring (Fig. 12.48B). When this animal is released into a population and mates with a wild-type animal, all offspring will inherit the gene drive, which will modify both copies of a target gene. Especially in animals that have a short generation time, such as insects, the mutation rapidly spreads in a population. In a similar manner, a transgene included on the gene drive can be transmitted to

Figure 12.47 Genetic modification with gene drives. **(A)** To disrupt the function of an endogenous gene (yellow boxes), a gene drive consisting of genes encoding an sgRNA and Cas9 endonuclease is inserted into one copy (allele 1) of the gene in an animal genome. When expressed, the sgRNA guides Cas9 to cleave at a target site within the other copy of the gene (allele 2). During homology-directed repair (HDR) of the double-stranded break, DNA synthesis extends the cleaved strands using the intact homologous chromosome as a template. The result is a transgenic animal that is homozygous for the mutation. **(B)** To generate a transgenic animal that is homozygous for one or more transgenes, the transgene is included on the gene drive.



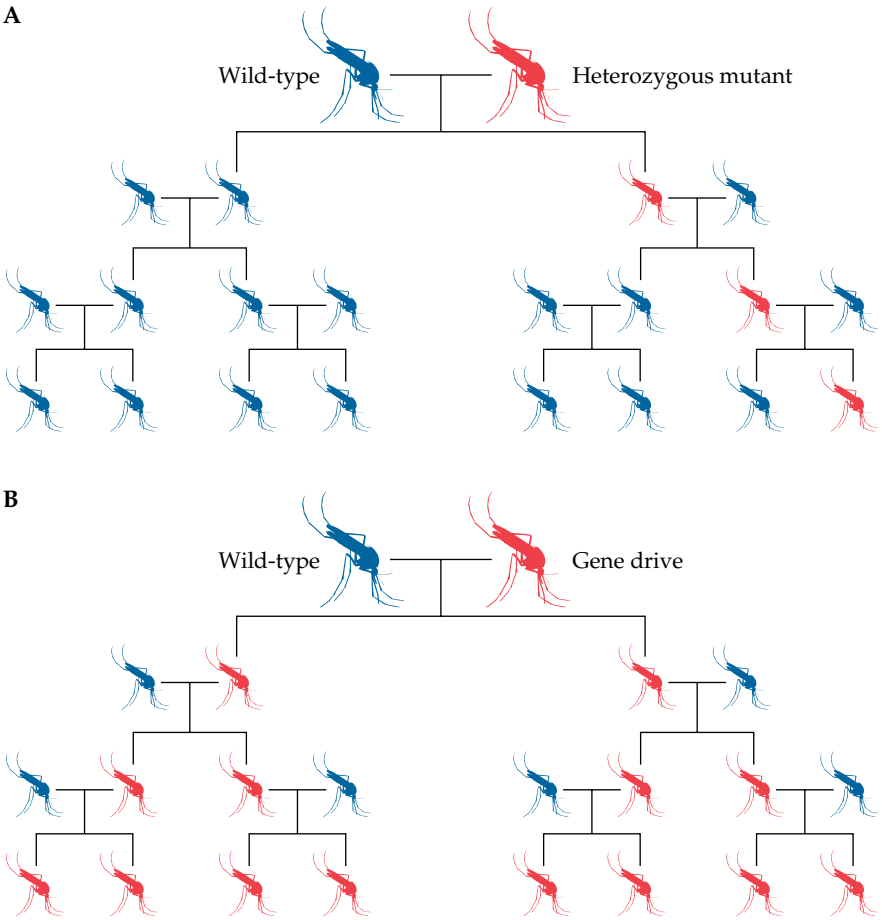


Figure 12.48 Inheritance of a gene drive. **(A)** In normal inheritance, when an animal that is heterozygous for a mutation mates with a wild-type individual, the offspring have a 50% chance of inheriting the mutation. **(B)** An animal that carries a gene drive is homozygous for a specific mutation or transgene, and therefore nearly 100% of the offspring of a mating with a wild-type individual will carry the gene drive. Mutations or transgenes introduced on a gene drive quickly spread through a wild population.

a population. Specific mutations or transgenes may be introduced into a population of insects that are vectors for human pathogens to prevent the transmission of infectious diseases such as mosquito-borne malaria or dengue fever. Contained laboratory experiments have established proof-of-principle; however, to date, transgenic insects carrying gene drives have not been released into wild populations.

Malaria Vector Population Suppression

Malaria is a major infectious disease that threatens half of the world's population. It is caused by several species of *Plasmodium* parasites and transmitted by mosquitoes (chapter 7, Fig. 7.20). Antimalarial drugs are available for prevention and treatment, although currently, the most successful efforts to reduce malaria are by using bed nets impregnated with

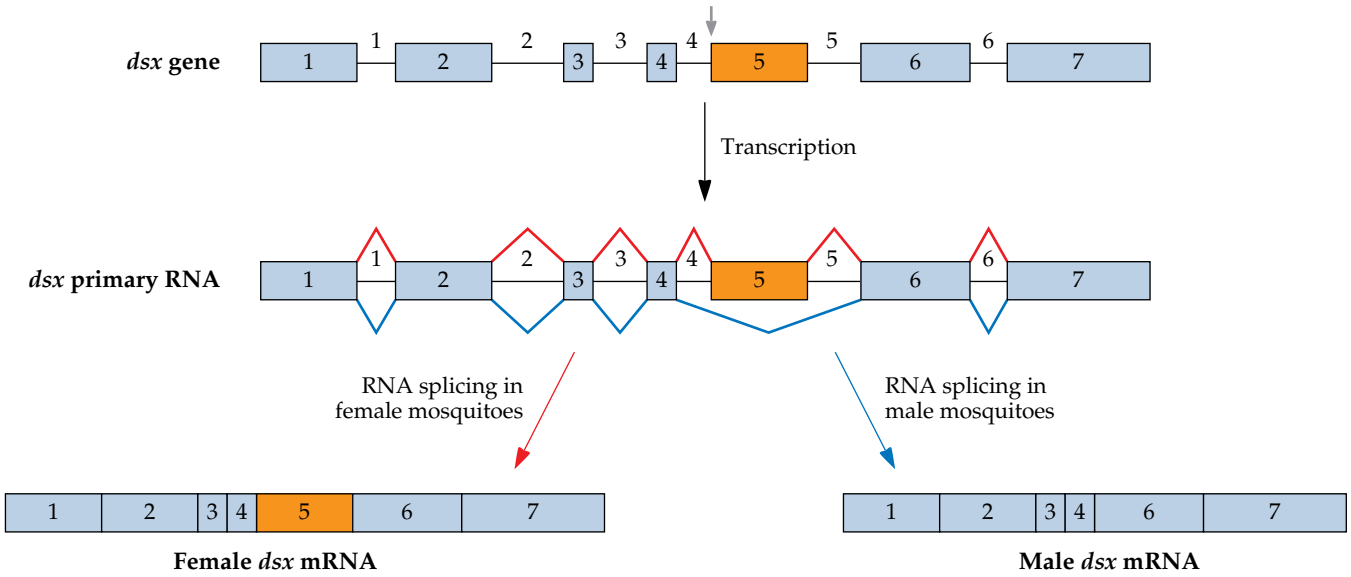


Figure 12.49 Mosquito population suppression with a gene drive that disrupts RNA splicing of the *doublesex* gene (*dsx*) in females. Normally, removal of introns (black lines) during splicing of the *dsx* primary transcript in female mosquitoes (red lines) generates mRNA containing exons 1 to 7 (numbered colored boxes). In male mosquitoes, RNA splicing (blue lines) results in excision of exon 5 from the mRNA. Insertion of a gene drive at the boundary of intron 4 and exon 5 in the *dsx* gene (gray arrow) prevents formation of female *dsx* mRNA and egg production in homozygous female mosquitoes, causing the population to collapse after several generations.

pyrethroid insecticides to prevent most bites from *Anopheles* mosquitoes, which tend to feed at night. However, there is evidence that pyrethroid resistance is spreading among mosquitoes, which threatens global programs to eliminate malaria. A malaria vaccine was introduced in 2019, although its efficacy is low, especially in young children, who account for 67% of deaths from malaria (Fig. 7.22).

An alternate method proposed to eradicate malaria is to suppress mosquito populations that transmit the *Plasmodium* parasite. A gene drive was constructed to prevent *Anopheles gambiae*, a mosquito species that commonly transmits malaria, from reproducing. The gene drive was designed to insert into the *doublesex* (*dsx*) gene that encodes two alternatively spliced transcripts involved in mosquito sex determination (Fig. 12.49). In female and male mosquitoes, the 5′ and 3′ ends of the *dsx* mRNA are similar and contain exons 1 to 4 and 6 to 7; however, female splice variants additionally carry exon 5, which is excised in male splice variants. The *dsx* mRNA variant found in females encodes a transcription factor that activates female gene expression, whereas the male variant blocks female gene expression. The sgRNA encoded on the gene drive targets the *dsx* gene sequence encoding the boundary between intron 4 and exon 5 (Fig. 12.49). Mutations at this site disrupt exon 5 and prevent the formation of female *dsx* mRNA, although male *dsx* mRNA is not affected. The sgRNA and Cas9 were expressed from a germline-specific promoter during meiosis to disrupt *dsx* at exon 5 and copy the

gene drive. Rates of transmission of the gene drive to offspring were 96% to 100%. Starting with a caged population of 300 wild-type female mosquitoes, 150 wild-type male mosquitoes, and 150 male mosquitoes carrying a single copy of the gene drive, the population collapsed in 8 to 12 generations. While female mosquitoes heterozygous for the gene drive are fertile, homozygous females did not bite anaesthetized mice (provided in test cages) to obtain a bloodmeal and are sterile (do not produce eggs). Because the *dsx* gene is essential for reproduction, it has high resistance to mutations that would prevent recognition by the sgRNA which targets a highly conserved region of *dsx*. Mutations in the sgRNA target site, whether from natural mutation mechanisms or from cleavage repair by nonhomologous end joining, will rise in frequency if they are favored by natural selection and are a major cause of failure of this technology.

Dengue Fever Virus-Resistant Mosquitoes

Another strategy to prevent transmission of infectious diseases by mosquito vectors does not exterminate a mosquito population, which could have ecological consequences, but rather, employs a gene drive carrying an antipathogen gene to render mosquitoes unable to harbor a human pathogen. An example of this approach is the propagation of a transgene encoding an antibody against dengue fever virus. Human infections with dengue fever virus are prevalent worldwide (390 million infections per year) and manifest as high fever and severe headaches which can progress to lethal dengue hemorrhagic fever and dengue shock syndrome. Four major serotypes of dengue fever virus are transmitted to humans by infected *Aedes aegypti* mosquitoes, whose habitat range is extensive and expanding. A monoclonal antibody that effectively neutralized all four viral serotypes was identified in a screen of monoclonal antibodies derived from vaccinated and naturally infected individuals. DNA sequences encoding the variable regions of the antibody light and heavy chains were obtained from hybridoma cells (see chapter 4), codon-optimized for expression in mosquitoes, and joined by a short linker for production of a single-chain variable fragment (scFv) (see chapter 5). The *A. aegypti* carboxypeptidase promoter was chosen to drive expression of the scFv gene in the midgut of female mosquitoes after a bloodmeal, when they would be exposed to the dengue virus (Fig. 12.50). The transgene construct was introduced into early-stage embryos to create a line of transgenic mosquitoes that had stably integrated the anti-dengue virus antibody gene into a chromosome. In dengue virus challenge experiments, viral RNA was not detectable in the midgut or saliva of female mosquitoes that were homozygous for the transgene, and heterozygous transgenic mosquitoes had reduced levels of midgut and saliva viral RNA compared to wild-type mosquitoes. Similar results were obtained following challenge with all four dengue virus serotypes. This result is proof of the concept that expression of an anti-dengue virus antibody in homozygous transgenic mosquitoes confers resistance to dengue virus infection. It is expected that propagation of the scFv gene on a gene drive will rapidly produce a homozygous *A. aegypti* population which will prevent dengue virus transmission without eliminating the mosquito population, although this remains to

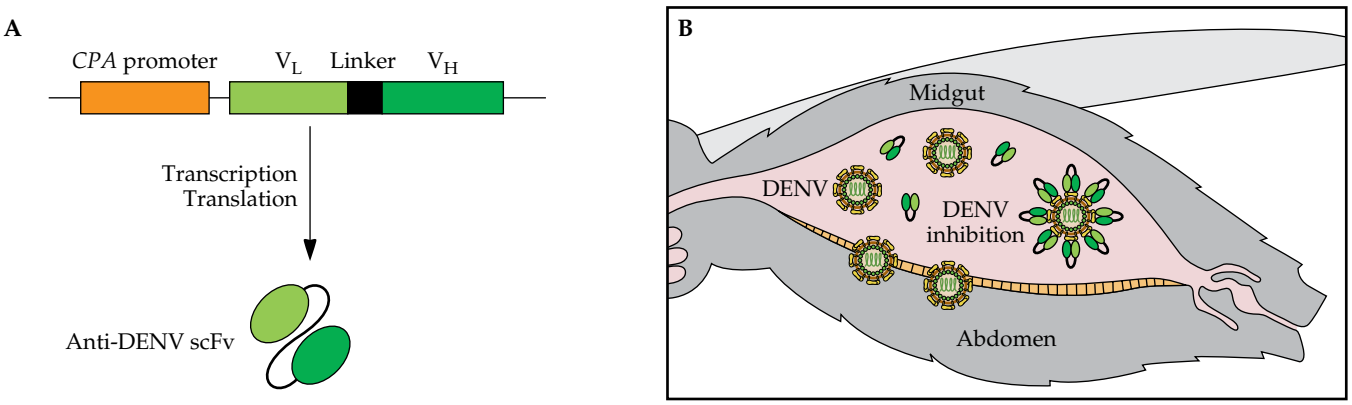


Figure 12.50 Dengue fever virus-resistant transgenic mosquitoes. **(A)** A single-chain variable fragment (scFv) was produced from a gene encoding the variable regions of the light (V_L) and heavy (V_H) chains, joined by a linker peptide, of a monoclonal antibody directed against dengue fever virus (DENV). In transgenic mosquitoes, the scFv is expressed from the carboxypeptidase gene (*CPA*) promoter in the midgut of female mosquitoes after they have fed on blood **(B)**. In the mosquito midgut, the scFv binds to and neutralizes DENV ingested in the bloodmeal.

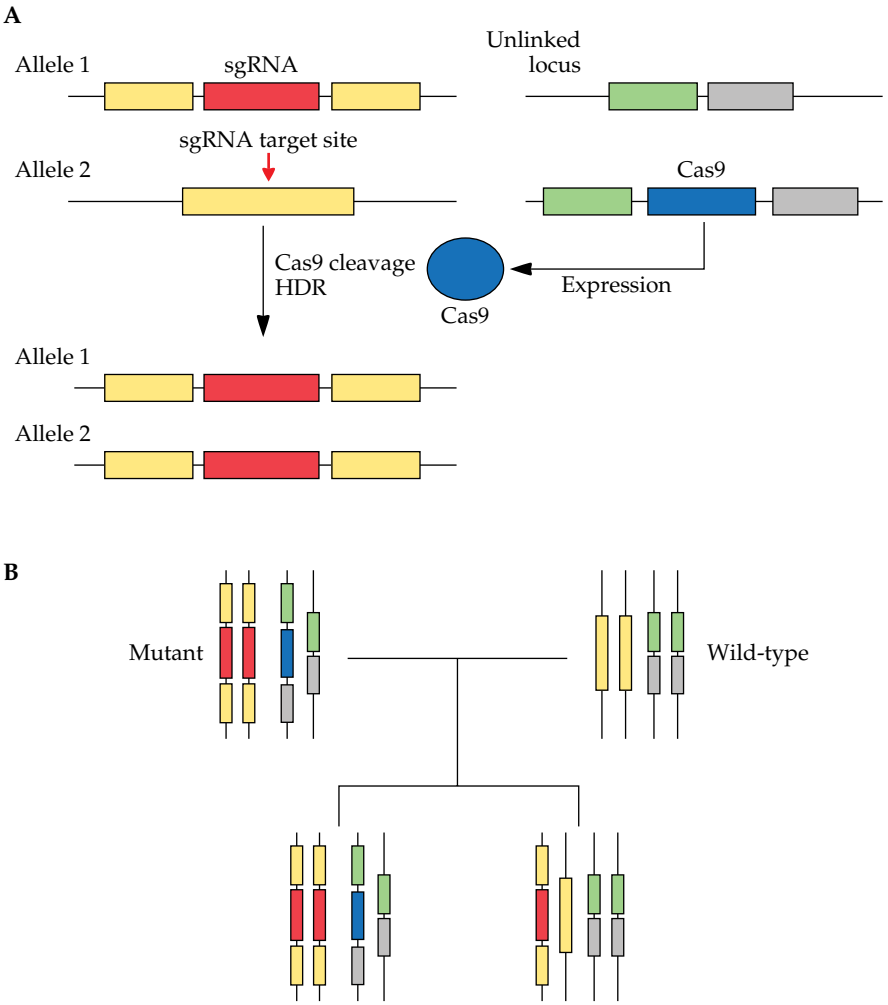
be tested. If effective, a similar strategy could be employed to reduce transmission of other human pathogens borne by *A. aegypti*, such as Zika virus and yellow fever virus.

Reversal Drives

Despite the potential of gene drives to eradicate devastating infectious diseases, there are concerns about unforeseen ecological consequences of elimination of an insect population and the potential for the gene drive to spread to nontarget organisms or populations. These concerns have so far precluded field-testing of the gene drive technology. Various reversal drives have been proposed to prevent propagation of the gene drive indefinitely, such as self-exhausting and self-eliminating drives. In a self-exhausting or split drive, Cas9 and sgRNA are encoded at separate loci in the genome (Fig. 12.51). While the sgRNA gene (in some cases together with a transgene) which is flanked by homology arms can be copied at a target site, the Cas9 gene is not. When both genes are present and expressed, the gene drive is active; however, if they are not inherited together, the gene drive cannot be propagated and will be lost over time, especially if the loss of a component increases fitness. A self-eliminating gene drive is excised from germ cells in a controlled manner, thus terminating its further spread. In one scenario, the gene drive is flanked by recognition sites for a recombinase (Fig. 12.52). The gene encoding the recombinase is included on the gene drive but is not expressed during the gene drive propagation phase. When the field trial is concluded, expression of the recombinase is induced in germ cells, and recombination results in excision of the gene drive. The restored allele will have lost the target sequence for the sgRNA encoded

on the gene drive and is therefore resistant to further modification, creating a pool of resistant alleles in the population. *In silico* modeling of this self-elimination strategy suggests that, even at low rates of efficiency, a population would eventually revert to wild type (Fig. 12.53); however, it has not yet been tested in the laboratory.

Figure 12.51 A split gene drive. **(A)** The gene drive components (sgRNA and Cas9 endonuclease genes) are integrated at separate loci in the insect genome. To disrupt the function of a gene (yellow boxes), the sgRNA gene is inserted into one allele and, when expressed, directs Cas9 expressed from an unlinked locus to cleave within the other allele (red arrow). Homology-direct repair (HDR) results in a copy of the sgRNA gene in allele 2 generating a homozygous mutant. The insect remains heterozygous for the Cas9 gene. **(B)** When the mutant insect mates with a wild-type individual, all offspring will inherit an sgRNA gene; however, they have only a 50% chance of inheriting the Cas9 gene. Offspring without the Cas9 gene will remain heterozygous for the mutation. Over several generations, the gene drive will be lost from the population.



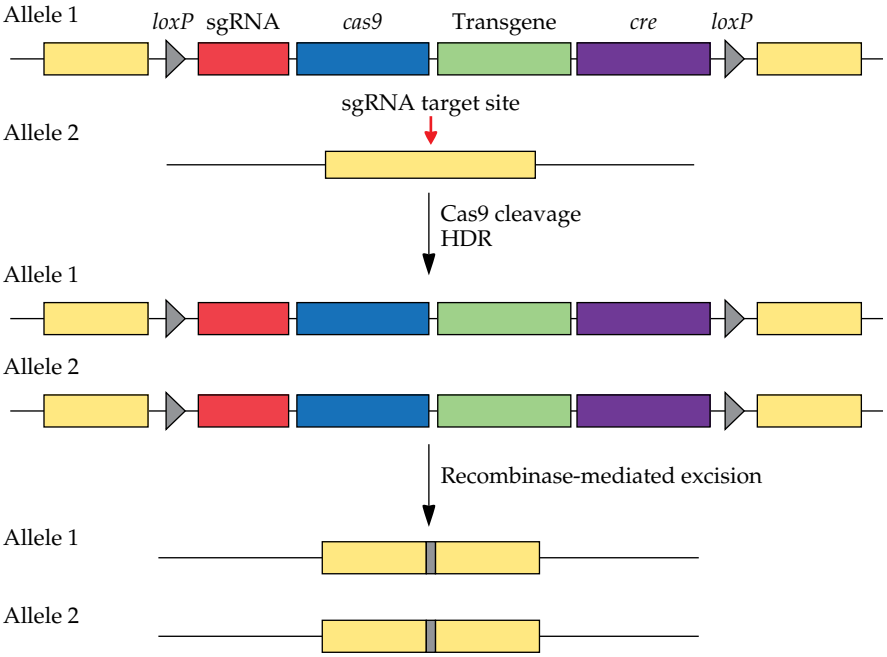
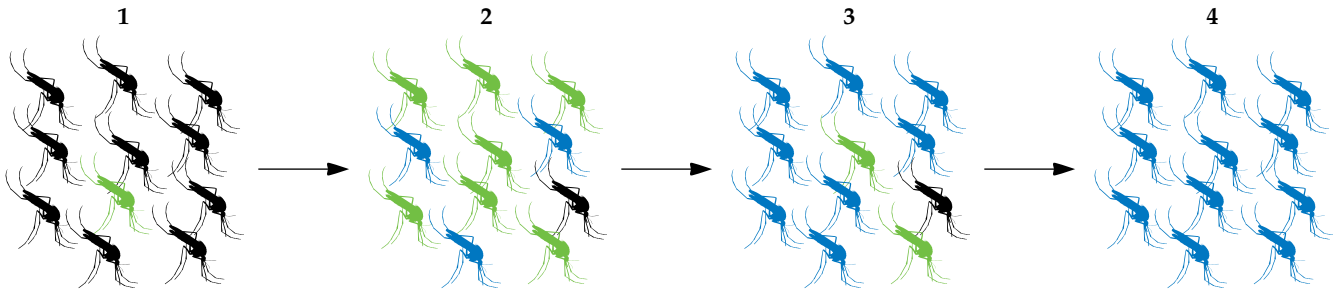


Figure 12.52 A self-eliminating gene drive. The gene drive, including an sgRNA gene, a *cas9* gene, a transgene, and the *cre* gene encoding Cre recombinase, is flanked by *loxP* sequences and inserted into one allele of a gene (yellow boxes) in an insect genome. During a field trial, propagation of the gene drive in a population is mediated by sgRNA and Cas9. At the conclusion of the trial, Cre recombinase expression is induced and excises the gene drive at the *loxP* sites to restore the alleles. A small "scar" sequence (gray boxes) remains at the excision site; however, this may not disrupt the reading frame of the restored gene.

Figure 12.53 Self-elimination of a gene drive from a population. Computer modeling indicates that a self-eliminating gene drive will eventually be lost by a population. **(1)** A small number of mosquitoes (green) carrying a gene drive are introduced into a wild-type population of mosquitoes (black). **(2)** The gene drive spreads in a population; however, in some mosquitoes (blue) the gene drive has undergone self-elimination, and these individuals are resistant to the gene drive. **(3)** The abundance of gene drive-resistant mosquitoes increases, and the gene drive cannot spread further, as few wild-type mosquitoes remain in the population. **(4)** Eventually, the gene drive is eliminated from the population.



summary

Transgenesis of animals involves introducing a foreign gene (transgene) into the genome of a cell that might, after proliferation and embryonic development in a receptive female, be present in the germ lines of some of the progeny. These founder animals are then used to establish true-breeding transgenic lineages. The transgene can be microinjected into the male pronucleus of a fertilized egg, delivered into an early-stage embryo using a retroviral vector, or transfected into embryonic stem cells. Transgenic livestock are often created by somatic cell nuclear transfer. This method entails transferring the nuclei from cells that have been genetically manipulated in culture to enucleated oocytes and obtaining among the progeny some animals that carry the transgene in all their cells. Transgenes that are introduced by microinjection or on retroviral vectors are incorporated into the genome at random sites, which is often undesirable. The CRISPR-Cas system enables insertion of a transgene at a specific site in the animal genome. The sgRNA guides Cas endonuclease to generate a double-stranded break at the target site that may be repaired by homologous recombination with donor DNA carrying the transgene.

Various strategies have been devised for regulating transgene expression. The Cre–*loxP* recombination system can be used to activate transgenes by selectively removing DNA elements that block transcription or to inactivate transgenes by excising part of the coding sequence. The tet-on and tet-off protocols use the tetracycline analogue doxycycline to either turn on or turn off the transcription of a transgene.

For some applications, it is desirable to inactivate (knock out) a specific gene in the genome of an animal. This can be accomplished by insertion of a selectable marker gene into the coding sequence of a gene by homologous recombination or by insertion and/or deletion of nucleotides by error-prone DNA repair (nonhomologous end-joining) following DNA cleavage at a targeted site using the CRISPR-Cas system. RNA interference is often used to decrease (knock down) expression of a target gene. These systems have advanced our understanding of gene activity during development and the consequences of loss of gene function in a particular tissue.

Transgenic mice are commonly employed as models for many human diseases, for example, Alzheimer's disease and Duchenne muscular dystrophy. However, other transgenic animals more accurately replicate the pathology of some human diseases. Lipid metabolism and cardiac physiology are similar in rabbits and humans, and transgenic rabbits are suitable models in which to study human cardiovascular diseases. Nonhuman primates, which have longer lifespans than most other laboratory animals, are better models of slow, progressive neurodegenerative diseases such as Huntington's disease. Zebrafish are transparent at specific stages of development and are therefore useful models for visualization of tumor development and progression.

The mammary glands of transgenic animals are used as bioreactors for the production of protein pharmaceuticals in milk. The human protease inhibitors antithrombin and C1 esterase inhibitor are secreted into the milk of goats and rabbits, respectively, from which they are purified for replacement therapy in patients with genetic deficiencies in the genes encoding these proteins. Eggs are also a repository for pharmaceutical proteins such as lysosomal acid lipase synthesized by transgenic chickens.

Genetic manipulation can lead to improved livestock with increased resistance to infectious diseases such as prions, tuberculosis, mastitis, and influenza. Other production traits that have been enhanced by transgenesis are the nutritional content of milk and meat products, growth rates, and muscle mass.

Laboratory experiments have demonstrated the feasibility of using gene drives to propagate specific mutations or transgenes in an insect population to eradicate insect vector-borne infectious diseases. A gene drive that disrupted a specific site in a mosquito gene for sex determination caused the mosquito population to collapse, which could eliminate the spread of malaria. A gene drive that introduces a transgene encoding a specific antibody fragment into a mosquito population would prevent mosquitoes from being infected with and transmitting dengue fever virus. Concerns about the safety of gene drives have precluded their release into natural populations and have prompted researchers to design reversal drives to control their propagation.

REFERENCES

Baron U, **Schnappinger D**, **Helbl V**, **Gossen M**, **Hillen W**, **Bujard H.** 1999. Generation of conditional mutants in higher eukaryotes by switching between the expression of two genes. *Proc Natl Acad Sci USA* **96:**1013–1018.

Bertolini LR, **Meade H**, **Lazzarotto CR**, **Martins LT**, **Tavares KC**, **Bertolini M**, **Murray JD.** 2016. The transgenic animal platform for biopharmaceutical production. *Transgenic Res* **25:**329–343.

Buchman A, **Gamez S**, **Li M**, **Antoshechkin I**, **Li H-H**, **Wang H-W**, **Chen C-H**, **Klein MJ**, **Duchemin J-B**, **Crowe JE Jr**, **Paradkar PN**, **Akbari OS.** 2020. Broad dengue neutralization in mosquitoes expressing an engineered antibody. *PLoS Pathog* **16:**e1008103.

Chan AWS, **Xu Y**, **Jiang J**, **Rahim T**, **Zhao D**, **Kocerha J**, **Chi T**, **Moran S**, **Engelhardt H**, **Larkin K**, **Neumann A**, **Cheng H**, **Li C**, **Nelson K**, **Banta H**, **Zola SM**, **Villinger F**, **Yang J**, **Testa CM**, **Mao H**, **Zhang X**, **Bachevalier J.** 2014. A two years longitudinal study of a transgenic Huntington's disease monkey. *BMC Neurosci* **15:**36–46.

Cibelli JB, **Stice SL**, **Golueke PJ**, **Kane JJ**, **Jerry J**, **Blackwell C**, **Ponce de León FA**, **Robl JM.** 1998. Cloned transgenic calves produced from nonquiescent fetal fibroblasts. *Science* **280:**1256–1258.

Cooper CA, **Garas Klobas LC**, **Maga EA**, **Murray JD.** 2013. Consuming transgenic goats' milk containing the antimicrobial protein lysozyme helps resolve diarrhea in young pigs. *PLoS One* **8:**e58409.

Cooper DKC, **Ekser B**, **Ramsoondar J**, **Phelps C**, **Ayares D.** 2016. The role of genetically engineered pigs in xenotransplantation research. *J Pathol* **238:**288–299.

Cowan PJ, **Aminian A**, **Barlow H**, **Brown AA**, **Chen CG**, **Fisicaro N**, **Francis DM**, **Goodman DJ**, **Han W**, **Kurek M**, **Nottle MB**, **Pearse MJ**, **Salvaris E**, **Shinkel TA**, **Stainsby GV**, **Stewart AB**, **d'Apice AJ.** 2000. Renal xenografts from triple-transgenic pigs are not hyperacutely rejected but cause coagulopathy in non-immunosuppressed baboons. *Transplantation* **69:**2504–2515.

Crispo M, **Mulet AP**, **Tesson L**, **Barrera N**, **Cuadro F**, **dos Santos-Neto PC**, **Nguyen TH**, **Crénéguy A**, **Brusselle L**, **Anegón I**, **Menchaca A.** 2015. Efficient generation of myostatin knock-out sheep using CRISPR/Cas9 technology and microinjection into zygotes. *PLoS One* **10:**e0136690.

Devlin RH, **Yesaki TY**, **Blagl CA**, **Donaldson EM**, **Swanson P**, **Chan W-K.** 1994. Extraordinary salmon growth. *Nature* **371:**209–210.

Galili U. 2001. The α-Gal epitope (Galα1-3Galβ1-4GlcNAc-R) in xenotransplantation. *Biochimie* **83:**557–563.

Gao Y, **Wu H**, **Wang Y**, **Liu X**, **Chen L**, **Li Q**, **Cui C**, **Liu X**, **Zhang J**, **Zhang Y.** 2017. Single Cas9 nickase induced generation of *NRAMP1* knockin cattle with reduced off-target effects. *Genome Biol* **18:**13–27.

Hadjantonakis AK, **Pirity M**, **Nagy A.** 2008. Cre recombinase mediated alterations of the mouse genome using embryonic stem cells. *Methods Mol Biol* **461:**111–132.

Kaufman CK, **Mosimann C**, **Fan ZP**, **Yang S**, **Thomas AJ**, **Ablain J**, **Tan JL**, **Fogley RD**, van **Rooijen E**, **Hagedorn EJ**, **Ciarlo C**, **White RM**, **Matos DA**, **Puller A-C**, **Santoriello C**, **Liao EC**, **Young RA**, **Zon LI.** 2016. A zebrafish melanoma model reveals emergence of neural crest identity during melanoma initiation. *Science* **351:**aad2197.

Kerr DE, **Plaut K**, **Bramley AJ**, **Williamson CM**, **Lax AJ**, **Moore K**, **Wells KD**, **Wall RJ.** 2001. Lysostaphin expression in mammary glands confers protection against staphylococcal infection in transgenic mice. *Nat Biotechnol* **19:**66–70.

Kubota C, **Yamakuchi H**, **Todoroki J**, **Mizoshita K**, **Tabara N**, **Barber M**, **Yang X.** 2000. Six cloned calves produced from adult fibroblast cells after long-term culture. *Proc Natl Acad Sci USA* **97:**990–995.

Kyrou K, **Hammond AM**, **Galizi R**, **Kranjc N**, **Burt A**, **Beaghton AK**, **Nolan T**, **Crisanti A.** 2018. A CRISPR-Cas9 gene drive targeting *doublesex* causes complete population suppression in caged *Anopheles gambiae* mosquitoes. *Nat Biotechnol* **36:**1062–1066.

Laird FM, Cai H, Savonenko AV, Farah MH, He K, Melnikova T, Wen H, Chiang H-C, Xu G, Koliatsos VE, Borchelt DR, Price DL, Lee H-K, Wong PC. 2005. BACE1, a major determinant of selective vulnerability of the brain to amyloid-cognitive, emotional, and synaptic functions. *Neurobiol Dis* **25:**11693–11709.

Leung C, **Jia Z.** 2016. Mouse genetic models of human brain disorders. *Front Genet* **7:**40.

Long C, **McAnally JR**, **Shelton JM**, **Mireault AA**, **Bassel-Duby R**, **Olson EN.** 2014. Prevention of muscular dystrophy in mice by CRISPR/Cas9-mediated editing of germline DNA. *Science* **345:**1184–1188.

Lowey S, **Bretton V**, **Joel PB**, **Trybus KM**, **Gulick J**, **Robbins J**, **Kalganov A**, **Cornachione AS**, **Rassier DE.** 2018. Hypertrophic cardiomyopathy R403Q mutation in rabbit β-myosin reduces contractile function at the molecular and myofibrillar levels. *Proc Natl Acad Sci USA* **115:**11238–11243.

Lyall J, **Irvine RM**, **Sherman A**, **McKinley TJ**, **Núñez A**, **Purdie A**, **Outtrim L**, **Brown IH**, **Rolleston-Smith G**, **Sang H**, **Tiley L.** 2011. Suppression of avian influenza transmission in genetically modified chickens. *Science* **331:**223–226.

McGowan E, **Sanders S**, **Iwatsubo T**, **Takeuchi A**, **Saido T**, **Zehr C**, **Yu X**, **Uljon S**, **Wang R**, **Mann D**, **Dickson D**, **Duff K.** 1999. Amyloid phenotype characterization of transgenic mice overexpressing both mutant amyloid precursor protein and mutant presenilin 1 transgenes. *Neurobiol Dis* **6:**231–244.

Nelson CE, **Hakim CH**, **Ousterout DG**, **Thakore PI**, **Moreb EA**, **Castellanos Rivera RM**, **Madhavan S**, **Pan X**, **Ran FA**, **Yan WX**, **Asokan A**, **Zhang F**, **Duan D**, **Gersbach CA.** 2016. *In vivo* genome editing improves muscle function in a mouse model of Duchenne muscular dystrophy. *Science* **351:** 403–407.

Niu D, **Wei H-J**, **Lin L**, **George H**, **Wang T**, **Lee I-H**, **Zhao H-Y**, **Wang Y**, **Kan Y**, **Shrock E**, **Lesha E**, **Wang G**, **Luo Y**, **Qing Y**, **Jiao D**, **Zhao H**, **Zhou X**, **Wang S**, **Wei H**, **Güell M**, **Church GM**, **Yang L.** 2017. Inactivation of porcine endogenous retrovirus in pigs using CRISPR-Cas9. *Science* **357:** 1303–1307.

Noble C, **Min J**, **Olejarz J**, **Buchthal J**, **Chavez A**, **Smidler AL**, **DeBenedictis EA**, **Church GM**, **Nowak MA**, **Esvelt KM.** 2019. Daisy-chain gene drives for the alteration of local populations. *Proc Natl Acad Sci USA* **116:**8275–8282.

Pfeifer A. 2006. Lentiviral transgenesis: a versatile tool for basic research and gene therapy. *Curr Gene Ther* **6:**535–542.

Petitte JN, Clark ME, Liu G, Verrinder Gibbins AM, Etches RJ. 1990. Production of somatic and germline chimeras in the chicken by transfer of early blastodermal cells. *Development* **108:**185–189.

Pirottin D, Grobet L, Adamantidis A, Farnir F, Herens C, Schrøder HD, Georges M. 2005. Transgenic engineering of male-specific muscular hypertrophy. *Proc Natl Acad Sci USA* **102:**6413–6418.

Rafferty SA, Quinn TA. 2018. A beginner's guide to understanding and implementing the genetic modification of zebrafish. *Prog Biophys Mol Biol* **138:**3–19.

Richt JA, Kasinathan P, Hamir AN, Castilla J, Sathiyaseelan T, Vargas F, Sathiyaseelan J, Wu H, Matsushita H, Koster J, Kato S, Ishida I, Soto C, Robl JM, Kuroiwa Y. 2007. Production of cattle lacking prion protein. *Nat Biotechnol* **25:**132–138.

Sykes M, Sachs DH. 2019. Transplanting organs from pigs to humans. *Sci Immunol* **4:**eaau6298.

Tait-Burkard C, Doeschl-Wilson A, McGrew MJ, Archibald AL, Sang HM, Houston RD, Whitelaw CB, Watson M. 2018. Livestock 2.0: genome editing for fitter, healthier, and more productive farmed animals. *Genome Biol* **19:**204–2014.

Urlinger S, Baron U, Thellmann M, Hasan MT, Bujard H, Hillen W. 2000. Exploring the sequence space for tetracycline-dependent transcriptional activators: novel mutations yield expanded range and sensitivity. *Proc Natl Acad Sci USA* **97:**7963–7968.

Van Dorpe J, Smeijers L, Dewachter I, Nuyens D, Spittaels K, Van Den Haute C, Mercken M, Moechars D, Laenen I, Kuiperi C, Bruynseels K, Tesseur I, Loos R, Vanderstichele H, Checler F, Sciot R, Van Leuven F. 2000. Prominent cerebral amyloid angiopathy in transgenic mice overexpressing the London mutant of human APP in neurons. *Am J Pathol* **157:**1283–1298.

van Veen HA, Koiter J, Vogelezang CJM, van Wessel N, van Dam T, Velterop I, van Houdt K, Kupers L, Horbach D, Salaheddine M, Nuijens JH, Mannesse MLM. 2012. Characterization of recombinant human C1 inhibitor secreted in milk of transgenic rabbits. *J Biotechnol* **162:**319–326.

Wall RJ, Powell AM, Paape MJ, Kerr DE, Bannerman DD, Pursel VG, Wells KD, Talbot N, Hawk HW. 2005. Genetically enhanced cows resist intramammary *Staphylococcus aureus* infection. *Nat Biotechnol* **23:**445-451.

Wilmut I, Schnieke AE, McWhir J, Kind AJ, Campbell KHS. 1997. Viable offspring derived from fetal and adult mammalian cells. *Nature* **385:**810–813.

Woodfint RM, Hamlin E, Lee K. 2018. Avian bioreactor systems. *Mol Biotechnol* **60:**975–983.

Yamamoto A, Lucas JJ, Hen R. 2000. Reversal of neuropathology and motor dysfunction in a conditional model of Huntington's's disease. *Cell* **101:**57–66.

Yang SH, Cheng PH, Banta H, Piotrowska-Nitsche K, Yang JJ, Cheng EC, Snyder B, Larkin K, Liu J, Orkin J, Fang ZH, Smith Y, Bachevalier J, Zola SM, Li SH, Li XJ, Chan AW. 2008. Towards a transgenic model of Huntington's's disease in a non-human primate. *Nature* **453:**921–924.

Yu S, Luo J, Song Z, Ding F, Dai Y, Li N. 2011. Highly efficient modification of beta-lactoglobulin (BLG) gene via zinc-finger nucleases in cattle. *Cell Res* **21:**1638–1640.

Yum S-Y, Youn K-Y, Choi W-J, Jang G. 2018. Development of genome engineering technologies in cattle: from random to specific. *J Anim Sci Biotechnol* **9:**16–24.

Zapletal J, Najmitabrizi N, Erraguntla M, Lawley MA, Myles KM, Adelman ZN. 2021. Making gene drive biodegradable. *Philos Trans R Soc Lond B Biol Sci* **376:**20190804.

review questions

1. Describe a method to create a transgenic mouse.
2. Describe the major features of a retroviral vector and its application in transgenesis.
3. What is a major advantage of using the embryonic stem cell method compared to DNA microinjection and the retroviral vector method for transgenesis?
4. How is the genomic site of transgene integration determined in a transgenic animal created by the embryonic stem cell method?
5. In the context of transgenesis, what are mosaic animals and how do they arise?
6. What is somatic cell nuclear transfer? What are some of the advantages and disadvantages of this method to produce transgenic animals?
7. Describe how the CRISPR-Cas system is used to insert a transgene into an animal genome.
8. Describe how the CRISPR-Cas system is used to disrupt a gene in an animal genome.
9. Discuss some of the limitations of the CRISPR-Cas system for modifying animal genomes.
10. How is the Cre-*loxP* recombination system used to excise a segment of DNA from a genome? Describe an application for this technology.
11. Describe the tetracycline-inducible system to control transgene expression.
12. What are knockout mice? How and why are they established?
13. What are knockdown mice? How are they generated?
14. Describe an application for a transgenic mouse.
15. Describe an animal model of a human disease. Justify the choice of animal used.
16. Discuss how the mammary gland could be used as a bioreactor for the production of commercial products.

17. Discuss how chicken eggs could be used as a bioreactor for the production of commercial products.

18. What approaches have been developed to produce transgenic chickens?

19. Discuss how transgenesis could be used to improve organ transplantation.

20. Describe a strategy to develop transgenic animals that are protected from infectious disease.

21. How can transgenesis of livestock animals lead to improved milk or meat production?

22. Discuss how transgenesis might improve fish aquaculture.

23. What is a gene drive? How is it applied to engineer an animal population?

24. Discuss an application for a gene drive.

25. Describe a strategy to limit the propagation of a gene drive in a population.

Development of Guidelines for Recombinant DNA Research

Deliberate Release of Genetically Modified Microorganisms

Environmental Concerns

Regulations

Regulation of Genetically Modified Foods

Food Ingredients Produced by Genetically Engineered Microorganisms

Genetically Modified Crops

Genetically Engineered Livestock

Societal Concerns about Genetically Modified Foods

Alteration of Nutritional Content of Food

Potential for Introducing Toxins or Allergens into Food

Potential for Transferring Transgenes from Food to Humans or Intestinal Microorganisms

Controversy about the Labeling of Genetically Modified Foods

Impact of Genetically Engineered Crops on Biodiversity

Who Benefits from the Production of Genetically Modified Foods?

Environmental Benefits of Genetically Modified Crops

How Do Views about Genetically Modified Organisms Impact Trade?

Regulation and Safety of Medical Products of Biotechnology

New Biological Drugs

Genetic and Genomic Testing

Economic Issues

Patenting Biotechnology

Patenting

Patenting in Different Countries

Patenting Nucleic Acid Sequences

Patenting Living Organisms

Patenting and Fundamental Research

SUMMARY

REFERENCES

REVIEW QUESTIONS

13 Molecular Biotechnology and Society

MAJOR TECHNOLOGICAL ADVANCES such as molecular biotechnology are seldom implemented without controversy. While many people recognize the enormous potential of recombinant DNA technology to address important problems facing humanity, such as infectious and genetic diseases, food insecurity, and the impact of climate change, others are apprehensive. Since its inception in 1973, concerns have been raised by various groups. Initially, scientists were concerned that recombinant microorganisms might pose a hazard to laboratory workers and others, and self-imposed a moratorium on certain types of recombinant DNA experiments until the adoption of regulatory guidelines. In the late 1990s, a vociferous and aggressive campaign was launched by critics against planting genetically modified crops and marketing products derived from them. The reasons for this response are complex and include concerns about human health and safety, environmental protection, corporate control of the food industry, world trade monopolies, trustworthiness of public institutions, integrity of regulatory agencies, and potential loss of individual choice. More recently, with the advent of genome editing technologies, there has been much thoughtful discussion about the medical and social implications of heritable genetic alteration of humans. There are no easy answers to the ethical, legal, and economic questions raised by the various applications of molecular biotechnology. However, because the stakes are so high, these issues have been examined extensively, and many governments have developed regulations that reflect the views of acceptable practices of the majority of society.

Molecular Biotechnology: Principles and Applications of Recombinant DNA,
Sixth Edition. Bernard R. Glick and Cheryl L. Patten.

© 2022 American Society for Microbiology.

Potential Biohazards of Recombinant DNA Molecules

milestone

After Cohen et al. (*Proc. Natl. Acad. Sci. USA* **70:**3240–3244, 1973) described the basic strategy for inserting a DNA molecule from any source into a plasmid, the safety and ethical implications of recombinant DNA technology were discussed at length in the scientific community. These informal discussions led to the formation of a temporary committee (Ad Hoc Committee on Recombinant Nucleic Acids) that was composed of leading molecular biologists who were charged with examining the concerns in more detail. In 1974, this group published a brief note, which was dubbed the "Berg letter," simultaneously in three major scientific journals: *Science*, *Nature*, and the *Proceedings of the National Academy of Sciences USA*. In this communication, Berg et al. recommended that researchers defer creating microorganisms with novel genes for drug resistance and toxin production and cloning cancer-causing genes from viruses into bacterial host cells. Moreover, they suggested that a government body, which would represent the interests of the public, should prepare guidelines and regulations for recombinant DNA technology and that an international meeting of scientists should be convened to "discuss appropriate ways to deal with the potential biohazards of recombinant DNA molecules." This letter had a significant impact. It led to a scientific meeting at the Asilomar Conference Center in California that attempted to define various recombinant DNA experiments in terms of their relative risk and suggested which laboratory safeguards should be implemented for experiments that were considered to have a minimal, low, moderate, or high risk. In turn, the Asilomar recommendations (Berg et al., *Science* **188:**991–994, 1975) contributed to the development of the 1976 NIH Guidelines that established stringent regulations for all types of recombinant DNA experiments. The Berg letter was responsible for the formulation of the regulatory framework for recombinant DNA technology. It focused public attention on the potential risks and benefits of recombinant DNA technology and ignited a heated public debate about recombinant DNA technology. Historically, it represents an interesting example of scientists trying to anticipate the possible risks of a new technology before there was any substantial evidence that there were any hazards.

Development of Guidelines for Recombinant DNA Research

In 1974, when it was realized that recombinant DNA technology could be used to engineer microorganisms with foreign genes, concerns about safety, ethics, and unforeseen consequences were raised by scientists, the public, and government officials. The major apprehension was that, either inadvertently or possibly deliberately for the purposes of warfare, unique microorganisms that had never previously existed would be developed and would cause epidemics or environmental catastrophes. In response, a group of prominent molecular biologists called for a moratorium on several kinds of recombinant DNA research, especially experiments that involved pathogenic microorganisms.

Subsequently, in 1976, the U.S. National Institutes of Health (NIH), the primary U.S. research grant agency in the medical and health sciences, issued "Guidelines for Research Involving Recombinant DNA Molecules" (NIH Guidelines). These rules and regulations rigorously defined physical (laboratory) containment levels for the conduct of different types of recombinant DNA experiments. They also required that biological containment be a component of any recombinant DNA experiment, i.e., the preferred hosts for foreign DNA would be those microorganisms considered least likely to proliferate outside the laboratory or to transfer their DNA to other microorganisms. For research with some known pathogenic organisms, controlled, negative-pressure, self-contained rooms were required. Research that was perceived to be less dangerous could be

contained within biological safety cabinets equipped with high-quality filter systems. One specific class of experiments that was "not to be initiated at the present time" under any circumstances was the "deliberate release into the environment of any organism containing a recombinant DNA molecule." Although the NIH Guidelines did not have legal status, most researchers, including those who did not receive NIH funding and private companies that were starting recombinant DNA technology research programs, voluntarily complied. Many other countries, using the NIH Guidelines as a model, adopted their own sets of restrictions for the conduct of recombinant DNA research.

The initial NIH Guidelines were very stringent, and many scientists thought that they were excessive. For example, the costs of the containment facilities required by the guidelines effectively prevented smaller companies and researchers with modest grant support from initiating programs using recombinant DNA technology. The NIH Recombinant DNA Molecule Program Advisory Committee (RAC) was created to oversee developments in recombinant DNA research and to modify the original guidelines, if necessary. The meetings, discussions, and decisions of the committee were, and still are, accessible to the public.

By 1980, the original NIH Guidelines were relaxed considerably by the RAC as a result of experience and specific experimental data from studies that the committee and the NIH had sponsored. For example, it was established that the common host organism *Escherichia coli* K-12 was unable to proliferate to any significant extent outside the laboratory, existing safety procedures for work with pathogenic organisms were already of a high standard, obviating the need for more stringent rules, and the risk of creating a pathogenic recombinant organism was low if the cloned gene had nothing to do with pathogenesis in the original organism. Today, the NIH Guidelines govern research involving transfer of nucleic acids to microorganisms, plants, and animals, although human gene therapy is regulated as a biological drug and falls under the purview of the U.S. Food and Drug Administration (FDA). The NIH Guidelines have recently been amended to include synthetic nucleic acids in light of the ability to synthesize and assemble DNA fragments into entire viral and bacterial genomes (see chapter 2), and the RAC was renamed the Novel and Exceptional Technology and Research Advisory Committee (NExTRAC) in 2019 to reflect an expanded focus on emerging biotechnologies. The risk of a recombinant DNA experiment is assessed according to its potential to cause disease in humans and includes consideration of disease severity and the availability of treatment. Good laboratory practices and containment conditions appropriate for the level of risk, as outlined in the guidelines, ensure the safety of laboratory workers constructing and handling recombinant and synthetic nucleic acids, and the organisms that contain them, and decrease the probability of dissemination of the nucleic acids outside the laboratory. The guidelines and this approach have been effective. No major laboratory-related accidents with or unintended releases of genetically modified organisms have been reported. Although laboratory-acquired infections occur, those involving genetically manipulated microorganisms are extremely rare. Only five have been reported, all involving recombinant vaccinia virus. Most of these were mild because the workers had been vaccinated with the smallpox vaccine, and all were successfully resolved.

Deliberate Release of Genetically Modified Microorganisms

Genetically modified bacteria, fungi, algae, and protozoa have the potential to solve many environmental problems and may be deliberately released into an environment to function as biofertilizers to promote crop growth, biopesticides to protect plants from insects or microbial pathogens, biosensors to detect environmental contaminants, or bioremediants to degrade pollutants (see chapter 8). Although society may benefit from these applications, the risks of unintended adverse consequences to human health and the ecosystem must be as low as possible.

Environmental Concerns

Several properties of genetically modified microorganisms are considered when evaluating the potential impact of their release into an environment. Living microorganisms can replicate and conditions in the environment may be suitable for their proliferation, the extent to which may be difficult to predict or control. If the introduced microbe is able to obtain and utilize the available resources efficiently, then it may outcompete indigenous microorganisms, thereby disrupting the structure and function of the ecosystem. This is evaluated experimentally before the microbe is released. In fact, most genetically engineered microorganisms are less fit than their wild-type counterparts and are unable to compete effectively with native microorganisms. This is due in part to repeated subculture in the laboratory under optimal conditions to which they become adapted. In addition, replication and expression of introduced DNA may pose a metabolic burden. However, in a few cases the introduced gene(s) may increase the competitiveness of the engineered strain under environmental conditions. An example is a bacterium that expresses an enzyme for utilization of an environmental pollutant as a carbon source.

Bacterial genes, whether native or introduced by genetic engineering, can be transferred among bacteria. The transfer of DNA from one bacterium to another in the natural environment may occur by conjugation in which a plasmid is transferred, transduction of bacteriophage DNA, transposition of a mobile genetic element such as a transposon, or transformation with free DNA. Frequencies of gene transfer may be higher in some environments, such as in biofilms and sites that host dense populations of bacteria. The stability of foreign DNA acquired by a new host may be limited by plasmid incompatibility and host range restrictions, digestion by host restriction endonucleases, and the metabolic load it imposes on the host.

For DNA transfer to pose a risk, the gene(s) newly acquired by an unintended host bacterium must enhance its competitiveness and/or pathogenicity. It is highly unlikely that a single transferred gene would confer these complex traits. Moreover, the transferred gene(s) must be expressed, which requires the host proteins to recognize and efficiently utilize the gene's transcription and translation signals. However, there have been cases where introduction of a single gene has led to unexpected virulence. As an example, introduction of a gene encoding a phytotoxin created a hypervirulent variant of a normally weakly phytopathogenic fungus, although this strain was not released to the environment. Also of concern is the spread of antibiotic resistance genes, which are commonly present on plasmids used to deliver a transgene. A solution to this concern is to stably integrate the foreign gene(s) into the host chromosome and avoid insertion of antibiotic resistance genes.

Regulations

In many countries, the laws related to the release of microorganisms, especially genetically engineered organisms, are amendments to existing laws that regulate the release of chemicals. These laws were established to ensure that human health and the environment are protected, while recognizing that microorganisms can be used to improve these and can also provide significant economic benefits. The regulations generally take a preventative approach, requiring that the proposed microorganism be assessed for safety prior to release, usually on a case-by-case basis. Included in the regulatory assessments are evaluations of the pathogenicity to humans and other nontarget organisms, toxin production, survival in the environment, the potential for gene transfer, and in cases where the microbe produces a pesticide, the potential for resistance to develop in the target organism.

In the United States, the Environmental Protection Agency (EPA) is responsible for regulating the release of new substances, including microorganisms, into the environment. Two acts set the regulatory framework. The Toxic Substances Control Act (TSCA) was established in 1976 to protect the environment and human health from the possible toxic effects of manufactured chemical substances. In 1997, genetically engineered microorganisms were deemed regulatable under the TSCA because they possess new traits and therefore their effects on health and the environment are uncertain. The TSCA also pertains to microorganisms that are manufactured or imported for industrial applications. Microorganisms that have been genetically modified to produce pesticides are regulated under the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA).

A new substance is defined as one that is not on the TSCA Chemical Substance Inventory, which lists substances deemed to be reasonably safe for humans and the environment. The definition includes genetically engineered microorganisms that result from the deliberate introduction of intergeneric genetic material, that is, DNA that was originally isolated from an organism belonging to a different genus. Microorganisms that contain intrageneric DNA are not included. This approach is based on the premise that intergeneric microorganisms are more likely to exhibit new traits and their behavior is therefore less predictable. Naturally occurring bacteria are not considered to be new and therefore are not subject to regulation. Exemptions are also made for host microorganisms that are on the approved list, for example, *Bradyrhizobium japonicum* and *Sinorhizobium meliloti*, and are transformed with DNA that is limited in size, well-characterized (including coding and regulatory sequences), poorly mobilizable, and free of certain sequences (e.g., toxin sequences).

Applications for approval and subsequent registration in the Chemical Substance Inventory require data obtained from field trials to support the safety and efficacy of the microorganism. These small-scale experimental releases are also authorized by the EPA. For example, field tests were recently approved to evaluate the ability of two genetically engineered bacteria to degrade hexahydro-1,3,5-trinitro-1,3,5-triazine, a chemical widely used in explosives. The strains will be assessed for their ability to proliferate in explosive-contaminated sludge and liquid wastes from a chemical depot, as well as to transfer the transgenes to indigenous microorganisms.

Under the FIFRA, the EPA has approved several genetically engineered microbial pesticides. These include insecticides such as *Bacillus thuringiensis* that expresses insecticidal crystal (Cry) proteins from a different strain, and bacteriocides such as *Agrobacterium radiobacter* that prevents crown

gall disease and contains a deletion to prevent DNA transfer. In evaluating proposals for commercial use, the risks to humans and the environment from exposure to a microbial pesticide are balanced with the benefits from its application. The infectivity, pathogenicity, and toxicity of the microbial pesticide to mammals, birds, fish, and representative insects, including those that are related to a target insect are considered, especially because the pesticides are often applied to food and feed crops. Toxicity is tested at levels of the pesticide up to 100 times greater than those expected in actual field applications. Persistence in the environment is also examined.

Regulation of Genetically Modified Foods

In the United States, the FDA is responsible for regulating the introduction of foods into the marketplace. The safety of new foods and food additives from microorganisms, crops, and livestock must be thoroughly assessed before they can be licensed for human consumption. The approach in the United States has been that food products produced through recombinant DNA technology are not inherently riskier than those derived through traditional forms of genetic improvement such as selective breeding or cross-pollination. In other words, new food products arising from genetic improvement should be evaluated for risks, regardless of the method by which the product was generated. For the most part, agencies in various countries that are responsible for regulating foods derived from genetically modified organisms have adopted a case-by-case approach. Each submission is considered separately, and depending on the judgment of the regulatory body, a series of tests is specified to ensure that the product is as safe to eat as those derived from their nonengineered counterparts.

Food Ingredients Produced by Genetically Engineered Microorganisms

The first recombinant protein approved by the FDA for use in food was chymosin, an enzyme used in cheese making. Chymosin cleaves the κ-casein protein of milk, creating curds, which in turn are processed into cheese. Traditionally, the milk-clotting agent for cheese making was derived from the fourth stomach of calves and consists of a mixture of substances, including chymosin, which collectively is called rennet. To ensure a reliable, convenient, and cost-effective industrial supply of chymosin, one of the bovine chymosin genes was cloned and expressed, and the product was harvested from *E. coli* K-12. The FDA required the petitioner to show that the recombinant chymosin was identical to the naturally occurring chymosin. The cloned and native DNA sequences of the chymosin gene were determined to be identical, the recombinant chymosin had the same molecular weight as purified calf chymosin, and the biological activities of the two forms of the enzyme were the same. Next, it was essential to establish that the recombinant chymosin preparation was safe. As part of its purification process, recombinant chymosin is extracted from inclusion bodies, and the company showed that the final preparation is free of whole bacterial cells, significant cell debris, and other impurities, including nucleic acids. Although the presence of minute amounts of *E. coli* K-12 cells in the final preparation of chymosin is undesirable, numerous studies have established that this strain is nontoxigenic and nonpathogenic to humans. To ensure that the recombinant chymosin preparations did not

contain an unexpected toxin, animal testing was conducted and showed no adverse effects. After considering all the information, in 1990 the FDA concluded that the recombinant chymosin could be licensed for commercial use. Currently, about 85% of all cheeses in the United States are produced with recombinant chymosin produced in *E. coli*, the filamentous fungus *Aspergillus niger*, and the yeast *Kluyveromyces lactis*. Chymosin preparations from these genetically engineered microorganisms have received the FDA designation Generally Recognized as Safe (GRAS) for use as a food ingredient.

Not all new food ingredients produced by recombinant DNA technology have been as well accepted as chymosin. The bovine somatotrophin (growth hormone) controversy illustrates the myriad of issues that can arise from introducing a recombinant product. Injection of bovine somatotropin into a dairy cow significantly increases its milk yield. Because natural bovine somatotropin is both difficult and costly to acquire in large quantities from the pituitary glands of cattle, it has not been used routinely as an augmenting agent by the dairy industry. Production in a microorganism offers a feasible alternative, and the gene for bovine somatotropin was cloned into *E. coli*, and recombinant bovine somatotropin was harvested, purified, and tested. Under trial conditions, milk production in dairy cows was increased by 20% to 25% after injection of recombinant bovine somatotropin.

The safety of natural bovine somatotropin in milk has been studied exhaustively. In treated cows, the levels of bovine somatotropin in milk are not higher than those in control cows. Moreover, bovine somatotropin is not active in humans, and all toxicity trials have shown that there are no adverse effects on test organisms. The FDA, after considering all the research results that it could assemble, concluded that there is no significant difference in the milk from untreated and recombinant bovine somatrophin-treated cows, that the meat and milk of treated cows are safe for human consumption, and that the recombinant hormone is safe for treated dairy cows. On 5 November 1993, the application to use recombinant bovine somatotropin as a milk production enhancer was granted, and it was marketed under the trade name Posilac.

An effective and powerful lobby group had been assembled to block governmental approval of recombinant bovine somatotropin. The principal reason for this opposition was based on the presumed economic consequences that recombinant bovine somatotropin would have on the dairy industry. The fear was that many small dairy farms would become unprofitable because fewer cows would be required to maintain current levels of milk production. In addition to the attrition of small dairy farmers, opponents of recombinant bovine somatotropin thought that the industry would become dominated by large corporate interests at the expense of independent producers. At the time, these economic arguments may have been justified, and certainly any group is entitled to protect itself against what it perceives to be a threat to its livelihood. However, the main issue in the advertising campaign waged against the use of recombinant bovine somatotropin was that "genetically engineered hormones" in "hormone-laced milk" would be harmful and cause cancer in humans.

Opponents of recombinant bovine somatotropin also contended that its use would increase the incidence of bacterial infection of milk glands (mastitis) in dairy cattle. It was further argued that larger-than-normal amounts of antibiotics would have to be used to maintain the health of recombinant

bovine somatotropin-treated animals, thereby resulting in increased levels of antibiotics in the cows' milk that may trigger allergic responses in some consumers. Increased antibiotic use could also heighten the selection pressure for drug-resistant pathogens. The Veterinary Medicine Advisory Committee of the FDA and others have studied this issue and concluded that the frequency of mastitis is not different between treated and untreated cows. On this basis, there is no reason to believe that the amounts of antibiotics in milk from recombinant bovine somatotropin-treated cows would be any greater than those from untreated cows. Additionally, by law, after any cow is treated with antibiotics, it is not milked for a specified period of time, to enable the medication to be cleared from its system.

The concern about an increased risk of cancer was based on the presumed increased concentration of insulin-like growth factor I in the milk of treated cows. The FDA maintains that the amount of insulin-like growth factor I in recombinant bovine somatotropin-stimulated milk is within the normal range observed for untreated milk. Any additional insulin-like growth factor I would add only a small amount to the existing pool of insulin-like growth factor I in human plasma. Taking into consideration these and other matters, the FDA reviewed the complete recombinant bovine somatotropin file in 1999 and found no reason to rescind its original decision.

Notwithstanding the certainty of both the FDA and the Joint Food and Agricultural/World Health Organization Expert Committee on Food Additives that recombinant bovine somatotropin poses no hazards for treated animals, the Canadian equivalent of the FDA, Health Canada, and the European Union both refused to approve recombinant bovine somatotropin on the grounds that it is detrimental to treated animals. In these instances, appointed committees concluded that recombinant bovine somatotropin supplementation increases the risk of mastitis, causes leg and foot disorders, decreases reproductive capabilities, and induces severe reactions at the site of injection. FDA analyses and more recent studies do not support these assertions. Clearly, the biological consequences of recombinant bovine somatotropin remain controversial. Currently, in the United States, about 15% of dairy producers use Posilac. There have been no reports since the onset of its commercial use that recombinant bovine somatotropin has had debilitating effects on treated cows. Studies also indicate that recombinant bovine somatotropin treatment has neither had an apparent impact on consumer prices nor led to extensive consolidation of small dairy farms. Moreover, a new market niche has been created. In many food stores, the consumer is given the choice between milk and milk products from either recombinant bovine somatotropin- or non-recombinant bovine somatotropin-treated cows.

By 1996, the initial apprehensions about regulating foodstuffs that were developed with recombinant DNA technologies appeared to have subsided in the United States as the FDA implemented traditional risk assessment procedures. Many food ingredients, including vitamins, amino acids, proteins, and oligosaccharides, and food enzymes such as lactase, phospholipase, amylases, and proteases produced by genetically engineered microorganisms have since been approved. While the food industry prefers a minimal set of regulations to speed up the transition from the developmental phase to the marketplace, governmental agencies, as representatives of the public, have a dual responsibility. They are entrusted

with protecting public health and ensuring that new developments are not needlessly curtailed. Presumably, any changes to the regulations will not be made for the sake of expediency at the expense of safety.

Genetically Modified Crops

While there is still polarization on the issue of genetically modified crops, many national governments have approved the commercialization of several transgenic plants (Table 13.1). The United States cultivated more than 70 million hectares, which represents 38% of the total global area planted with transgenic crops, in 2019. In total, 29 countries now grow transgenic crops on 190 million hectares. Four crops, soybean, corn, cotton, and canola, represent about 99% of the genetically engineered crops, with alfalfa, sugar beets, sugarcane, papaya, safflower, potatoes, eggplant, squash, apples, and pineapple making up the balance. Most have been engineered to increase productivity by incorporating genes that confer tolerance to herbicides and resistance to insect predation, although crops engineered to mitigate abiotic stresses, such as drought-tolerant corn, have increased and are expected to become more widespread as the impact of climate change increases. Other traits that have been approved but are marketed to a lesser extent include resistance to viral infection (especially in papaya) and altered plant quality (e.g., flower color in carnations).

While recognizing the need to increase agricultural production, the overarching purpose of regulating the commercialization of transgenic crops, as for any crop, is to ensure that the crops are safe to eat by humans and livestock and safe to grow. Not only must the product be nontoxic for consumption, but there must also be consideration of its potential impact on other organisms and on environments outside the area of cultivation because the plants are usually grown in an open field environment. For example, there is potential for plants carrying insecticidal toxins to adversely affect nontarget insects or to cross-breed with wild plants that could enhance their invasiveness. In many countries, regulations governing transgenic crops are still being developed and are evolving.

Table 13.1 Land area under cultivation with transgenic crops in countries that cultivated more than 1 million hectares in 2019[a]

Country	Area (million hectares)	Transgenic crop(s)
United States	71.5	Corn, soybeans, cotton, alfalfa, canola, sugar beets, potatoes, papaya, squash, apples
Brazil	52.8	Soybeans, corn, cotton, sugarcane
Argentina	24.0	Soybeans, corn, cotton, alfalfa
Canada	12.5	Canola, soybeans, corn, sugar beets, alfalfa, potatoes
India	11.9	Cotton
Paraguay	4.1	Soybeans, corn, cotton
China	3.2	Cotton, papaya
South Africa	2.7	Corn, soybeans, cotton
Pakistan	2.5	Cotton
Bolivia	1.4	Soybeans
Uruguay	1.2	Soybeans, corn

[a]Data from James, *Global Status of Commercialized Biotech/GM Crops: 2019*. International Service for the Acquisition of Agri-Biotech Applications (ISAAA) Brief no. 55. ISAAA, Ithaca, NY, 2019.

In the United States, three agencies are responsible for assessing and approving applications for the development and release of genetically modified crops. The U.S. Department of Agriculture (USDA) is responsible for protecting agriculture and the environment from pests, the FDA is responsible for the safety of human food and animal feed, and the EPA is responsible for regulating pesticides, including plants that are engineered to produce pesticides. In establishing regulations, the general view among the national regulatory agencies has been to consider the characteristics of a transgenic rather than the process by which it was created. Plants derived by transgenesis are not intrinsically less safe than those from traditional breeding experiments. In practice, however, genetically engineered plants tend to be scrutinized more closely than conventionally bred plants. Currently, plants developed through genome editing technologies such as the clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated endonuclease (Cas) system, transcription activator-like effector nucleases (TALENs), and zinc-finger nucleases are not subject to regulation by the USDA if they do not contain any foreign DNA (genes outside the plant's gene pool) and therefore are not considered to be genetically engineered. In most of these plants the function of a gene has been abolished by a small number of nucleotide changes.

Many other countries have taken a similar approach when adopting legislation and follow general guidelines provided by the Organization for Economic Cooperation and Development and the Cartagena Protocol on Biosafety, which was established to develop international standards for biosafety. The regulatory system in the European Union differs somewhat from that in the United States in placing greater emphasis on the process by which the crops have been developed. For example, in the European Union all genome-edited organisms are considered to be genetically engineered and therefore are subject to regulation. There is also a requirement for additional information that is used for labeling and to enable tracking of the origin of the plant. Draft decisions on new genetically modified food applications are made by the European Commission and then by majority rule of the member states. However, there are strong and opposing views on genetically engineered foods among European countries, which has thwarted approval of transgenic crops. To address the gridlock in the approval process, a directive was introduced to allow individual European countries to reject the cultivation of genetically modified crops after an approval has been adopted by the European Commission. As of 2021, only corn carrying an insecticide transgene is cultivated in the European Union, mainly in Spain and to a lesser extent in Portugal, although millions of tons of transgenic seeds are imported legally for animal feed.

In general, each new transgenic plant is considered on a case-by-case basis. Even when a previously approved transgene is introduced into a different plant variety or into the same plant genotype, each transgenesis event is assessed in a new application because the site of insertion into the plant genome is usually random depending on the method used, and therefore the potential for disruption of the function or regulation of endogenous or introduced genes must be considered. Other traits that are assessed include the rate and method of reproduction, the potential for transfer of genetic material to other organisms, toxicity to other organisms, and sexual compatibility with wild relatives. For plants engineered with insecticidal proteins, strategies to manage insect resistance must also be in place (Table 13.2).

Table 13.2 Some of the studies submitted to the U.S. EPA as part of the approval process for a variety of Bt corn[a,b]

Molecular characterization of insect-protected transgenic corn
Evaluation of transgenic corn line in U.S. field trials
Assessment of the equivalence of a *B. thuringiensis* subsp. *kurstaki* protein produced in commercial Bt corn varieties and microorganisms
Dietary toxicity study with Bt corn meal in the northern bobwhite
Aerobic soil degradation study of Bt protein
Acute oral toxicity study of Bt corn tryptic protein in albino mice
Assessment of the *in vitro* digestive fate of Bt protein
Stability of Bt protein in sucrose and honey solutions under nonrefrigerated temperature conditions
Evaluation of the dietary effects of purified Bt proteins on honeybee larvae
Evaluation of the dietary effects of purified Bt proteins on honeybee adults
Dietary toxicity study of activated Bt protein with green lacewing larvae
Dietary toxicity study of activated Bt protein with the parasitic hymenopteran *Brachymeria intermedia*
Dietary toxicity study of activated Bt protein with ladybird beetles
Evaluation of Bt corn feed as ingredient for catfish feed
Acute toxicity study of Bt protein with the earthworm in an artificial soil substrate
Effects of Bt protein on *Folsomia candida* and *Xenylla grisea* (Insecta: Collembola)
Expression of Bt protein in Bt corn
Tissue expression and corn earworm efficacy of Bt protein in Bt corn
Chronic exposure of *F. candida* to corn tissue expressing Bt protein
Corn pollen containing Bt protein: 48-h static-renewal test with *Daphnia magna* (Cladocera)

[a]Data from National Research Council, *Genetically Modified Pest-Protected Plants: Science and Regulation*, p. 245. National Academies Press, Washington, DC, 2000, with permission.

[b]Bt corn, corn genetically engineered with the *Bacillus thuringiensis cry* gene to make it resistant to pests.

To date, in the United States, more than 200 transgenic crops have been approved for commercialization, cultivation, and importation for food and feed, although not all are currently produced. Of the genetically modified crops planted, about 90% are soybeans and corn, mainly resistant to herbicides and pesticides. Some argue that the lengthy and costly pre- and post-commercialization regulatory processes are prohibitive to the development of new transgenic crops with traits other than herbicide and pest resistance, which would be produced on a smaller scale. They also argue that the high cost of meeting regulatory requirements, estimated to be about $35 million per product, precludes development by anyone other than large multinational companies. Proponents maintain that after more than 20 years of scrutiny and incorporation of many of the safeguards to ensure that transgenic crops are safe for consumption and release into the environment, the regulatory process could be streamlined. To this end, in 2020 the USDA deregulated genetically engineered crops that contain minor genetic changes that could be made through traditional breeding or that are new varieties of already approved transgenic crops.

Somewhat more challenging is the regulation of transgenic plants that produce pharmaceutical proteins. Frequently, these are agricultural crop plants such as corn, tobacco, potato, rice, and safflower, although so far, they have mainly been developed for production in contained facilities, for example, in cell cultures. Pharmaceutical plants grown in greenhouses, cell culture systems, and other contained facilities are regulated as drugs. The plants are genetically manipulated, often with several different genes, to

maximize yields of proteins that are intended to be biologically active in humans or other animals. Thus, the plants potentially pose a greater risk to human health and the environment when grown in the field and therefore are given special consideration by regulators. For example, there are concerns that material derived from the plants could inadvertently end up in the food chain, perhaps through seed dispersal. Prevention of such problems may require special confinement measures such as dedicated farm machinery or containment netting to separate pharmaceutical crops from food crops. Employing nonfood crops such as tobacco to produce pharmaceutical proteins would circumvent some of these problems and also provide tobacco farmers, who are facing income losses due to a decline in tobacco consumption, with an alternate market.

Genetically Engineered Livestock

The regulation of genetically engineered livestock is not fundamentally different from that of genetically modified crops. Generally, the same agencies in the United States are responsible for regulatory oversight of both. Similar to cultivated plants, the goal for the commercial production of animals, regardless of the method of production, is to ensure that the animal products are safe for use and that the impact of the animals on the environment is as low as possible. In addition, the well-being of the animals must be protected. For new, rapidly developing and expanding animal biotechnologies, regulators are challenged with ensuring that these goals are met while recognizing the power of the technologies to solve important problems in agriculture and medicine (see chapter 12). Moreover, for acceptance of the products of these animals and for commercial viability, society must be confident that safety issues have been adequately addressed.

The regulatory issues for cloned livestock that have not undergone modification of their genome are separated from those for transgenic animals. In the United States, the Center for Veterinary Medicine of the FDA considers that the process of cloning livestock through somatic cell nuclear transfer is not different from other assisted reproductive technologies that have been practiced for many years and does not pose greater risks. Many cloned animals, particularly cattle and sheep, are unhealthy; for example, the mortality of born bovine clones in the first six months is 14 to 42%. This is most likely because the genome transferred from a differentiated somatic cell fails to undergo epigenetic reprogramming. Epigenetic reprogramming returns the genomes of differentiated egg and sperm cells to a nascent state in an early embryo that is required for normal development. However, unhealthy animals are generally destroyed, and their tissues are not consumed as food by humans or other animals. Healthy cloned animals are used primarily as breeding stock. Their offspring, which are consumed as food, are produced through sexual reproduction and therefore develop normally (i.e., undergo proper epigenetic programming) and are generally born healthy. The FDA takes the position that healthy animals produce safe foods. After extensive assessment of meat and milk composition, it was found that meat and milk derived from cloned animals are not different and are as safe to eat as those from animals produced using conventional agricultural practices. Additional regulations specific for cloned animals were determined to be unnecessary because any problems associated with the products from cloned animals can be identified through food

inspections in current practice. This view is in sharp contrast to that in the European Union, where the European Parliament voted to ban the cloning of farm animals and the sale of products from cloned animals and their offspring over concerns about the welfare of the cloned animals.

All genetically modified animals must be approved before they are commercialized. In the United States, genetically engineered animals, whether containing heritable or nonheritable recombinant DNA, and regardless of their intended use, are considered to be "new animal drugs" by the FDA. Approval must be obtained each time a gene is introduced into an animal. As for transgenic plants, it is generally recognized that the site of insertion into the animal's genome is difficult to control and may impact the level of expression of the transgene or the animal's native genes and could therefore affect the health of the animal. All animals containing a recombinant DNA construct are regulated, including subsequent generations of animals that contain the construct that were derived from the original manipulated parent animals by breeding with a nontransgenic animal.

A distinction is made between transgenic animals that are to be used for food production and nonfood transgenic animals. Transgenic animals that are used for nonfood research purposes (e.g., laboratory mice) or to produce human therapeutic proteins are contained within controlled environments and may not be as strictly regulated as those that are grown in open environments for food production, unless safety issues are raised. Laboratory animals and animals raised in contained facilities for production of pharmaceuticals pose a low risk for unintended release into the environment. Therapeutic proteins produced in the mammary glands or eggs of three different transgenic animals have been approved by the FDA; human antithrombin protein produced in a goat and human C1 esterase inhibitor protein produced in a rabbit were approved in 2009 and 2014, respectively, and human lysosomal acid lipase produced in chicken eggs was approved in 2015. Genetically engineered animals that have been developed for nonfood uses and may be released into uncontrolled environments are evaluated by the FDA, USDA, or EPA, depending on their intended use. To date, only small-scale releases of transgenic insects have been conducted. For example, in 2021, male mosquitoes carrying a transgene that kills female progeny were released in the Florida Keys to test their potential to specifically suppress the population of *Aedes aegypti* mosquitoes. *A. aegypti*, an invasive species, makes up about 4% of the mosquito population in this region and transmits Zika, dengue, chikungunya, and yellow fever viruses to humans. Data from this field trial will be used by the EPA to determine whether the mosquitoes can be released in other regions of the United States.

To gain approval, a genetically engineered food animal must be demonstrated to be safe for production and consumption, and the introduced characteristic must be shown to be effective as intended. The risk assessment by the FDA and the regulatory agencies of many other countries is consistent with those of the international standards adopted by the Codex Alimentarius Commission of the Food and Agriculture Organization and World Health Organization. The FDA considers several characteristics, including the method for introduction and detection and the nature of the introduced DNA (e.g., expression, stability, and potential for the construct to be acquired by pathogens); the traits conferred by the introduction of the transgene(s) and any alterations from the normal

phenotype of the corresponding nontransgenic animal, especially where it may impact the composition of the animal product (e.g., production of allergens, toxins, and novel metabolites; potential for unintended cell dysregulation); and methods for production, processing, and disposal of transgenic animals and their tissues. Regulations also take into consideration the risk associated with the release of a genetically modified animal into the environment, whether intentional or accidental. Measures to prevent escape of farm animals, and their products, from barns and pastures are required. Transgenic animals must be prevented from mating with nontransgenic relatives, which could spread the transgene, disrupt an ecosystem, or reduce biological diversity. Aquaculture facilities, for example, that raise transgenic fish, must ensure that the fish cannot escape and poachers cannot get in. To date, two transgenic food animals have received approval for commercial sale in the United States. Following 20 years of regulatory scrutiny, in 2015 the FDA approved an Atlantic salmon engineered with a Chinook salmon growth hormone gene for human consumption. The transgenic salmon are all females that are unable to reproduce and are grown in contained, land-based facilities in Panama from eggs produced in Canada. In 2020, a line of pigs engineered to abolish production of the allergen α-Gal on the surface of their cells was approved for human consumption.

Societal Concerns about Genetically Modified Foods

All foods produced today are derived from plants and animals that have been genetically manipulated to enhance desirable characteristics through either traditional selective breeding or recombinant DNA technology. The genetic changes introduced through recombinant DNA technology are vastly smaller and much better understood than those from traditional breeding, as only one or a few well-characterized genes are introduced. Also, because known regulatory elements are included on the introduced genetic constructs, spatial and temporal expression of recombinant genes can be controlled. Despite these advantages, some consumers are concerned that manipulating the genomes of food plants and animals through recombinant DNA technology can lead to the production of foods that are unsafe for human consumption. This fear has had a powerful effect on the food industry. For example, consumer concerns over the safety of transgenic foods led major baby food producers to stop using them in their products, and several fast-food chains have removed them from their menus. On the other hand, proponents of genetic engineering argue that by rejecting transgenic ingredients, producers are marketing foods that are less safe. For example, conventional food is more likely to be contaminated with potent mycotoxins produced by molds, which could be prevented by growing crops that are engineered to be mold-resistant. In addition, allergens, naturally present in many foods, could be removed through genetic manipulation.

Alteration of Nutritional Content of Food

All genetically engineered foods derived from plants and animals are tested on a case-by-case basis for "substantial equivalence" before approval for commercialization by regulatory agencies. This means that the safety and nutritional content of genetically engineered food must be substantially the same as that of the corresponding conventional food. Nutrients

Table 13.3 Comparison of nutritional content of glyphosate-tolerant transgenic corn grain (Roundup Ready corn line GA21) with that of the nontransgenic parental control line[a]

Nutrient	Transgenic corn	Nontransgenic corn	Range (avg.) for conventional corn
Proximates (% dry weight)			
Protein	11.05	10.54	6.67–14.69 (10.18)
Fat	3.90	3.98	2.03–4.90 (3.48)
Carbohydrates	83.7	83.8	77.4–89.5 (84.8)
Fatty acids (% total fatty acids)			
Palmitic (16:0)	10.70	10.72	8.57–17.46 (11.87)
Stearic (18:0)	1.68	1.67	1.02–2.86 (1.94)
Oleic (18:1)	24.2	24.1	17.4–38.5 (25.6)
Linoleic (18:2)	61.4	61.5	47.7–64.2 (57.5)
Minerals (% dry weight)			
Phosphorus	0.326	0.326	0.160–0.533 (0.321)
Calcium	0.0039	0.0043	0.0022–0.0163 (0.0043)
Amino acids (% total amino acids)			
Methionine	2.16	2.17	(2.05)
Cysteine	2.22	2.28	(2.12)
Lysine	3.11	3.02	(3.09)
Leucine	12.98	12.87	(12.91)
Tryptophan	0.61	0.61	(0.62)
Phenylalanine	5.31	5.33	(5.10)

[a]Data from Chassy et al., *Comp Red Food Sci Food Saf*. **3**:35–104, 2004. Only some of the nutrients measured are listed. Also shown are the ranges of nutrient values found in conventional corn grains in the United States. The values in the rightmost column were provided by the International Life Sciences Institute Crop Composition database (http://www.cropcomposition.org). Note: the ranges for individual amino acids could not be calculated from available data.

that are measured include carbohydrates, proteins, fats, amino acids, fatty acids, vitamins, and minerals (Tables 13.3 and 13.4). The natural variation in nutritional content among the conventional varieties of the food product due to varietal, developmental, or environmental factors is taken into consideration, and therefore there is usually a range of levels that is considered acceptable for a given food. Also considered are any changes in antinutritional factors that occur naturally in foods. These include chelators such as phytic acid that effectively remove calcium, iron, zinc, and magnesium, protease inhibitors that prevent protein digestion, and lectins that reduce the bioavailability of carbohydrates. The measurements are used to determine whether undesirable changes in nutrient content have occurred as a consequence of genetic manipulation, that is, whether levels of important nutrients are reduced or levels of potentially harmful compounds have increased.

Numerous feeding trials in which insect-resistant or herbicide-tolerant transgenic corn, rice, potatoes, soybeans, or tomatoes were fed to laboratory animals or livestock for prolonged periods, and often for several generations, have found no adverse effects related to nutrient deficiencies. There were no significant differences in the composition, quality, or digestibility of the food or in the development, health, or performance of animals fed genetically engineered versus conventional plants. For example, more than 20 studies have shown that the yield and quality (fat, lactose, and protein content) of milk produced by lactating dairy cows fed corn genetically engineered to be insect-resistant or corn, soybeans, or beets genetically modified to be tolerant to the herbicide glyphosate were the same as those from cows fed a nontransgenic diet.

Table 13.4 Comparison of nutritional content of a genetically engineered Atlantic salmon that expresses a growth hormone transgene with that of nontransgenic Atlantic salmon[a]

Nutrient (per 100 g)	Transgenic salmon	Nontransgenic salmon
Proximates (g)		
Protein*	19.13	20.16
Total fat*	14.42	9.14
Carbohydrate	0.38	0.37
Vitamins (μg)		
Folic acid	0.22	0.25
Niacin	97.46	88.66
Vitamin B_{12}	0.03	0.03
Minerals (mg)		
Calcium	27.57	30.03
Iron	0.52	0.48
Sodium	32.53	35.81
Amino acids (g)		
Alanine	1.10	1.17
Leucine	1.42	1.52
Phenylalanine	0.74	0.79
Fatty acids (g)		
Palmitic (16:0)*	1.79	1.07
Linoleic (18:2)*	0.74	0.51
Eicosenoic (20:1)	0.53	0.46

[a]Data from Veterinary Medicine Advisory Committee, FDA. 2010. *VMAC Briefing Packet Aqu Advantage Salmon*. Center for Veterinary Medicine, Rockville, MD. Only some of the nutrients measured are listed. *, nutrients that are statistically significantly different between transgenic and nontransgenic triploid female salmon; all are within the ranges of values for conventional salmon composition. (From Health Canada. 2016. Determination of the safety of AquaBounty Technologies Inc.'s Salmon (*Salmo salar* L.) event. Decision document DD2016-117.

Detailed analyses of the composition of products derived from cloned animals have led the U.S. FDA Center for Veterinary Medicine to conclude that there is no difference between the nutritional content of meat or milk from clones and conventionally bred animals. These studies measured milk yields and milk and meat nutrient content. It is acknowledged that an exhaustive analysis of tissue composition is impossible due to the complexity of the molecules present and that composition is impacted appreciably by the diet of the animal and its environment.

Several plants have been engineered to improve the nutritional value of food (Table 13.5). In these cases, the nutritional content of the genetically engineered food is intentionally made different from the conventionally bred plant and in this regard does not meet the criteria of substantial equivalence. An important example of a nutritionally enhanced food is "Golden Rice," which was created to address dietary deficiencies of vitamin A. Insufficient consumption of vitamin A is a significant problem in developing nations, where 250 million people suffer from vitamin A deficiency and each year hundreds of thousands of children are blinded due to retinal and corneal damage, suffer from infectious diseases, or die as a consequence of this nutritional deficiency. Carrots, tomatoes, meat, and milk are good dietary sources of vitamin A; however, in developing countries these are not always readily available. Rather, rice grains, which do

Table 13.5 Some genetically modified plants with enhanced nutritional value

Plant	Enhanced trait	Transgene product
Banana	Increased α- and β-carotene (vitamin A precursors)	Phytoene synthase
Canola	Increased vitamin E	γ-Tocopherol methyl transferase
Canola	Increased γ-linolenic acid (omega-6 fatty acid)	Δ^6 and Δ^{12} desaturases
Cassava	Decreased cyanogenic toxins	Hydroxynitrile lyase
Cassava	Increased iron and zinc	Iron transporter, ferritin
Coffee	Decreased caffeine	Antisense xanthosine-*N*-7-methyltransferase
Corn	Increased vitamin C	Dehydroascorbate reductase
Corn	Increased iron	Ferritin, phytase
Potato	Decreased solanine (glycoalkaloid toxin)	Antisense sterol glycotransferase
Rice	Added β-carotene	Phytoene synthase, phytoene desaturase, lycopene cyclase
Rice	Increased iron	Ferritin, metallothionein, phytase
Tomato	Increased β-carotene and lycopene	Lycopene cyclase, phytoene desaturase
Tomato	Increased flavonoids (antioxidants)	Chalcone isomerase
Tomato	Increased folate	GTP cyclohydrolase I, aminodeoxychorismate synthase
Wheat	Reduced gluten	Hairpin RNA targeting γ-gliadin mRNA

not contain vitamin A, are a primary food source. In 2000, after 7 years of research, scientists successfully engineered rice to produce β-carotene, a biosynthetic precursor of vitamin A, by introducing genes from daffodil and the bacterium *Erwinia uredovora* (see chapter 11). In the initial strain, the levels of β-carotene were too low to acquire the recommended amounts through diet, a problem highlighted by opponents of the genetically engineered rice. However, a subsequent version, announced in 2004 and called Golden Rice 2, produced 20 times more β-carotene, thereby reducing the amount of rice that would need to be eaten in order to obtain sufficient vitamin A to stave off malnutrition.

Among other extensive environmental and biosafety analyses, Golden Rice, like other nutritionally enhanced foods, was assessed for substantial equivalence in nutritional composition to nontransgenic rice. Except for β-carotene, the compound that was intentionally altered, levels of all nutrients measured in Golden Rice 2, grown at four sites in the Philippines over two seasons, were similar to those of nontransgenic control rice grown at the same sites and were within the range of variability of conventional rice varieties (Table 13.6). Based on mean β-carotene content, 40% of the recommended daily intake for children under 5 years old could be met with 100 g of Golden Rice 2, the average amount consumed daily by this age group. Despite the promise of this nutritionally enhanced food to allay vitamin A malnutrition, more than 20 years after it was first developed, Golden Rice has been approved in only a few countries (see chapter 11). Opponents argue that consumption of rice containing the vitamin A precursor is unnecessarily risky because other methods to deliver vitamin A are available, including vitamin tablets and vitamin-fortified foods and growing vegetables rich in vitamin A. However, these methods are often costly and difficult to sustain. The obstacles notwithstanding, holders of the various patented technologies that contributed to the development of Golden Rice have provided licenses to developing countries at no charge in anticipation of its eventual commercialization, as a humanitarian gesture to improve the health of billions of people.

Table 13.6 Nutritional content of grain from Golden Rice 2 and conventional rice[a,b]

Nutrient	Golden Rice 2	Conventional Rice[c]	Range[d]
Proximates (% dry weight)			
Protein	8.1	8.3	5.9–11.8
Carbohydrates	84.6	84.4	80.0–86.4
Fat	1.42	1.34	0.56–3.47
Total dietary fiber (% dry weight)	17.0	16.9	11.4–23.0
Minerals (mg/100 g dry weight)			
Calcium	22.5	21.4	10–150
Iron	3.96	4.57	1.6–9.1
Potassium	346	339	170–472
Amino acids (mg/100 g dry weight)			
Methionine	169	166	130–310
Lysine	299	294	210–430
Glycine	389	393	290–510
Fatty acids (% total fatty acids)			
Palmitic (C16:0)	19.5	18.5	14.9–31.2
Oleic (C18:1)	39.7	40.2	30.9–45.9
Vitamins (mg/kg dry weight)			
Thiamine (B1)	3.13	3.08	2.35–6.25
Niacin (B3)	36.9	32.6	20.2–65.0
Folic acid (B9)	0.91	0.88	0.39–1.54
Carotenoids (mg/kg dry weight)			
All-trans-α-carotene	0.71	<LOQ[e]	
All-trans-β-carotene	3.57	<LOQ	
Total carotenoids	5.88	<LOQ	

[a]The mean levels of only some of the nutrients measured are listed.
[b]Data from Swamy et al., *J. Agric. Food Chem.* **67:**7986, 2019.
[c]Parental rice strain from which Golden Rice 2 was derived.
[d]International Life Sciences Institute Crop Composition database.
[e]LOQ, limit of quantification.

Potential for Introducing Toxins or Allergens into Food

It is important to bear in mind that no foods can be guaranteed to be 100% safe. Many foods contain trace amounts of natural toxins that are harmless when consumed at low levels. For example, two common food plants, the potato and tomato, are members of the nightshade (*Solanaceae*) family that produce glycoalkaloids that can cause serious illness when consumed in relatively large quantities. Also, many foods naturally contain proteins that elicit an allergic response in some consumers. Extensive health safety tests in animals that compare genetically engineered foods to their nonengineered counterparts have concluded that the process of creating genetically engineered foods does not make the food different in digestibility or detrimental to the health of the consuming animal. However, it is necessary to ensure that the introduction of a specific transgene into a food product does not increase the risk of producing a toxin or allergen (Table 13.7). The regulatory agencies consider each food product derived from a transgenic plant or animal on a case-by-case basis and assess the risks against those associated with consuming the corresponding nontransgenic product.

Table 13.7 Some examples of animal feeding studies conducted to assess the potential toxicity of genetically modified (GM) foods[a]

GM food	Transgene[b]	Test animal	Duration of feeding	Health effects
Corn	*epsps*	Rat	13 weeks	No adverse effects; animals were similar in overall health, body weight, food consumption, organ weights, blood and urine chemistry, and tissue appearance
Corn	*cry1Ab*	Chicken	38 days	No significant difference in survival and body weight
Corn	*cry1Ab*	Pig	91 days	No difference in nutrient digestibility and energy content
Soybean	*epsps*	Rat	13 weeks	No difference in animal activity, body weight, blood chemistry, and urine chemistry; no gross abnormalities
Soybean	*epsps*	Mouse	Gestation and lactation	No difference in litter size, body weight of pups, and testicular development
Potato	*bar*	Rat	10 weeks	No significant difference in body or organ weights, food consumption, sperm motility, litter size, survival and body weight of pups; significant reduction in male and female fertility

[a]Adapted from Domingo, *Crit. Rev. Food Sci. Nutr.* **47:**721–733, 2007. The health effects are compared to those in animals fed non-genetically modified food.

[b]*epsps* confers tolerance for glyphosate herbicide, *cry1Ab* confers resistance to lepidopteran insects, and *bar* confers tolerance for glyfosinate herbicide.

A transgene that is commonly introduced into crops encodes the Cry protein, an insecticidal protoxin, from the bacterium *B. thuringiensis* (Bt). When ingested by target insect larvae, the protoxin is converted into an insecticidal toxin by specific insect proteases, and then the toxin interacts with specific receptors in the epithelium of the larval gut. The protein inserts into the epithelium, forming a pore through which gut contents leak, leading to death of the insect. The Bt toxin has been used safely for over 40 years to protect crops from a variety of insect pests, initially as an agricultural spray and more recently produced by transgenic plants. The Cry protein is produced by *B. thuringiensis* in an inactive form as a full-length protein. Proteases produced specifically by the insect's digestive system cleave the protein, which activates the toxin. The proteases are active in the slightly alkaline environment of the insect gut. The protein is safe for human consumption because humans do not produce the specific protease required to activate the toxin or the epithelial receptors that bind the toxin, and the toxin is rapidly degraded under the acidic conditions found in the mammalian digestive system. Nonetheless, some consumers have questioned whether the Cry protein may be a potential allergen. This concern is fueled in part by reports that are often based on misinterpreted scientific studies. For example, in a 2005 campaign against Bt rice, Greenpeace stated that rice genetically engineered to produce the Cry1Ac protein "could cause an allergic reaction, as it did when tested on mice." Several scientific studies were cited to support the claim that the protein elicited an allergic reaction in mice. However, it was not explained that the purpose of the cited studies was to test the Cry1Ac protein as an adjuvant to increase the efficacy of a vaccine or that the Cry1Ac protein was chosen because it was known to

have low toxicity in vertebrates. While anti-Cry1Ac antibodies typical of an immune response were produced in the immunized mice, these were not immunoglobulin E antibodies involved in the allergic response, nor was evidence of an allergic reaction presented. It was also not mentioned that Cry1Ac was already widely produced in crops consumed by humans and other animals without harmful effects. Interestingly, there has been almost no objection to the widespread spraying of the Bt protoxin, as a natural alternative to chemical insecticides, over the last four decades; opposition seems only to be voiced against its production in transgenic plants.

To further support the allergenicity of Cry1Ac, opponents made reference to concerns among regulators regarding the potential allergenicity of another Bt product, known as StarLink corn, an unfortunate association due to a regulatory breach by StarLink. The Bt gene used to generate StarLink corn encodes the Cry9C protein, which differs from Cry1Ac and other Bt toxins. Although Cry9C is not similar to known protein allergens and is not derived from an organism known to produce allergenic proteins, it has greater heat stability and is not as readily digested, which are characteristics of some allergenic proteins. For these reasons, the FDA did not approve StarLink corn for human use but did allow it to be sold as animal feed and used for industrial purposes. However, in July 2000, Larry Bohlen, director of the Community, Health, and Environment Program for the U.S. affiliate of the international environmental organization Friends of the Earth, went grocery shopping and bought a large number of products that contained corn flour, including taco shells, corn flakes, and muffin mixes. He had these tested for the StarLink Cry9C gene using the PCR test, and evidence of the gene was found in some of the taco shells. Clearly, food safety regulations had been broken. It was estimated that StarLink contaminated, at most, about 0.125% of all corn-based foods. The source of the contamination has never been traced. Aventis, the company that developed StarLink, has assumed that some farmers, probably inadvertently, mixed it with other varieties of corn despite written agreements that committed them to keeping their harvested Bt crops segregated. The seed is no longer sold.

There were a number of ramifications from the StarLink episode. The critics of biotechnology pointed out that the regulatory system was seriously flawed and stressed that consumers should be aware of the hazards of genetically modified organisms. The biotechnology industry representatives noted that there was no public health concern, but the regulatory issue had to be corrected. The FDA declared that it would no longer allow a product to be approved only for animal feed. A number of scientific advisory panels were convened to study in detail the consequences of Cry9C in human food, the possibility of consumer maladies due to eating products that may have been derived from StarLink, and how to improve the regulation and surveillance of transgenic crops. By 2001, despite the fact that there were no definitive cases of allergic reactions to Cry9C, individuals who had symptoms compatible with a possible newly acquired allergy were monitored. No serious illness has been documented as a result of eating food contaminated with StarLink.

The potential for introduction of allergenic proteins into a food product that does not otherwise elicit an allergenic response is a major

Table 13.8 Some foods containing allergenic proteins that may be reduced through genetic engineering

Food plant	Allergenic protein	Strategy to reduce allergenic protein
Rice	14–16 kilodalton allergen	Antisense RA17 gene silencing
Soybean	Gly m bd 30 K (P34; cysteine protease)	P34 sense cosuppression
Apple	Mal d 1	Mal d 1 RNA interference
Tomato	Lyc e 1 (profillin)	Lyc e 1 RNA interference
Tomato	Lyc e 3 (lipid transfer protein [LTP])	LTP RNA interference
Peanut	Ara h 2 (conglutin 7)	Ara h 2 RNA interference

concern. The FDA requires labeling to indicate genetically engineered foods containing proteins from organisms known to produce proteins that are allergens to humans. Major allergens are proteins from eggs, milk, shellfish, fish, tree nuts, soybeans, wheat, and peanuts. In most cases, such transgenic food products are not commercialized. For example, soybeans engineered to express a protein from Brazil nut that was intended to increase the methionine content of soybeans was found to react with the sera from individuals known to be allergic to Brazil nuts. As a consequence, further development of the transgenic soybeans was voluntarily terminated by the developer. Also of concern is the potential for a genetically engineered food to have increased levels of an endogenous allergen as an unintended consequence of transgene expression, although there may be a large variation in the amount of such a protein in conventionally bred foods.

Food allergies are a significant problem in developed countries. In the United States about 5% to 8% of children and 1% to 2% of adults have food allergies, some of which are fatal. In fact, molecular biotechnology has the potential to reduce the allergenicity of some foods by preventing the synthesis of the allergenic protein (Table 13.8). For example, a hypoallergenic variety of peanut is currently under development using RNA interference to reduce levels of the seed storage protein Ara h 2, a potent peanut allergen to which most hypersensitive individuals respond, often fatally.

Potential for Transferring Transgenes from Food to Humans or Intestinal Microorganisms

We consume a great amount of DNA, and the encoded protein products, in the plant and animal tissues that make up our daily diet. Indeed, these supply part of the nutrition from food. It has been estimated that an average adult consumes approximately 0.1 to 1 gram of DNA per day. The DNA is partially degraded during food processing at high temperature or low pH and is further degraded into small fragments through chewing and the activities of nucleases in saliva and the gastrointestinal tract. Only a very small amount of the total DNA consumed remains in fragments that are capable of carrying an intact gene as the digested DNA passes through the small intestine. This was shown in patients who had an ileostomy, an operation in which the lower portion of the intestinal tract, the ileum, is severed from the upper portion of the small intestine, which allowed researchers to collect digestion products before they completed passage

through the small intestine and colon. After eating soybeans genetically engineered with the *epsps* gene (encoding 5-enolpyruvylshikimate-3-phosphate synthase, which confers tolerance to the herbicide glyphosate), up to 3.7% of the consumed transgene was detected by PCR as a 180-bp fragment in the digestion products of the ileostomists. However, the *epsps* gene fragments were not detected in the feces of healthy volunteers with intact intestinal tracts and therefore were completely digested in the small and large intestines.

If the genes that are present in the food that we eat remain intact in our intestinal tract, can they be incorporated into our genomes, where they could disrupt the function of a normal gene or where they could be expressed, resulting in the production of a foreign protein? Several studies have shown that neither transgenic DNA, nor the recombinant proteins encoded by the transgenes, present in food are found in the tissues of humans and livestock. PCR and Southern hybridization failed to detect recombinant DNA in the milk, eggs, skin, muscle, and other tissues of several livestock animals fed genetically engineered crops. Nonrecombinant chloroplast DNA, which is naturally present in multiple copies in plant cells, was detected in some tissues. This may have implications for transgenes that are incorporated into the chloroplast genome. Even when present in tissues, integration of functional genes from food DNA, transgenic or otherwise, into the genome of the consuming animal has never been found and is likely restricted by (i) the small fragment sizes of the DNA remaining in the gastrointestinal tract following food digestion, (ii) the presence of cytosolic nucleases that further digest foreign DNA that is taken up by intestinal epithelial cells, (iii) restricted access to the nucleus, (iv) insufficient sequence homology required for integration by homologous recombination, (v) the requirement for appropriate transcription and translation signals for expression of an integrated gene, and (vi) transcriptional silencing of integrated foreign DNA by methylation of cytosines at CpG dinucleotides.

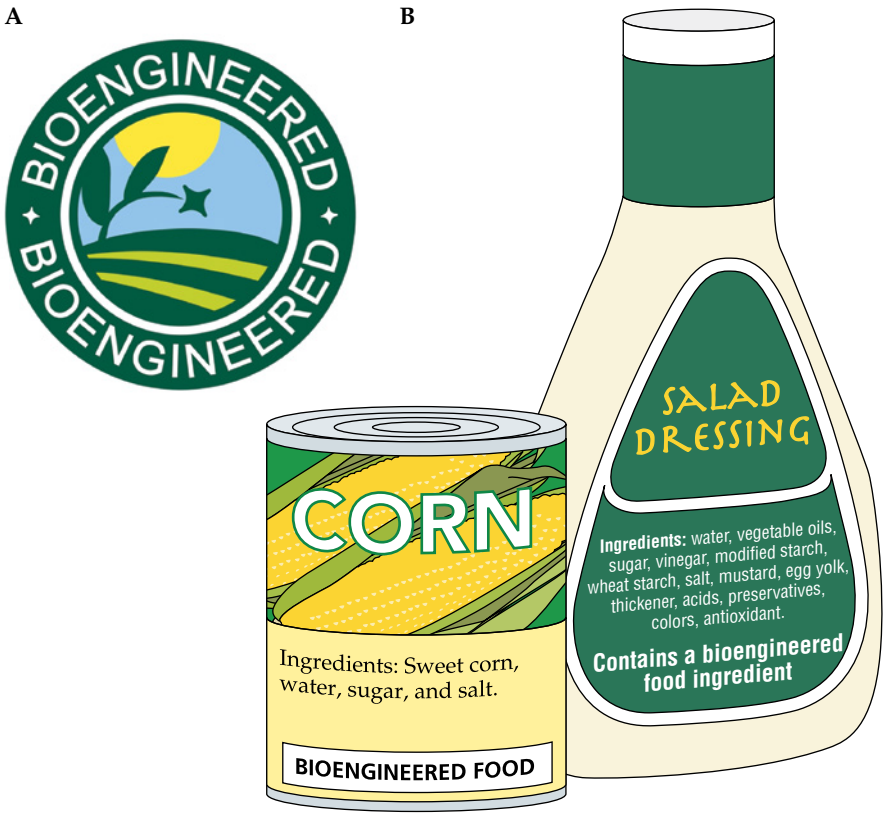
The possibility that antibiotic resistance genes used as selectable markers during the process of creating transgenic plants could be transferred to microorganisms in the intestinal tract has also been investigated. Because most of the DNA derived from food that reaches the intestines, and particularly the colon, where most gut microorganisms reside, has been digested into small fragments, the risk of an intestinal microbe taking up and expressing a functional antibiotic resistance gene is extremely low. For example, an antibiotic resistance gene in transgenic corn was undetectable after one minute in sheep rumen fluid and in the intestines of chickens. Although there is no evidence from long-term studies of a variety of animals that antibiotic resistance genes consumed in transgenic food have been transferred to intestinal bacteria, to alleviate public concerns, methods that avoid the use antibiotic resistance genes have been developed (see chapter 10).

Controversy about the Labeling of Genetically Modified Foods

The U.S. regulations have long held that the composition of a food product, not the process by which the product is produced, determines whether specific information should be added to a label. For example, corn syrup derived from Bt corn is identical to that from conventional corn.

Consequently, until recently, this type of product did not require a special label. Labeling was required if the nutrient content of a genetically modified food is substantially different from that of the traditional product, if the food is novel (i.e., it never has been produced before), if it is likely to contain a potential allergen, or if it has an increased level of a toxic agent. However, the labeling of biotechnology-derived foods remains a contentious issue. In 2014, the U.S. state of Vermont passed legislation requiring the labeling of raw and processed foods containing genetically modified ingredients, and similar legislation was under consideration by several other states. Subsequently, to create a national standard, in July 2016 the National Bioengineered Food Disclosure Law was passed requiring mandatory labeling of foods containing genetic material that has been modified through recombinant DNA techniques and cannot be obtained through conventional breeding or found in nature. The food must be labeled with a symbol (Fig. 13.1A), a text (Fig. 13.1B), an electronic link such as a barcode that leads to a website, or a text message. Refined foods derived from genetically engineered organisms in which the modified

FIGURE 13.1 Labels for foods that contain genetically modified ingredients. Under the U.S. National Bioengineered Food Disclosure Standard, mandatory labels for genetically modified foods must include either the approved symbol **(A)**, text **(B)**, an electronic or digital link that when scanned leads to a website, or a text message ("Text [command word] to [number] for bioengineered food information."). In addition to including a bioengineered food disclosure, the product information web page or text message must not contain marketing or promotional information or collect information about consumers.



genetic material is undetectable in the final product are excluded from the regulations. Foods served in restaurants, produced by small food manufacturers, and from animals that consumed genetically modified feed are exempted. Opponents argue that science has proven the foods to be safe and that labeling poses a burden on the food industry. For example, mandatory labeling may cause manufacturers to reformulate their products using nongenetically modified ingredients to avoid adopting a label that may be perceived as a warning and thereby reduce sales. Proponents believe that consumers have a right to know what they are eating.

Similar regulations are in place in other countries. It is mandatory for all European Union countries to label any food as "genetically modified" if more than 0.9% is from a genetically modified organism. The genetically modified food must also be traceable from its source on the farm through all stages of processing, storing, and transporting. This is to facilitate removal of the product should it be found to have adverse effects. Australia and New Zealand require the same designation if "novel DNA and/or protein is present in the final product" but not if the food is highly refined or there are no genetically modified ingredients in the final product derived from a transgenic organism. In most cases, the stringency of the labeling is not solely dependent on health concerns, but often, economic and philosophical issues are also important considerations.

Impact of Genetically Engineered Crops on Biodiversity

There are some concerns that our increasing dependence on genetically engineered crops is adversely affecting the environment by decreasing biodiversity and harming unintended organisms. Agriculture in general has decreased biodiversity. Crop plants are selected for high yields of their edible parts, hardiness, and other traits at the cost of loss of richness not only of cultivated species but also of other wild plants as cultivation has expanded over a greater area. Newer, higher-yielding varieties have tended to replace traditional crops. The diversity of insects, birds, and other animals has also decreased as more land is dedicated to agriculture and host and food plants are lost. Application of agricultural chemicals to eradicate weeds and pests is a contributing factor. There is some concern that the trend toward extensive cultivation of crops that have been genetically engineered to resist environmental factors that normally limit plant growth (e.g., herbicide tolerance and insect resistance) could further decrease biodiversity. This concern stems mainly from the possibility that the transgenic plants could become weedy or invasive of natural habitats or that the transgenes can be transferred from genetically engineered crops to wild relatives or to nongenetically engineered crops and could confer on them a selective advantage. Enhanced growth of the unintentionally engineered plants could cause them to be invasive or to outcompete plants without the transgene.

Several species of weeds have been found to be resistant to the herbicide glyphosate. For the most part, this is due to overuse of the herbicide which selects for natural mutations that increase resistance in weeds rather than to the transfer of the transgene through cross-breeding (hybridization), although the latter has been shown to occur at a low frequency. Pollination of wild relatives by genetically engineered crops can occur when they are within range of pollen distribution by wind, insects, etc., and when pollen

production by the genetically engineered crop and flower production by the wild relative occur at the same time. Transfer of genes among plants that mainly self-pollinate (pollination of a flower by its own pollen), such as rice and soybeans, occurs at a low frequency. For example, transfer of an herbicide resistance gene from genetically engineered rice to wild red rice was found to occur at a frequency of less than 1%. The frequency of transfer is higher for plants that cross-pollinate (the pollen from one plant fertilizes the ovule of another plant). The glyphosate resistance gene from transgenic canola (*Brassica napus*), an important oilseed crop, was transferred to a wild relative (*Brassica rapa*, field mustard) at a frequency ranging from 7% to 14%, depending on the *B. rapa* population. This was determined by collecting seeds from herbicide-sensitive *B. rapa* plants that had been interplanted with herbicide-tolerant transgenic *B. napus* and then assessing herbicide tolerance and production of the specific protein encoded by the transgene in the plants that developed from the *B. rapa* seeds. While cross-pollination occurred at a relatively high frequency between *B. napus* and *B. rapa*, hybridization between *B. napus* and three other wild relatives was extremely rare.

The spread of herbicide resistance genes reduces the ability of farmers to control weeds, and the weedy plants have the potential to become invasive, as they can proliferate in areas where the herbicides are used. Herbicide-resistant plants are unlikely to have an advantage in nonagricultural areas where herbicides are not applied. Moreover, hybridization does not always confer an advantage on the hybrid progeny. Hybrids resulting from cross-pollination by transgenic rice resistant to glyfosinate (an herbicide that disrupts glutamine biosynthesis) were less fit than their wild-type parents because flowering occurred too late to produce seed, which prevented proliferation of the hybrids. Hybrids that express a Bt toxin transgene may have a significant advantage when insect pressure is high.

The potential for gene flow to nontransgenic cultivars and wild relatives and its consequences are included in risk assessments required by regulatory agencies before a genetically engineered crop is commercialized, and strategies to manage the risks must be in place prior to cultivation. The assessment considers the mechanism by which the plant reproduces and pollen is disseminated, the nature of the transgene and any selective advantage it may impose, and the geographic context in which the crop will be grown, which includes the identification of sexually compatible wild relatives. Management strategies can include avoiding cultivation of a transgenic crop in areas where indigenous wild varieties are found and controlling embryo and/or seed viability through genetic engineering.

There are also fears that cultivation of transgenic crops may have a detrimental impact on other (nontarget) organisms, especially insects. In May 1999, a report titled "Transgenic Pollen Harms Monarch Larvae" by Losey et al. was published in the scientific journal *Nature*. In this study, pollen from Bt corn was sprinkled on milkweed leaves, the sole source of food for the monarch butterfly, to simulate wind distribution from transgenic crops. Monarch butterfly larvae that fed on leaves treated with transgenic pollen had a higher mortality rate than those that consumed leaves treated with nontransgenic pollen or leaves without pollen. The authors concluded that extensive Bt crop acreage could have "profound implications for the conservation of monarch butterflies." The study, which was conducted in the laboratory rather than under natural field conditions,

incited much criticism. Scientists pointed out that the amount of pollen on the leaves was much higher than would be found on leaves of field-grown milkweed and that the larvae were provided only with pollen-covered leaves and therefore could not choose to avoid eating them. These technical concerns were largely ignored because the monarch butterfly has iconic and aesthetic status in North America. Consequently, the study was widely reported and the threat to the monarch butterfly formed the cornerstone of the campaign against biotechnology in general. However, two years of detailed studies showed that the risk of Bt toxicity to the monarch butterfly in the field was negligible.

More recently, reports in the popular media have emerged suggesting that pollen containing the Bt toxin was responsible for severe reductions in the honeybee population in North America and Europe. Declines in the honeybee population, a phenomenon known as colony collapse disorder, were occurring at an alarming rate, and the consequences would be very serious if pollinators were not available to produce fruits. However, an analysis of 25 independent studies concluded that pollen containing the Cry proteins that are present in transgenic crops currently under cultivation and are effective against coleopteran and lepidopteran predators are not toxic to honeybees, a hymenopteran insect. After a great deal of investigation, the general consensus among researchers is that because the Bt toxins are highly specific for target insects and are confined to plant tissues, the impact on nontarget insects is minimal. Several studies have shown that insect diversity and abundance have actually increased with the cultivation of Bt crops compared to application of chemical insecticide sprays that have a much broader target range.

Who Benefits from the Production of Genetically Modified Foods?

Developers of biotechnology products that have applications in agriculture have benefited financially from recombinant DNA technology. The global value of biotechnology-derived crops was $18.9 billion in 2018. Many smaller developers complain that the high cost of developing and commercializing a genetically engineered product prevents all but large multinational companies from profiting. For example, in the agricultural biotechnology sector, where the cost of developing a genetically engineered product can exceed $20 million due in part to stringent regulatory requirements, four companies own or co-own about 80% of the genetically engineered products that have received commercial approval: Bayer CropScience (which acquired Monsanto in 2018; Germany), Syngenta (acquired by ChemChina in 2016; Switzerland), Corteva Agriscience (a division of Dow Chemical and DuPont, which merged in 2019; United States), and BASF (Germany). High research and development costs have impeded the commercialization of products that are not marketable on a large, profitable scale. This means that transgenic crops other than herbicide-tolerant and insect-resistant crops that are not widely planted do not provide the revenues that are necessary to justify the high cost of product development. Improved varieties of crops that are important for resource-poor nations have not been developed as extensively, although this is beginning to change as exemplified by the approval of pod-borer-resistant cowpea in Nigeria in 2019. In addition, product development has

been limited by antibiotechnology markets such as the European Union, which in turn influences development in countries that rely on export to these markets.

Farmers in both developed and developing nations have perhaps been the greatest beneficiaries of recombinant DNA crop technology not only due to increased yields from genetically engineered crops but also from cost savings due to fewer applications of agricultural chemicals and to reduced labor and machinery costs. Profits are generally higher even after the higher cost of the transgenic seed is taken into consideration. Over the past 20 years, genetic engineering has increased crop yields by 22%. Much of this is due to increased resistance to insects. Overall, profits to farmers have increased by 68%.

Most of the economic benefit of this technology has accrued to farmers in developing countries, where 90% of farmers growing biotechnology-derived crops are from small, resource-poor farms. In particular, farmers in developing countries profited from cultivation of herbicide-tolerant soybeans and insect-resistant cotton. For example, a multiyear comparison of the productivity and profitability of small, resource-poor farms in South Africa growing nontransgenic or bollworm-resistant Bt cotton found consistent yield increases of 156% to 185% for those growing Bt cotton. Although the Bt seed costs were roughly twice those of the conventional cotton, profits from the insect-resistant cotton were at least double due to higher yields, reduced pesticide costs, and reduced labor costs to spray the pesticide. In 2019, farmers in Bangladesh growing Bt brinjal (eggplant) saw yield increases of 20% and income increases of 22% compared with those growing conventional brinjal. Since its adoption in 2014, the number of Bangladeshi farmers cultivating Bt brinjal has increased 1,000-fold. For most insect-resistant crops, the greatest economic benefits are reaped in seasons when insect infestation is high. Because damage from insect predation is lower for the Bt crops, fungal infestations, which are often facilitated by insect predation on plant tissues, are also reduced. Cereals, oilseeds, and nuts contaminated with fungi are not suitable for sale due to increased risk of contamination with mycotoxins such as aflatoxin, a carcinogenic toxin produced by *Aspergillus* spp. The economic benefits due to reduced mycotoxin contamination are significant at $30 million.

Environmental Benefits of Genetically Modified Crops

Ultimately, it may be all consumers that benefit from enhanced productivity from genetically engineered plants as we struggle to feed an increasing global population that is expected to reach over 10 billion by 2050. Already the burden is high, with only 0.2 hectares of arable land available per person, when the minimum is calculated to be 0.5 hectares per person. The amount of land suitable for crop cultivation may be further reduced in the future due to major climatic shifts, increased global conflicts, competition with nonfood crops such as those for the bioenergy market, pollution, and deterioration of soil quality from overuse of agricultural chemicals. Production of transgenic crops can offer solutions to some environmental problems by increasing the resistance of important crop plants to infectious disease, insect predation, and climate change and reducing applications of pesticides.

Cultivation of plants that produce heterologous proteins that provide protection from pests has led to a global reduction in the application of

agricultural chemicals, many of which have toxic effects on the environment. Spraying of insecticides has been reduced by 42% worldwide since the introduction of insect-resistant crops. Cultivation of insect-resistant cotton and corn accounts for much of this reduction. For example, the western corn rootworm is a devastating and pervasive insect pest of corn and causes huge economic losses in the United States and elsewhere. Because the predator is difficult to control, farmers routinely apply insecticides to their cornfields before infestation of the worm is apparent. The cultivation of genetically engineered varieties of corn that produce the Cry3Bb1 toxin, which is effective against the western corn rootworm, have significantly reduced the spraying of insecticides against that insect.

Although some would argue that the widespread cultivation of herbicide-tolerant crops has increased our dependence on herbicides, it has actually resulted in a reduction in herbicide applications to transgenic crops compared to conventional crops. For example, cultivation of herbicide-tolerant cotton reduced global herbicide use by 9.6% from 1996 to 2018. Cultivation of herbicide-tolerant corn reduced herbicide use by 7.3% over the same period. While the use of glyphosate, the herbicide to which most crops are engineered to be tolerant, has increased sharply over the last decade, the use of other herbicides has decreased. The herbicides that are currently applied are generally of lower strength and are less toxic than the previous generation of herbicides. In addition, the herbicides can be applied to fields containing tolerant crops later in the season, after the crops have grown. This has dramatically reduced (and on many farms eliminated) soil tilling, the practice of turning the soil to remove weeds, and has thereby increased soil quality by reducing erosion and organic carbon loss and decreased fuel consumption by farm machinery. This latter benefit is not trivial in its impact on the environment, as the estimated annual reduction in emissions of carbon dioxide, a greenhouse gas, as a consequence of growing herbicide-tolerant soybeans, corn, and canola, is the equivalent of removing 1.4 million cars from the road for one year.

Because different agricultural chemicals have different environmental toxicities, a better measure of their impact on the environment takes into consideration not only the amount of active ingredient applied, but also the toxicity of the chemical to farm workers, the consumer, and other organisms such as birds and insects, its persistence in the soil, and its potential to leach into groundwater. These parameters are used to determine the relative environmental impact of each pesticide. Using these values, the global reduction in environmental impact resulting from changes in herbicide and pesticide applications to genetically modified crops was estimated to be 18.5% over the last 22 years.

Many other examples have been presented throughout this book that illustrate the potential of genetic engineering to reduce damage to the environment. Plants, and plant-associated bacteria, can be genetically engineered to more effectively remove toxic compounds such as heavy metals from contaminated soils. Moreover, plants and animals can be engineered to more efficiently utilize nutrients and to grow under nonoptimal conditions. These features could enable cultivation of crops on less land or on land that would otherwise not be usable, thereby meeting the food demands of an increasing global population with a reduced impact on limited resources.

How Do Views about Genetically Engineered Organisms Impact Trade?

Differences in consumer acceptance and regulatory views have often made trade in agricultural products between regions difficult. For example, sales of Hawaiian papaya to Japan decreased from $15 million in 1996 to $1 million in 2010 after papaya engineered to resist the devastating ring-spot virus was widely grown in Hawaii and was waiting for approval for importation by the Japanese regulatory agency. In 1998, the European Union placed a moratorium on the importation of genetically modified corn, cotton, and soybean products. This prompted a complaint in 2003 by the United States, Canada, and Argentina to the World Trade Organization, which ruled in 2008 that the additional risk assessments and approvals required by the European Union for some of the genetically engineered products were not based on scientific evidence and unfairly delayed trade. In contrast to the United States, only a few genetically engineered crops have been approved for cultivation and importation in the European Union, and to avoid labeling foods that contain more than 0.9% of an approved genetically engineered ingredient, exporters to the European Union have been forced to separate genetically modified crops from conventional crops. The inadvertent presence of genetically modified varieties in exports can be costly. For example, genetically engineered flaxseed was detected in Canadian flax shipments to the European Union in 2009, resulting in suspension of all flax shipments from Canada for several months, costing the Canadian flax industry about $25 million and the European industry about $48 million. Despite the view in many other countries that genetically modified foods are safe, Europeans remain strongly opposed to the products. The resistance of the European Union to transgenic products has forced many developing countries to forego the benefits of planting genetically engineered crops for fear that trade with the European Union will be jeopardized.

Regulation and Safety of Medical Products of Biotechnology

There is some level of risk of unintended, harmful effects associated with all medical interventions. For example, all medications may cause side effects. The risk is reduced as much as possible by the requirement for clinical trials before all new human therapeutic agents are made available to the public. However, there may be consequences that cannot be predicted from testing a drug on this smaller group. Adverse reactions may be due to an inappropriate dose of a drug, an unforeseeable reaction to a drug such as an allergic reaction, interactions with other drugs, or the state of health or genetic background of a patient. Thus, health care professionals must continue to monitor new therapies after they are released. In most countries, regulations for the approval of new pharmaceuticals for commercial use are sufficient to ensure the efficacy and safety of a drug regardless of how it is produced. In the United States, the FDA is responsible for regulating the introduction of drugs and medical devices into the marketplace; in the European Union, the European Medicines Agency fulfills this role.

New Biological Drugs

The FDA's Center for Drug Evaluation and Research evaluates new drugs under the Federal Food, Drug, and Cosmetic Act before they are made available to the public. Along with a diverse range of chemically synthesized drugs, this includes biological molecules, whether recombinant or not, such as hormones (e.g., insulin, human growth hormone), therapeutic proteins (e.g., interferon, hyaluronidase), and monoclonal antibodies. While the FDA assesses applications for drug approval, the producer must test the drug and provide evidence that it is safe and effective. The drugs are first tested in laboratory animals (preclinical trials), and those that show promise are then tested in humans in clinical trials for safety, including toxicity and the severity of side effects (phase I), treatment efficacy in those affected with the disease (phase II), and optimal dosage, interaction with other drugs, and efficacy in a larger number of patients and in different populations (phase III). Approval is usually limited to specific applications for which the drug was shown to be effective in clinical trials. For example, the FDA approved a drug called Kadcyla (ado-trastuzumab emtansine) for treatment of breast cancer. The drug comprises the monoclonal antibody trastuzumab, which specifically targets HER2, a protein that is elevated in 20% of breast cancers and contributes to cancer cell proliferation, and DM1, which inhibits cell growth. The new therapy is approved only for HER2-positive women with metastatic breast cancer who have been previously treated with trastuzumab (not conjugated to DM1) and the chemotherapeutic drug taxane. Because it is intended for use in patients for whom there is no satisfactory alternative treatment available, the drug was reviewed under an expedited FDA program.

Premarket regulation of therapeutic agents derived from living organisms, including genetically engineered organisms, is the mandate of the FDA Center for Biologics Evaluation and Research under the Public Health Service Act. A genetically engineered animal must also receive approval as a new animal drug under the Federal Food, Drug, and Cosmetic Act, as described above for genetically engineered livestock. Products from living organisms are known as biologics and include vaccines, blood and blood products (including recombinant plasma proteins such as clotting factors), tissue and cell-based treatments, and gene therapies (Table 13.9). Many vaccines have been licensed for immunization against bacterial and viral pathogens in the United States. All of these have undergone rigorous testing, including preclinical and clinical testing, to ensure that they are safe, free of contaminants, and effectively protect people against the targeted infection. In addition, they continue to be monitored for side effects once they are in use. For example, a subunit vaccine containing the major capsid proteins of four different types of human papillomaviruses (HPV) was approved by the FDA in 2006 for immunization of young females (9 to 26 years of age) against genital warts and cervical, vulvar, vaginal, and anal cancers caused by human papillomaviruses (see chapter 7). It was approved for boys in 2011. Safety and efficacy were established in clinical trials involving 21,000 girls and women. From 2006 to 2014, approximately 25,000 adverse events out of 67 million doses administered were reported, less than 8% of which were serious enough to require hospitalization or result in permanent disability or death. Serious adverse events included hypersensitivity reactions, Guillain-Barré syndrome (a neurological disorder that causes muscle weakness), blood clots, and seizures. Review of the adverse events reports by the FDA and the Centers for Disease Control and Prevention led to the conclusion

Table 13.9 Examples of biologics approved by the FDA in 2019–2021[a]

Proper name	Trade name	Application[b]
Coagulation factor VIIa	Sevenfact	Recombinant protein for treatment of bleeding episodes in hemophiliacs
Antihemophilic factor, glycopegylated	Esperoct	Recombinant protein for treatment of bleeding episodes in hemophiliacs
Immune globulin	Xembify	Antibody replacement therapy for primary humoral immunodeficiencies
Meningococcal conjugate vaccine	MenQuadfi	Vaccine for prevention of meningococcal disease caused by *Neisseria meningitidis*
Influenza A (H5N1) monovalent vaccine	Audenz	Vaccine to prevent infection with influenza A virus H5N1 subtype
Lisocabtagene autoleucel	Breyanzi	CAR T-cell immunotherapy to treat large B-cell lymphoma
Brexucabtagene autoleucel	Tecartus	CAR T-cell immunotherapy to treat mantle cell lymphoma
Onasemnogene abeparvovec	Zolgensma	Gene therapy for spinal muscular atrophy

[a]Source: www.fda.gov/vaccines-blood-biologics/development-approval-process-cber/biological-approvals-year.
[b]CAR, chimeric antigen receptor.

that rates for these events reported following HPV immunization were not greater than expected for the general population of young females, and there was no evidence that the deaths were caused by the vaccine. Despite data supporting the safety of this vaccine, vaccination rates are low (roughly half of adolescents in the United States have received the complete 3-dose series), while more than 500,000 new cases of cervical cancer arise worldwide each year, with 270,000 deaths. Reluctance to be immunized has contributed to a large increase in the number of cases of other preventable infectious diseases such as pertussis and measles (see Box 13.1).

The first human gene therapy (Luxturna) was approved by the FDA in 2017, and since then several others have received approval in the United States and the European Union to treat certain cancers and genetic disorders (Table 13.10). Some of these are *ex vivo* gene therapies in which target cells are harvested from a patient, transfected with a transgene, and then reintroduced into the patient (see chapter 6). Examples include T cells (a type of white blood cell) engineered to express a surface chimeric antigen receptor (CAR) that, when reintroduced into the patient, direct the T cells to destroy specific, often cancerous, cells that display the antigen (CAR T-cell immunotherapy). Others are *in vivo* gene therapies in which a functional copy of a gene is injected directly into a target tissue to replace a nonfunctional copy of the gene in the patient (see chapter 6). One such *in vivo* gene therapy, onasemnogene abeparvovec (sold under the trade name Zolgensma), consists of the human *SMN1* gene encoding survival motor neuron protein that is normally expressed at high levels in the spinal cord and is important for maintenance of motor neurons. Individuals with mutations in the *SMN1* gene that result in a deficiency of the protein lose motor neurons, which leads to muscle weakness and wasting, a disease known as spinal muscular atrophy. Following phase I clinical trials to establish safety and dose, in an ongoing phase III clinical trial, the *SMN1* gene carried on an adeno-associated virus vector was administered by intravenous infusion to 21 symptomatic patients under 6 months old. At 14 months old, 91% of these patients were alive and did not need respiratory support, compared to 25% of those who did not receive treatment. Normally, 18-month-old infants with this disease cannot sit without support; however, following *SMN1* gene therapy, 59% could sit independently for at least 30 seconds,

Antivaccine Movement

box 13.1

During the 2020–2021 COVID-19 pandemic, protesters gathered across North America and Europe to voice their opposition to COVID vaccines, in some cases disrupting vaccination clinics. These protests began before vaccines were available and continued even when SARS-CoV-2 infections surpassed 100 million worldwide. Despite the fact that thousands of people were dying from COVID every day in the United States, 20% of Americans claimed that they would not take the vaccine, and more than 30% were not sure if they would (Cornwall, *Science* **369:**14, 2020). Most of these are generally opposed to all vaccines. The antivaccine movement is currently small but is growing and widespread; the World Health Organization lists vaccine hesitancy as a top 10 global health threat. Antivaccine advocacy groups organize campaigns to spread doubts about vaccines, usually by appealing to emotions (largely fear), and in some cases, to promote conspiracy theories. Social media is an effective conduit for dissemination of false information about vaccines.

The reasons for vaccine opposition are varied but typically include the belief that vaccines are unsafe, ineffective, and/or unnecessary, despite a large body of scientific data that indicate otherwise. Misconceptions about vaccines are not new, although fears about their health risks were mobilized after publication of an article in February 1998 by British physician Andrew Wakefield suggesting a causal link between the measles, mumps, and rubella (MMR) vaccine and autism in children (Wakefield et al., *Lancet* **351:**637). In associating the MMR vaccine with a "regressive developmental disorder" reported as autism in eight children, Wakefield and colleagues stated that "the average interval from exposure to first behavioral symptoms was 6.3 days." The research received much attention in the media. It has since been revealed that there were many discrepancies in the report (described in detail by B. Deer, *BMJ* **342:**c5347, 2011), and an association between the MMR vaccine and autism has not held up in subsequent scientific studies (e.g., Jain et al., *JAMA* **313:**1534, 2015). Wakefield's article was retracted in 2010; however, the damage was done.

Vaccine opponents argue, without credible substantiation, that not only do vaccines pose a health risk that outweighs the benefits, but they impede natural development of immunity, which is claimed to be more effective, and they are unnecessary because infectious diseases are no longer a great threat. Underlying these beliefs is often a suspicion that governments and large corporations, especially pharmaceutical companies, conceal the real risks of vaccines. There is a distrust among many Americans, particularly those from minority groups, of medical authorities, in large part due to a history of medical research abuses, such as the Tuskegee Study conducted by the U.S. Public Health Service, in which hundreds of African American men known to have syphilis were not treated with penicillin, in order to study the natural progress of the disease. Others feel that mandatory vaccines for school-age children impinges on personal freedoms, which they believe are slowly being eroded. These protesters feel that they are defending individual liberties, and many have connections to far-right groups.

The antivaccine campaigns have been effective, as vaccination rates have plateaued in the last 20 years. While 94% of parents in 2001 felt it was important to vaccinate their children, this dropped to 84% in 2019. Certainly, vaccine opponents have a right to express their concerns, but false information about vaccine safety and efficacy can undermine efforts to reduce infection rates. Vaccine-preventable infections and deaths have increased. For example, in 2019 there were 663 cases of measles, a highly infectious and often fatal disease that was declared eliminated in the United States in 2000; 88% of cases were in underimmunized communities. Public health officials are concerned that fears raised by antivaccine campaigns will thwart efforts to achieve the herd immunity necessary to end the COVID-19 pandemic.

To counter the antivaccine movement, accurate information about vaccine safety and efficacy must be more effectively disseminated to correct misconceptions. Strategies similar to those used by antivaccine advocacy groups that appeal to emotions may be more effective than relying solely on facts and statistics, which tend to be less memorable. False claims should be responded to promptly because, as a quote attributed to American writer Mark Twain states, "a lie can travel halfway around the world while the truth is putting on its shoes." The truth is that vaccines save lives.

and remarkably, one could stand and walk. In 2019, the U.S. FDA approved the gene therapy for pediatric patients under 2 years old with biallelic mutations in the *SMN1* gene. Because the long-term consequences of gene therapies are unknown, patient follow-up is required for several years after treatment. About 400 patients have been treated with Zolgensma and

Table 13.10 Some approved gene therapies[a]

Trade name	Transgene	Target cells	Application	Approval year (agency)
Strimvelis	Human *ADA* gene encoding adenosine deaminase, carried on a retroviral vector	Hematopoietic stem cells (*ex vivo* therapy)	Severe immunodeficiency caused by adenosine deaminase deficiency	2016 (EMA)
Luxturna	Human *RPE65* gene encoding retinoid isomerohydrolase, carried on AAV2	Retinal cells (*in vivo* therapy)	Retinal dystrophy, an inherited blindness caused by a biallelic *RPE65* mutation	2017 (FDA) 2018 (EMA)
Kymriah	CAR gene encoding an scFv that targets B-cell CD19 surface antigen	Patient's T cells (*ex vivo* therapy)	B cell lymphoblastic leukemia in children	2017 (FDA) 2018 (EMA)
Yescarta	CAR gene encoding an scFv that targets B-cell CD19 surface antigen	Patient's T cells (*ex vivo* therapy)	Large B-cell lymphoma	2017 (FDA) 2018 (EMA)
Zolgensma	Human telomeric *SMN1* gene encoding a motor neuron protein, carried on AAV9	Motor neurons (*in vivo* therapy)	Spinal muscular atrophy in children under 2 years old	2019 (FDA) 2020 (EMA)
Zynteglo	Human *HBB* gene encoding β-globin	Hematopoietic stem cells (*ex vivo* therapy)	β-thalassemia caused by a hemoglobin deficiency	2019 (EMA) 2021 (FDA)
Abecma	CAR gene encoding an scFv that targets B-cell maturation antigen	Patient's T cells (*ex vivo* therapy)	Multiple myeloma	2021 (FDA)

[a]EMA, European Medicines Agency; FDA, Food and Drug Administration, USA; CAR, chimeric antigen receptor; scFv, single-chain antibody variable fragment; AAV, adeno-associated virus.

continue to be monitored for adverse reactions, which include acute serious liver injury, thrombocytopenia, and increased levels of cardiac troponin-1.

Given the enormous potential of genome editing technology to treat human diseases, it is essential that guidelines for its research and clinical use be established. A major safety concern with this technology is the potential for mutations at unintended sites in the genome. Early in December 2015, prominent biomedical researchers, legal experts, ethicists, and advocacy groups met at an international summit to discuss the nature of these guidelines and recommended that approval of clinical applications for editing of somatic cells continue under the existing regulations. The U.S. FDA has adopted this approach and regulates therapeutic somatic cell genome editing as it does other gene therapies. Human germ line editing is currently illegal in the United States (Box 13.2). Several clinical trials have been conducted to disable genes in somatic cells by CRISPR-Cas genome editing (see chapter 6) as an *ex vivo* therapy to treat cancers, human immunodeficiency virus infections, and blood disorders. The first clinical trial to replace a defective gene with a functional copy using this technology was approved in 2021. In the trial, hematopoietic stem cells were harvested from bone marrow of a small number of patients with severe sickle cell disease and transfected with the β-globin gene (*HBB*) *ex vivo*. Stem cells carrying the *HBB* gene were transplanted back into the patient, whose remaining bone marrow was destroyed with chemotherapy. The stem cells mature into different types of blood cells and are released into the bloodstream to restore hemoglobin function. An advantage of this approach is that genetically modified cells can be selected and characterized prior to transplantation, while a major obstacle is establishment of a sufficient number of edited cells following transplantation to obtain clinical benefit. It is expected that replacement of 20% of defective *HBB* genes with functional copies will improve hemoglobin function. Safety concerns associated with the transplantation of stem cells include the ability of the cells to differentiate into unintended cell types and their ability to proliferate

Human Germ Line Editing

box 13.2

In November 2018, He Jiankui, a scientist at the Southern University of Science and Technology in Shenzhen, China, announced the birth of the first babies, twin girls, from genome-edited embryos. He claimed to have disrupted the *CCR5* gene, which encodes a membrane protein that enables the human immunodeficiency virus (HIV) to enter cells, using CRISPR-Cas genome editing technology. In a small proportion of Europeans, a natural 32-bp deletion in the coding region of the *CCR5* gene generates a nonfunctional receptor which confers resistance to HIV. According to Dr. He, immediately following *in vitro* fertilization, Cas endonuclease was targeted to one end of the 32-bp region and generated a 15-bp in-frame deletion in one of the copies of the *CCR5* gene in one twin and a 1-bp insertion in one copy and a 4-bp deletion in the other copy of the gene (both resulting in reading frame shifts) in the second twin. The details of the experiment were not published in a scientific journal, but rather, the babies' births were announced first on YouTube and then presented at an international summit on human genome editing.

The announcement shocked the scientific community, which condemned the experiment as reckless and called for the immediate cessation of human germ line editing. Germ line editing refers to genomic changes to reproductive (germ) cells that develop into gametes (egg or sperm) or to early-stage embryos. In contrast to somatic cell genome engineering, genomic changes in the germ line can be passed on to future generations. Many scientists believe that the health and safety issues and social consequences of human germ line editing are not fully known, and therefore its application for reproductive purposes is premature. One important concern is the potential for mutations to occur at unintended (off-target) sites in the genome. In the case of He's twins, also unknown are the consequences of the mutations in the CCR5 protein, which normally functions as a chemokine receptor involved in immune signaling; the specific mutations generated in the edited embryos were not tested in animal models before the embryos were implanted in the mother. Will the mutations protect against HIV infection or produce proteins with novel or toxic functions, and will they increase susceptibility to other infections, such as West Nile virus and influenza, as is the case for individuals with the natural 32-bp deletion? It is likely that neither twin is protected from HIV infection, because one has a functional copy of the gene that encodes the HIV receptor and, if the genomic changes occurred after cells in the embryo had divided, the other twin may be a genetic mosaic and retain some normal copies of the *CCR5* gene. Some argue that the risks of embryo and germ line editing are unnecessary, in this case because HIV infections are preventable and controllable with drugs, and in general because other methods are available to avoid passing on genetic disorders, such as preimplantation genetic screening of embryos that are fertilized *in vitro*. That is, there is not an unmet medical need that justifies the risks. Ethical concerns range from the question of whether He adequately informed the prospective parents of the risks to the potential of germ line editing to change our species.

While some scientists and ethicists have called for a global moratorium to allow discussion of technical, medical, and ethical issues, many believe that once safety issues are resolved, genome editing of embryos should be used to prevent serious genetic diseases for which effective alternatives are not available. Indeed, some suggest that it would be unethical not to do so. Few support its application for enhancing cognitive or physical traits. Regulations for the safe and ethical implementation of human genome editing should be developed in consultation with the general public, taking into consideration diverse perspectives, and with international cooperation.

Following his announcement, Dr. He was fired from his position at the university, fined the equivalent of US$430,000, jailed for 3 years for practicing medicine illegally, and banned from working with human reproductive technology. Researchers who assisted He received lesser penalties.

inappropriately, which can lead to tumor formation. Researchers hope to eventually introduce the functional *HBB* gene *in vivo*.

Genetic and Genomic Testing

Genetic tests analyze the sequence or expression of genes for the purpose of diagnosing human disease and predicting disease susceptibility, prognosis, or response to therapy. They are currently regulated as medical devices in the United States under the Food, Drug, and Cosmetic Act and in some cases as biological products under the Public Health Services Act. Quality

standards for clinical laboratories that perform these tests have also been established. Medical devices are classified in one of three groups depending on the potential risk of harm to patients from their intended use and are subject to different levels of control depending on their classification. In this context, the definition of harm includes the potential consequences to a patient's health from the interpretation of the test results. Most molecular diagnostic assays for human diseases are considered to be of either moderate risk (class II), for example, a test that is substantially equivalent to an existing test for influenza virus infection, or high risk (class III), for example, tests that utilize new technologies or screen for tumor markers or for highly virulent pathogens. Both require premarket evaluation and either premarket notification for tests (generally class II tests) that are substantially equivalent to a predicate test or FDA approval (class III tests) before they are made available to the general public. Scientific evidence, usually including clinical studies, must be presented to show that the assay is safe and effective for its intended use. The analytical performance of a molecular diagnostic assay that uses antibodies, receptor proteins, nucleic acids, and other biological molecules is based on its specificity and sensitivity, that is, on a test's ability to accurately and consistently detect a target molecule. Assessment of risk takes into consideration the consequences of false-positive or -negative results on a patient's health. Several nucleic acid-based diagnostic tests that analyze a genetic sequence or measure levels of gene expression have been approved by the FDA for a variety of human diseases and conditions (Table 13.11).

More than two decades ago, when the first human genome sequence was published, considerable discussion focused on the next revolution in medicine: personalized genomic medicine (also referred to as precision medicine). It was anticipated that a patient's genome sequence could be used to predict susceptibility to disease, to tailor interventions to reduce the risk of developing a disease, and to prescribe effective drugs at the correct dosage that will yield fewer side effects and will be more cost-effective. There is little doubt that genome sequencing has contributed to our understanding of human diseases. Genome-wide association studies have identified more than 3,500 genes that are associated with susceptibility to human diseases; however, the predictive power of these data is still quite low. The genetic variants that have been identified explain at best 20%, and more typically less than 10%, of the risk of developing a common heritable disease. The absence of a disease-associated allele does not necessarily indicate reduced risk. On the other hand, interpretation of genome sequences may lead to false-positive conclusions, that is, the diagnosis of a disease that is not present. While some alleles confer a very high risk of a disease, such as some cancer-associated variants of the *TP53* gene encoding the tumor suppressor protein p53, other genetic variants confer only a low risk of a disease. For some diseases, the presence of an allele associated with high risk of the disease in the genome sequence of a healthy individual may have predictive value only when other characteristics are present that contribute to the development of that disease. For example, the C282Y mutation is associated with high risk for hemochromatosis, an iron metabolism disorder, only in individuals who exhibit elevated blood iron levels. Moreover, often nothing is known about the biological function of a disease-associated locus in the disease process, and therefore the information cannot be used beyond determining associations. For example, the locus may not be an effective target for treatment.

Table 13.11 Some FDA-approved nucleic acid-based tests for human genetic and infectious diseases[a,b]

Test name	Disease or condition detected	Type of test	Intended use
Vysis EGR1 FISH probe kit	Acute myeloid leukemia	Fluorescence *in situ* hybridization of probes to detect deletion of *EGR1* on chromosome 5q	To determine disease prognosis
MammaPrint	Breast cancer	Microarray-based gene expression profiling of tumors	To assess risk of metastasis
xTAG cystic fibrosis kit	Cystic fibrosis	Multiplex PCR/allele-specific primer extension to detect mutations in the cystic fibrosis transmembrane regulator gene	To identify carriers of gene variants associated with the disease and for screening newborns
Infiniti CYP2C19 assay	Drug metabolism	Microarray-based assay to detect variants of the cytochrome P450 2C19 gene associated with drug metabolism	To determine therapeutic strategies and treatment dose for drugs that are metabolized by the *CYP2C19* gene product
AlloMap molecular expression testing	Heart transplant	Quantitative real-time PCR assay to determine expression profiles (RNA levels) of several genes	To identify heart transplant recipients with a low probability of moderate to severe acute rejection
NADiA ProsVue	Prostate cancer	Immunoassay to measure prostate-specific antigen	To determine the prognosis and assess risk of recurrence of prostate cancer
Verigene *Clostridium difficile* nucleic acid test	*Clostridium difficile* infection	Multiplex PCR to detect *C. difficile* toxin genes in stool samples	To diagnose *C. difficile* infection
Cobas CT/NG test	*Chlamydia trachomatis* and *Neisseria gonorrhoeae* infection	TaqMan assay to detect specific *C. trachomatis* and *N. gonorrhoeae* DNA sequences	To diagnose chlamydial and gonococcal disease
MultiCode-RTx HSV-1 and -2 kit	HSV-1 and -2 infection	Real-time PCR to detect specific HSV-1 and -2 DNA sequences	To diagnose HSV-1 and -2 infections
JBAIDs influenza virus subtyping kit	Influenza A virus infection	Reverse transcriptase real-time PCR assay targeting the viral hemagglutinin and nucleocapsid protein genes	To detect and differentiate among seasonal influenza A/H1 and A/H3 and 2009 H1N1 influenza virus infections
BD Max MRSA assay	MRSA	TaqMan assay to detect MRSA-specific DNA sequences	To prevent and control MRSA in health care settings

[a]From http://www.fda.gov/MedicalDevices/ProductsandMedicalProcedures/InVitroDiagnostics.
[b]*EGR1*, early growth response 1 gene; HSV-1 and -2, herpes simplex viruses 1 and 2, respectively; MRSA, methicillin-resistant *Staphylococcus aureus*.

Thus, although the current cost of sequencing a human genome is low and direct-to-consumer genomic tests are accessible to many individuals, caution must be taken in interpreting the sequences. There is the potential to reveal risks for disease for which there are no treatments. For example, while identification of *APC* (adenomatous polyposis coli) alleles associated with colon cancer may lead to regular monitoring by colonoscopy, and identification of certain alleles of breast cancer genes *BRCA1* and *BRCA2* may lead to prophylactic mastectomies and ovariectomies, there are no interventions available for individuals who possess alleles associated with Huntington's disease. Patients must be counseled about the medical, emotional, and social risks and benefits of genomic testing and understand the scope and limitations of the predictive ability of genome sequence data.

Many are also concerned about the potential for discrimination against individuals who carry putative deleterious alleles. In 2008, the U.S. Genetic Information Nondiscrimination Act was passed to prevent health insurance providers and employers from requesting genetic information and subsequently discriminating against individuals based on that information. However, many countries are without similar legislation to address genetic discrimination.

Economic Issues

Many biologic drugs are costly. For example, the cost to treat a patient for rheumatoid arthritis with etanercept (Enbrel), a fusion protein that inhibits the cytokine tumor necrosis factor and is produced by cultured mammalian cells, is $72,000/year. The average cost of treatment with therapeutic monoclonal antibodies is $97,000/year, and the gene therapy Zolgensma for spinal muscular atrophy, a one-time treatment, is $2,125,000, one of the most expensive drugs. Why are the costs so high? Contributing factors are the human and technological resources and the strict and lengthy regulatory process required to bring new drugs to market. It is a risky business. According to the pharmaceutical industry, the average cost over the last 10 years to bring a new drug to market was $1.3 billion. This takes into account the total annual research and development costs of a pharmaceutical company and the investment in drugs that did not reach the market. Only about 10% of drugs under investigation enter clinical trials, and clinical trials, which typically involve thousands of patients (most in phase III), can cost tens of millions of dollars each. On average, 66% of new drugs are successful following phase I clinical trials. Phase II and III trials have success rates of 49% and 59%, respectively, and 83% are finally approved by the regulatory agency. Overall, only about 14% of new drugs that enter clinical trials are approved for commercialization. Oncology drugs have a much lower success rate (3%), while vaccines for infectious diseases have a higher success rate (33%). On average, it takes about 8 years to bring a new drug to market, although this varies depending on the type of drug. For example, new drugs developed to treat diseases of the central nervous system take a longer time to approval (about 10 years) than antivirals for treatment of AIDS (about 5 years). Many stakeholders have advocated for a reduction in the length of clinical trials; however, a major factor for their long duration is recruitment of a sufficient number of patients, especially from diverse populations. As drug patents expire (see below), the availability of biosimilar drugs that are highly similar to, and therefore interchangeable with, an FDA-approved biological drug, is expected to reduce the costs of these drugs, in one estimate by up to 40%. For approval, biosimilar drugs must have the same mode of action, route of administration, and strength and be used to treat the same disease as the original drug.

Despite the high cost of these drugs, their efficacy has reduced hospital costs. For example, the number and costs of emergency department visits and hospitalizations decreased for patients with the inflammatory bowel diseases ulcerative colitis and Crohn's disease in the year following initiation of monoclonal antibody therapy (most commonly infliximab and adalimumab). Since the introduction of biologic drugs, the number of days that cancer patients spend in the hospital has declined by 23%, and by 59% for breast cancer patients alone. The use of newer drugs has also reduced the number of workdays lost due to illness and the number of workers receiving disability insurance benefits.

Some people are concerned that the high cost and low approval rates will slow the development of new drugs, especially the so-called orphan drugs that are used to treat rare diseases or diseases found mainly in low-income populations. In the United States, an orphan drug is defined as one used to treat, diagnose, or prevent a disease that affects fewer than 200,000 people or that will not generate enough sales to recover the costs of development and marketing (Table 13.12). More than 7,000 diseases are considered to be

Table 13.12 Some orphan drugs recently approved by the FDA[a,b]

Orphan drug	Drug description	Orphan disease
Evinacumab	Monoclonal antibody targets angiopoietin-like protein 3	Familial hypercholesterolemia
Umbralisib	Small molecule inhibitor of phosphatidylinositol 3-kinase delta	Marginal zone and follicular lymphoma
Lisocabtagene maraleucel	CD19-directed CAR T-cell immunotherapy	Large B-cell lymphoma
Daratumumab	Monoclonal antibody targets CD38 glycoprotein on myeloma cells	Light chain amyloidosis
Ivacaftor	Small molecule targets cystic fibrosis transmembrane conductance regulator channel	Cystic fibrosis
Ansuvimab	Monoclonal antibody targets Ebola virus surface protein	Ebola virus infection
Rilonacept	Fusion protein blocks interleukin-1 signaling	Deficiency of interleukin-1 receptor antagonist
Anakinra	Recombinant human interleukin-1 receptor antagonist protein	Deficiency of interleukin-1 receptor antagonist
Tasimelteon	Melatonin receptor agonist	Smith-Magenis sleep disorder
Viltolarsen	Antisense oligonucleotide targets dystrophin pre-mRNA	Duchenne muscular dystrophy

[a]Source: www.accessdata.fda.gov/scripts/opdlisting/oopd/index.cfm.
[b]CAR, chimeric antigen receptor.

rare, including multiple sclerosis, cystic fibrosis, and Duchenne muscular dystrophy. Most rare diseases are genetic disorders. In the European Union, the definition of an orphan drug also includes those used to treat diseases that are primarily found in low-income countries. To encourage the development of orphan drugs, many countries facilitate the approval process and provide financial incentives. The average time for clinical testing and regulatory approval for orphan drugs is about 6.6 years, compared to 8.0 for nonorphan drugs. Financial incentives may include tax credits or special grant programs for orphan drug producers, a reduction in the number of patients required for clinical trials, or an extended period of patent protection. These incentives have resulted in a greater number of drugs to treat rare diseases. Since the enactment of the U.S. Orphan Drug Act in 1983, more than 600 orphan drugs have been developed and approved, in contrast to 38 prior to the legislation. In 2020, 58% of FDA novel drug approvals were for orphan drugs. Among these are treatments for rare cancers, such as the monoclonal antibodies naxitamab for treatment of neuroblastoma and daratumumab for treatment of multiple myeloma.

Has the high cost of drugs forced some patients to seek cheaper, illegitimate drugs and companies to manufacture them? In this context, illegitimate drugs are those made to deliberately deceive the consumer and include substandard and falsified drugs, where the quality, source, or identity of the drug is falsely represented. Many different types of drugs have been counterfeited, ranging from treatments for life-threatening diseases to analgesics (Table 13.13). The contents may contain incorrect, low concentrations of, or no active ingredients. Thus, the medicines may be ineffective, impure, or toxic. Furthermore, illegitimate antimicrobial agents often contain insufficient concentrations of antibiotics that select for the growth of drug-resistant bacteria. Low-dose antibiotics are a major contributing factor to the emergence of drug-resistant strains of *Mycobacterium tuberculosis* and *Plasmodium* spp. that cause tuberculosis and malaria, respectively.

Illegitimate drugs can be found throughout the world but are particularly problematic in many developing countries, where the marketing of pharmaceuticals is poorly regulated or the regulations are not enforced, and adherence to international quality standards is financially difficult for many small manufacturers. In developed countries such as the United States,

Table 13.13 Some examples of recently circulating illegitimate drugs[a]

Drug	Intended use	Illegitimate drug type[b]	Year and place of circulation
BNT162b2 COVID-19 vaccine	Protection against SARS-CoV-2 virus	Falsified drug	2021; Mexico
Harvoni (Ledipasvir/Sofosbuvir)	Treatment for chronic hepatitis C	Falsified drug	2020; Brazil, Turkey
Fluzone quadrivalent influenza vaccine	Protection against influenza virus	Falsified drug	2020; Mexico
Defibrotide (Defitelio)	Treatment for hepatic veno-occlusive disease	Lacks active ingredient, contaminated with mold	2020; Argentina, Australia, Latvia, Malaysia, Saudi Arabia
Uni-Gold HIV	Diagnosis of human immunodeficiency virus	Falsified diagnostic test	2020; Guyana, Kenya
Quinine sulphate	Treatment of malaria	Lacks active ingredient	2020; West and Central Africa
Verorab rabies vaccine	Prevention of rabies	Falsified drug	2019; Philippines
ICLUSIG (ponatinib)	Treatment of leukemia	Contains a different drug (paracetamol)	2019; global
Dukoral vaccine	Protection against *Vibrio cholera*	Falsified drug	2019; Bangladesh
Amoxicillin plus clavulanic acid	Treatment of bacterial infections	Falsified drug	2019; Haiti

[a]Source: https://www.who.int/medicines/publications/drugalerts/en/.
[b]Falsified drug indicates that the identity, composition, or source are deliberately or fraudulently misrepresented.

most European Union countries, Canada, Australia, New Zealand, and Japan, illegitimate drugs represent less than 1% of the market. But in many African, Asian, and Latin American countries, the problem is greater. However, unlicensed and unregulated vendors often sell illegitimate drugs over the Internet, mainly to individuals in middle- to high-income countries. In the United States, online pharmacies must be accredited by the National Association of Boards of Pharmacies and comply with standards of quality. In addition, globalization of the pharmaceutical production and distribution system contributes to the problem. Ingredients for medicines are often produced, combined, and packaged in different countries and therefore pass through many hands before reaching the patient. This provides many opportunities for substandard and falsified products to enter into the manufacturing and distribution chain. International cooperation is necessary to establish and enforce regulations for manufacturing quality and marketing to ensure a safe supply of drugs for patients. Measures can also be taken to reduce the costs of essential drugs for patients in developing countries that cannot afford them and therefore turn to the illegitimate market. This could include discounted pricing, bulk purchasing, or local production.

Patenting Biotechnology

Biotechnology products are produced by companies with the expectation that they will profit from their sale. Many of these products require substantial funds and many years to develop and receive regulatory approval, and such high-risk investments are not made without some assurance of legal protection from competition. Society also has a large stake in technological advancements that improve the quality of life, which are facilitated by disclosure of the details of new technologies. Intellectual property rights are granted by governments to protect the interests of both inventors and society at large. That is, they provide inventors with exclusive rights to the novel products or processes that they develop while encouraging innovation. For biotechnologies, the most important form of intellectual property is a patent. A patent is a legal document that gives the patent holder

exclusive rights to make, use, or sell the described invention for a period of 20 years from the date that the application is filed. Although patents are awarded to individuals, the patent rights may be assigned to a company. A patent is also a public document that contains a detailed description of the invention. However, while the patent holder may develop other products from the original invention, competitors would have to license the right to use the invention in order to develop a product based on it.

Patenting

Product and process patents are the two major categories of patents (Table 13.14). Products include homogeneous substances, complex mixtures, and various devices, while processes include preparative procedures, methodologies, or actual uses. Generally, for either a product or a process to be patentable, it must satisfy four fundamental requirements:

1. The invention must be novel. The product or process must not already exist or be described in another patent or publication, even in another country, prior to the submission of the patent application.
2. The invention must be nonobvious. A patent cannot be granted for something that was merely previously unknown, i.e., it cannot be a discovery. Rather, the invention must contain an inventive step and be sufficiently different from products or processes already in use or previously described.
3. The invention must be useful in some way, whether it is a process, an instrument, a compound, a microorganism, or a multicellular organism.
4. The invention must be adequately described in the patent application such that a person knowledgeable in the same field can implement it.

A patent cannot be granted for anything that is "a product of nature." The notion here is that it is not appropriate for society to give a monopoly to someone for something that occurs naturally, that has merely been discovered, and that therefore belongs to the public. Often companies and individuals circumvent this constraint by applying for a patent that covers the process of purification of a product, thereby avoiding the direct question of

Table 13.14 Common types of patent categories with examples of biotechnology inventions

Category	Examples
Product patents	
Substance	Recombinant proteins, monoclonal antibodies, plasmids, promoters, vectors, cDNA sequences, antigens, peptides, RNA constructs, antisense oligonucleotides, peptide nucleic acids, ribozymes, and fusion proteins
Composition of matter	Multivalent vaccines, biofertilizers, bioinsecticides, host cells, microorganisms, transformed cell lines, and transgenic organisms
Devices	Pulsed-field gel electrophoresis apparatus, DNA-sequencing units, and micropro-jectile gene transfer machine
Process patents	
Process of preparation	DNA isolation, synthesizing double-stranded DNA, vector-insert construction, PCR applications, and purification of recombinant proteins
Method of working	Nucleic acid hybridization assays, diagnostic procedures, and mutation detection systems using PCR
Use	Applying biofertilizers and bioinsecticides, fermentation of genetically modified microorganisms, and nontherapeutic animal treatment systems

ownership of either a natural substance or an organism that produces the product. According to the U.S. Supreme Court, virtually "anything under the sun that is man-made" is patentable; however, in other countries, including member states of the European Union, therapeutic and diagnostic procedures are not patentable (although therapeutic substances are patentable).

In the United States, an inventor has 1 year following public disclosure of their invention in which to apply for a patent. A patent application must include the background of the invention that describes the current "state of the art" in the field of the invention, an explanation of the nature of the invention and a description of how the invention works, including figures and schematic representations, where necessary, and a list of claims about the invention and how the invention may be used. The application is sent to the U.S. Patent and Trademark Office (USPTO), where it is reviewed by an examiner for novelty, nonobviousness, utility, feasibility, and general acceptability as a patentable invention.

If an examiner agrees that the invention meets all the criteria for patentability, then a patent is awarded. Usually, it takes 2 to 5 years following the filing of the initial application before a patent is granted. However, the receipt of a patent is not a license to produce and sell the invention. All statutory regulations must be met before any product can be marketed. For example, if a patent is granted for a genetically engineered microorganism, the manufacturer must satisfy the regulations for its production, distribution, and release. Protection of patent rights is the responsibility of the patent holder, and generally, that means bringing a lawsuit(s) against those who are presumed to be infringing on the patent. Patent disputes are decided by the courts and not by the patent office. Similarly, if a person or company feels that an awarded patent is inappropriate, the legitimacy of the patent can be challenged by a lawsuit (Box 13.3).

If a patent application is rejected by an examiner, then the applicant can appeal the decision to a Patent Appeals Board. If this appeal is turned down, then the decision can be challenged legally. There are a number of cases in which the stakes are considered to be so high that costly court battles go on for years. The American inventor Thomas Edison, who held more than a thousand patents in his lifetime, once described a patent as "an invitation to a lawsuit."

Patenting in Different Countries

The rights given by a patent extend only throughout the country in which the application was filed. Therefore, to protect an invention, a patent application must be filed separately in each country, and although the World Intellectual Property Organization is attempting to develop international standards, patent offices in different countries often reach quite different conclusions about the same patent application. For example, in 1989 the scientists working for the biotechnology company Genentech applied for a patent in the United Kingdom for, among other things, the production of human tissue plasminogen activator (tPA) by recombinant DNA processes. This protein exists in small amounts in the human body and converts plasminogen to plasmin. Plasmin is an active enzyme that degrades the fibrin of a blood clot, and human tPA is used for the prevention and treatment of coronary thrombosis. Although the procedures are fairly routine now, in the initial development of this drug, the Genentech scientists expended considerable effort to assemble a complete version of a human tPA complementary

CRISPR Patent Controversy

box 13.3

Throughout this book, we have highlighted some of the enormous number of applications for CRISPR-Cas genome editing technology, especially in crop development and the modeling and treatment of human diseases. Clearly, a patent for this technology would be extremely lucrative for the patent holder. Since the filing of the first CRISPR patent applications in 2012, the exclusive rights to this technology have been in dispute. The crux of the dispute is who was the first to demonstrate the utility of CRISPR-based genome editing. One group, led by the University of California, which includes Nobel laureates Jennifer Doudna and Emmanuelle Charpentier (University of Vienna), published research in 2012 that showed that Cas9 endonuclease, guided by a single chimeric RNA, cleaved purified DNA at a target sequence (Jinek et al., *Science* **337:**816, 2012). The University of California-led group filed a patent application with the USPTO claiming this technology in May 2012. In February 2013, another team of researchers, led by Feng Zhang of the Broad Institute of MIT and Harvard, reported that Cas9 guided by two RNA molecules generated mutations at a target site when introduced into mammalian (human embryonic kidney and mouse neuroblast) cells (Cong et al., *Science* **339:**819, 2013). A patent application filed by Zhang and the Broad Institute in December 2012 was reviewed by the USPTO under an expedited program and granted (U.S. patent no. 8,697,359) in April 2014. Subsequently, the University of California group argued that the Broad patent interferes with the claims in their patent application, which was still being considered. However, in 2017, the Patent, Trial, and Appeal Board ruled that the claims in the two patent applications concerned different inventions; while the University of California group demonstrated efficacy of the CRISPR-Cas9 system *in vitro*, they did not provide evidence of gene editing in eukaryotic cells, which the Broad group did. The decision was upheld by a federal appeals court in 2018. In 2019, the USPTO granted a patent (U.S. patent no. 10,428,352) to the University of California-led group for methods of genome editing of bacterial cells using CRISPR-Cas9; they also hold patents for CRISPR-Cas9 genome editing in all cell types in Europe, the United Kingdom, China, and several other countries.

DNA (cDNA) and clone this cDNA into *Escherichia coli* for the production of large amounts of pure tPA. In its patent application, the Genentech scientists claimed rights to human tPA as a product based on certain procedures of recombinant DNA technology that they developed, the cloning vector system, and the transformed microorganism. As a part of the "process" category, protection for the use of human tPA as a pharmaceutical agent was also sought. A total of 20 claims were presented in the original patent application. Some of these were broad and others were narrow in scope. The patent was rejected by the U.K. patent office. Genentech, on behalf of its scientists, then appealed to the U.K. Court of Appeals, which, after considerable deliberation, invalidated all of the claims for a variety of reasons. The judgment concluded that the patent was novel, but some of the judges argued that the submission was obvious; therefore, it could not be patented.

In contrast, the Genentech scientists were readily awarded a patent for human tPA in the United States. The U.S. patent not only protected the form of human tPA that was to be marketed by Genentech but also gave the Genentech scientists exclusive rights to all similar, but not identical, active forms of human tPA. Genentech won a lawsuit against two other biotechnology companies that were found to be infringing on its tPA patent, even though they were selling nonidentical forms of tPA.

The Japanese version of the tPA patent was limited to the amino acid sequence of the human tPA that was cloned and patented by the Genentech scientists. In Japan, other companies could sell variant forms of human tPA. Thus, basically the same patent application was rejected, approved and given a broad interpretation, and approved and given a narrow

interpretation by three different patent offices. This illustrates the divergent views about what is or is not a patentable invention.

Patenting Nucleic Acid Sequences

Since the development of technologies to sequence, isolate, and synthesize nucleic acids, the eligibility of these sequences for patent protection has been hotly contested. Thousands of patent applications for whole genes have been approved by patent offices throughout the world. In the United States, more than 40,000 DNA-related patents have been issued, and about 20% of human genes have been patented. Some of these human genes are used to make therapeutic proteins such as recombinant erythropoietin, which stimulates formation of red blood cells and is used to prevent anemia in some patients. Many of the patented gene sequences are used as diagnostic probes to detect disease-related genes. One example results from the discovery that particular mutations in the human genes *BRCA1* and *BRCA2* are linked to breast and ovarian cancer. Patents issued to Myriad Genetics, Inc., originally claimed the isolated *BRCA1* and *BRCA2* genes, cDNAs, and methods to detect mutations in these genes to diagnose a predisposition to breast and ovarian cancers.

The issue of patenting gene fragments was first broached in 1991 when scientists from the NIH filed for the patent rights for 315 partially sequenced human cDNAs. Two additional filings brought the total number of partial sequences to 6,869. In 1994, in a preliminary ruling, the USPTO notified the NIH that it would reject the patent application on the grounds that the functions of the sequences were not known. In other words, partial sequences by themselves did not fulfill the requirement of utility and were not patentable. However, by 1997, over 350 patent applications for more than 500,000 partial DNA sequences had been filed, mostly by private companies, which purportedly met the standard for usefulness. One of these patent proposals sought protection for about 18,500 expressed sequence tags (ESTs; short cDNAs derived from messenger RNA [mRNA] fragments). Consequently, serious concerns were raised about granting patents for large numbers of sequenced genes and partially sequenced DNA fragments with broadly based applications.

The central argument of the debate is whether isolated DNA sequences are discoveries, and therefore not patentable, or inventions, and therefore patentable. Individuals who opposed the patenting of DNA fragments of unknown or loosely defined function contend that genes and partial DNA sequences are discoveries or products of nature, and definitely not inventions. Others conceded that some of these sequences might eventually be useful but that it was premature and speculative to award patents without additional information about the functions of the sequences. On the other side, those who favored patenting ESTs maintained that these collections were novel because they defined the normal mRNA complement of various tissues and organs and consequently had utility because each collection could be used as a diagnostic assay to determine the extent to which a disease alters the normal complement of mRNAs in various organs.

After developing some ad hoc rules, the USPTO examined in more detail a full range of issues and concluded that genes and partial DNA sequences were patentable. In 2001, a set of guidelines for gene patenting was released. The key requirement for this type of application was that

each DNA sequence must have "specific, substantial and credible utility." Moreover, the written specifications and claims for each sequence must be thorough and demonstrate the actual use of each sequence and not merely a potential function. These guidelines established the criteria for patenting incomplete DNA sequences in the United States. Similarly, in the European Union, the Biotechnology Directive established that DNA sequences are patentable as long as there is an indication of function.

The patenting of DNA sequences found in humans and other organisms remains controversial. For example, the validity of the *BRCA1* and *BRCA2* diagnostic patents was challenged by the American Civil Liberties Union and the Public Patent Foundation on the basis that Myriad Genetics' exclusive rights to the genes made it impossible for women to confirm the test results using a similar diagnostic test and that the high cost of the test precluded many women from determining their risk for breast and ovarian cancers. In 2010, a U.S. district court invalidated the patent on the grounds that isolated genes are not patentable; however, a federal appeals court overturned the decision in 2011, finding that the chemical nature of isolated DNA is different from DNA in cells. For example, cDNA generated from mRNA *in vitro* lacks intron sequences found in natural DNA, and isolated DNA is cleaved from the original molecule and thus is modified. In other words, the act of isolating the sequences requires human intervention, and therefore the molecules are "engineered." Moreover, the isolated DNA sequences are used in a manner that is fundamentally different from their natural use, in much the same way that the use of purified antibiotics to treat bacterial infections is different from their natural use by soil microorganisms. The decision was appealed and, in 2013, the U.S. Supreme Court ruled that isolated nucleic acid sequences are not patentable, as they are "natural phenomena," that is, isolated DNA is not sufficiently different from genomic DNA to be eligible for patenting. On the other hand, cDNA, primers, and probes are eligible because they are synthesized in the laboratory and therefore are not naturally occurring. The decision leaves thousands of existing nucleic acid patent claims open to invalidation should they be challenged. Immediately following the Supreme Court decision, several companies began producing or using screening panels for the *BRCA* genes.

Also controversial is the increased scope of the DNA patent claims in recent years. Often included in the claims to DNA sequences and the proteins they encode are antibodies against the protein, even when the antibody has not actually been produced. Antibodies against human proteins are important from a commercial perspective because they are used for diagnostic purposes and as therapeutic agents, and therefore, there is considerable incentive to include them in patent applications. However, some argue that they do not meet the criteria for patentability because in many cases the antibodies have not been produced, and therefore their characteristics are not specifically described and working examples are not provided.

Patenting Living Organisms

Cell lines and genetically engineered unicellular organisms are considered to be patentable. For example, in 1873, Louis Pasteur received two patents (U.S. patents 135,245 and 141,072) for a process for fermenting beer that included the living organism (yeast) used in the process. However, the first time a scientist attempted to patent a genetically modified

Transgenic Nonhuman Mammals

milestone

In 1980, the U.S. Supreme Court defined a patentable invention as one that included "anything under the sun that is made by man." In 1988, a transgenic mouse was the first genetically engineered animal to be patented. In this case, the transgene consisted of a cancer-causing gene (oncogene) driven by a promoter in the long terminal repeat of the mouse mammary tumor virus (MMTV LTR). The oncogene was the *myc* gene from the chicken myelocytomatosis OK10 virus. The invention entailed cloning an MMTV LTR–*myc* fusion gene into a plasmid, injecting linearized plasmid DNA into the male pronuclei of fertilized one-celled mouse eggs, identifying offspring that expressed the *myc* gene, and establishing transgenic mouse lines. In some of these lines the *myc* gene was expressed in several different tissues, and in other lines it was limited to one or a few tissues. The integration of the MMTV LTR–*myc* gene construct, according to Leder and Stewart in U.S. patent 4,736,866, "increases the probability of the development of neoplasms (particularly malignant tumors) in the animal." These transgenic organisms can be used to test whether a compound either causes or prevents cancer and as a source of cell lines from cells of various tissues, such as the heart, that are difficult to culture. Since 1989, Du Pont has been selling one of these lines of transgenic mice under the trade name OncoMice. More generically, others prefer to call this mouse line the "Harvard oncomouse" or, for short, just "oncomouse."

The granting of the oncomouse patent was contentious, with much of the concern directed at the ethical implications of such patents. Those who oppose the patenting of transgenic animals argue that this type of patent violates the sanctity of life, threatens the integrity of species, and fosters inhumane treatment of animals. Notwithstanding these allegations, since 1988, hundreds of patents have been granted in the United States for various transgenic organisms. For example, there are now patents for transgenic animals that act as models for benign prostatic disease, inflammatory disease, altered fat tissue metabolism, and thrombocytopenia, to name a few. To date, neither the U.S. courts nor the U.S. government has suggested that, in principle, any of these patents is inappropriate. The patenting of transgenic organisms is no longer an issue in the United States. And, after much discussion and litigation, transgenic animals are now patentable in most jurisdictions throughout the world.

microorganism that was engineered by introducing different plasmids, the case was highly controversial. Each of the plasmids carried the genes for a separate hydrocarbon-degradative pathway enabling the genetically modified bacterium to break down many of the components of crude oil and therefore potentially to be used to clean up oil spills. The bacterium was developed by A. Chakrabarty, an employee of the General Electric Corporation at the time. The patent application for this bacterium was rejected by the USPTO on the grounds that microorganisms are products of nature and, as living things, are not patentable. In 1980, in a landmark decision, the U.S. Supreme Court decided that this organism was patentable according to the U.S. Patent Statute, arguing that "a live, human-made microorganism is patentable subject matter . . . as a manufacture or composition of matter."

The argument against patenting this genetically engineered microorganism tended to center on how the organism was developed. In the past, induced mutation followed by selection for novel properties was an acceptable way to create a patentable living organism. However, genetic engineering was considered by some to be "tampering with nature." Consequently, it was argued that no inventor should benefit from manipulating products of nature. This position was not upheld. In the United States from 1980 onward and later in other countries, organisms, regardless of the means that were used to develop them, must be judged by the standard criteria of novelty, nonobviousness, and utility to determine if they are patentable.

Living multicellular organisms that meet these criteria are also eligible for patenting in the United States and most other developed countries, although this continues to raise ethical and social concerns. The exclusive ownership of multicellular organisms is not a new concept. Since 1930, the U.S. Plant Patent Act has given plant breeders the right to own various plant varieties. In 1988, the first genetically engineered animal, a transgenic mouse, was patented. The transgene consisted of a cancer-causing gene (oncogene) driven by a viral promoter that increased the probability that the mouse would develop tumors. The "oncomouse" can be used to test whether a compound either causes or prevents cancer. The granting of the oncomouse patent (U.S. patent 4,736,866) was contentious, especially on moral grounds. However, from a historical perspective, it is unlikely that a position based on ethical considerations will be completely successful in preventing the patenting of all transgenic animals. For example, if an invention purports to facilitate a new treatment for human disease, the current prevalent view in most countries is that human rights and needs supersede those of animals. Since the decision on the patentability of the oncomouse, hundreds of patents have been granted in the United States for various transgenic animals, including those that act as models for human diseases or produce human therapeutic proteins.

Patenting and Fundamental Research

One of the major goals of the patent system is to encourage innovation by providing an incentive to engage in research and disclose inventions that others can improve on. However, some opponents believe that awarding a monopoly restricts competition, leads to higher prices, curtails new inventions, and favors large corporations at the expense of individual inventors or small companies. For example, one of the concerns about gene patenting is that it will impede new medical biotechnologies because researchers must obtain permission to use the genes, which can be expensive and cumbersome if multiple patents are involved. Researchers may avoid research that requires the use of patented materials or processes. Others counter that there is no evidence to support the notion that patents are serious impediments to innovation. For example, U.S. patent 4,237,224, which was granted to Stanley Cohen and Herbert Boyer in 1980 for recombinant DNA technology, for both the use of viral and plasmid vectors and the cloning of foreign genes, obviously did not seriously constrain the development of recombinant DNA technology.

Some members of the academic scientific community are concerned that patents and the consequences of patenting may be detrimental to established scientific values. Traditionally, university-based scientific research has been an open system with a free exchange of ideas and materials. The belief is that the growth of scientific knowledge and the development of technologies benefit from public availability of research articles and databases. However, some scientists have begun to feel that the integrity of traditional scientific inquiry has become secondary to financial gain from innovations. Many scientists are now advised by patent lawyers not to disclose the results of their research until a patent application is filed. Licensing fees and royalties from patents can be significant sources of income for financially constrained, nonprofit institutions such as universities. For example, during its lifetime from 1980 to 1997, the Cohen-Boyer patent for recombinant DNA technology earned about

$45 million for Stanford University and the University of California. Also, the Massachusetts Institute of Technology files more than 100 patents annually in all research fields and generates about $5.5 million per year from licensing patent rights. Most universities have established patent policies and offices that facilitate patenting and the transfer of technology, at a price, to industry. Faculty members usually receive a portion of the income from their inventions. Clearly, entrepreneurial activity is a fact of life at many universities.

summary

Significant technological advances, such as molecular biotechnology, are seldom implemented without controversy. The issues and concerns raised by the ability of scientists to genetically engineer organisms have had far-reaching implications. Guidelines for the proper conduct of recombinant DNA experiments were established by the NIH in the late 1970s and continue to be modified as the technologies advance. Regulations have also been established to ensure that the introduction of recombinant DNA products into the marketplace does not adversely affect human or animal health or the environment.

For the most part, in developing regulations, authorities have taken the view that the nature of the product and its properties, not the process by which it was manufactured, are what matter. Thus, products of biotechnology are generally evaluated on a case-by-case basis. Prior to their release into the environment, microorganisms that have been engineered to carry intergeneric DNA are evaluated by the U.S. EPA for their potential to disseminate, outcompete, or transfer DNA to indigenous microorganisms and adversely affect nontarget organisms, including humans. Microbial pesticides must also be tested for toxicity, persistence, and the potential for resistance to develop in the target organism before they are marketed.

The production of genetically modified foods has greatly increased over the last decade. More than 40% of the total global area devoted to agriculture is now planted with biotechnology-derived crops, and the first transgenic food animal was approved for commercial sale in 2015. Before they are available for consumption, the safety of new foods and food additives from microorganisms, crops, and livestock must be approved by the U.S. FDA or comparable regulatory bodies in other countries. Foods from transgenic crops and livestock undergo compositional analyses to establish that the nutrient content is substantially equivalent to that of their nontransgenic counterparts and that they do not contain toxins or allergens. For transgenic plants grown in open fields, the potential for unintended impacts on other organisms in the cultivation area, such as insects, or in adjacent fields, such as weed plants, must also be evaluated by the USDA. This is also the case for genetically engineered animals that may escape from contained facilities such as pastures or aquaculture pens. Foods containing genetic material that has been modified through recombinant DNA techniques must be labeled in the United States. In Europe, foods containing 0.9% genetically modified ingredients must be labeled. Differences in consumer acceptance and regulatory views have impacted global trade in agricultural products.

Most countries have regulations in place to ensure that new drugs, including biologics that are derived from living organisms, are safe and effective before they are commercially available. A key regulatory requirement is the series of clinical trials, in which treatments that show promise after testing in animals are tested further on different populations of humans. The long and costly regulatory process contributes to the high cost of new drugs. To encourage development of orphan drugs to treat rare diseases, many countries offer financial incentives and an expedited approval process. The high cost of some drugs has forced some patients to seek illegitimate drugs that may contain no or low concentrations of active ingredients or may contain toxic contaminants.

Commercial genetic and molecular tests to diagnose or predict susceptibility to a disease and the clinical laboratories that perform them are also regulated. Consideration is given to the potential for harm to a patient's health from interpretation of the tests, for example, false-positive (diagnosis of a disease that is not present) or false-negative (erroneous indication of low disease risk) results. Currently, the low predictive power of genetic sequences limits the clinical application of human genomic sequences for predicting disease susceptibility

(continued)

summary *(continued)*

and tailoring therapies to reduce disease risk or increase treatment efficacy.

Companies that produce biotechnology products often invest a great deal of time and financial resources to develop the products to the commercialization stage. Patents are a means to protect their investment by giving the patent holder the exclusive right to make, use, or sell the product for 20 years. At the same time, public disclosure of the invention is expected to encourage innovation; however, opponents believe that awarding a monopoly curbs new inventions, as other researchers may be concerned about infringing on a patent or must obtain permission to use patented materials or processes. For a patent to be granted, an invention must be novel, not obvious, and useful. The patenting of nucleic acid sequences is controversial. Recently, the U.S. Supreme Court ruled that DNA sequences are products of nature and therefore are not patentable, although cDNA, which is synthesized in the laboratory from RNA, is patentable. Cell lines and unicellular and multicellular organisms, including genetically engineered plants and animals, may be patented in the United States as long as they meet the standard criteria of novelty, nonobviousness, and utility.

REFERENCES

Bender E. 2018. Regulating the gene-therapy revolution. *Nature* **564:**S20–S22.

Berg P, **Baltimore D**, **Brenner S**, **Roblin RO III**, **Singer MF.** 1975. Asilomar conference on recombinant DNA molecules. *Science* **188:**991–994.

Brookes G, **Barfoot P.** 2020. Environmental impacts of genetically modified (GM) crop use 1996–2018: impacts on pesticide use and carbon emissions. *GM Crops Food* **11:**215–241.

Brunham LR, **Hayden MR.** 2012. Whole-genome sequencing: the new standard of care? *Science* **336:**1112–1113.

Chassy B, **Hlywka JJ**, **Kleter GA**, **Kok EJ**, **Kuiper HA**, **McGloughlin M**, **Munro IC**, **Phipps RH**, **Reid JE.** 2004. Nutritional and safety assessments of foods and feeds nutritionally improved through biotechnology. *Compr Rev Food Sci Food Saf* **3:**35–104.

Church G. 2015. Perspective: encourage the innovators. *Nature* **528:**S7.

Cohen SN, **Chang AC**, **Boyer HW**, **Helling RB.** 1973. Construction of biologically functional bacterial plasmids in vitro. *Proc Natl Acad Sci USA* **70:**3240–3244.

Cong L, **Ran FA**, **Cox D**, **Lin S**, **Barretto R**, **Habib N**, **Hsu PD**, **Wu X**, **Jiang W**, **Marraffini LA**, **Zhang F.** 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339:**819–823.

Cornwall W. 2020. Officials gird for a war on vaccine misinformation. *Science* **369:**14–15.

Cyranoski D. 2019. What's next for CRISPR babies? *Nature* **566:**440–442.

Deconinck K. 2020. Concentration in seed and biotech markets: extent, causes, and impacts. *Annu Rev Resour Econ* **12:**129–147.

Deer B. 2011. How the case against the MMR vaccine was fixed. *BMJ* **342**(jan05 1)**:**c5347.

Domingo JL. 2007. Toxicity studies of genetically modified plants: a review of the published literature. *Crit Rev Food Sci Nutr* **47:**721–733.

Doudna J. 2015. Perspective: embryo editing needs scrutiny. *Nature* **528:**S6.

Enserink M. 2008. Tough lessons from Golden Rice. *Science* **320:**468–471.

EFSA GMO Panel Working Group on Animal Feeding Trials. 2008. Safety and nutritional assessment of GM plants and derived food and feed: the role of animal feeding trials. *Food Chem Toxicol* **46**(Suppl 1)**:**S2–S70.

European Food Safety Authority. 2007. EFSA statement of the fate of recombinant DNA or proteins in meat milk and eggs from animals. https://www.efsa.europa.eu/en/efsajournal/pub/744.

FitzGerald MJ, **Spek EJ.** 2020. Microbiome therapeutics and patent protection. *Nat Biotechnol* **38:**806–810.

Gibbs JN. 2011. Regulating molecular diagnostic assays: developing a new regulatory structure for a new technology. *Expert Rev Mol Diagn* **11:**367–381.

Guerrini CJ, **Majumder MA**, **McGuire AL.** 2016. Persistent confusion and controversy surrounding gene patents. *Nat Biotechnol* **34:**145–147.

Hanlon P, **Sewalt V.** 2021. GEMs: genetically engineered microorganisms and the regulatory oversight of their uses in modern food production. *Crit Rev Food Sci Nutr* **61:**959–970.

Hashimoto K, **Aida T.** 2008. Antibody patenting without antibodies: a global trend. *Nat Biotechnol* **26:**1341–1343.

Hoffman NE. 2021. Revisions to USDA biotechnology regulations: the SECURE rule. *Proc Natl Acad Sci USA* **118:**e2004841118.

Holman CM. 2012. Debunking the myth that whole-genome sequencing infringes thousands of gene patents. *Nat Biotechnol* **30:**240–244.

Howard T, **Rifkin J.** 1977. *Who Should Play God? The Artificial Creation of Life and What It Means to the Human Race*. Dell Publishing Co, New York, NY.

International Service for the Acquisition of Agri-Biotech Applications (ISAAA). 2019. *Global Status of Commercialized Biotech/GM Crops: 2019*. Brief no. 55. ISAAA, Ithaca, NY.

Jain A, **Marshall J**, **Buikema A**, **Bancroft T**, **Kelly JP**, **Newschaffer CJ.** 2015. Autism occurrence by MMR vaccine status among US children with older siblings with and without autism. *JAMA* **313:**1534–1540.

Jinek M, **Chylinski K**, **Fonfara I**, **Hauer M**, **Doudna JA**, **Charpentier E.** 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337:**816–821.

Kimman TG, **Smit E**, **Klein MR.** 2008. Evidence-based biosafety: a review of the principles and effectiveness of microbiological containment measures. *Clin Microbiol Rev* **21:** 403–425.

Klümper W, **Qaim M.** 2014. A meta-analysis of the impacts of genetically modified crops. *PLoS One* **9:**e111629.

Koch MS, **Ward JM**, **Levine SL**, **Baum JA**, **Vicini JL**, **Hammond BG.** 2015. The food and environmental safety of Bt crops. *Front Plant Sci* **6:**283.

Lander ES, **Baylis F**, **Zhang F**, **Charpentier E**, **Berg P**, **Bourgain C**, **Friedrich B**, **Joung JK**, **Li J**, **Liu D**, **Naldini L**, **Nie J-B**, **Qiu R**, **Schoene-Seifert B**, **Shao F**, **Terry S**, **Wei W**, **Winnacker E-L.** 2019. Adopt a moratorium on heritable genome editing. *Nature* **567:**165–168.

Lemaux PG. 2008. Genetically engineered plants and foods: a scientist's analysis of the issues (Part I). *Annu Rev Plant Biol* **59:**771–812.

Lemaux PG. 2009. Genetically engineered plants and foods: a scientist's analysis of the issues (part II). *Annu Rev Plant Biol* **60:**511–559.

Long KC, **Alphey L**, **Annas GJ**, **Bloss CS**, **Campbell KJ**, **Champer J**, **Chen C-H**, **Choudhary A**, **Church GM**, **Collins JP**, **Cooper KL**, **Delborne JA**, **Edwards OR**, **Emerson CI**, **Esvelt K**, **Evans SW**, **Friedman RM**, **Gantz VM**, **Gould F**, **Hartley S**, **Heitman E**, **Hemingway J**, **Kanuka H**, **Kuzma J**, **Lavery JV**, **Lee Y**, **Lorenzen M**, **Lunshof JE**, **Marshall JM**, **Messer PW**, **Montell C**, **Oye KA**, **Palmer MJ**, **Papathanos PA**, **Paradkar PN**, **Piaggio AJ**, **Rasgon JL**, **Rašić G**, **Rudenko L**, **Saah JR**, **Scott MJ**, **Sutton JT**, **Vorsino AE**, **Akbari OS.** 2020. Core commitments for field trials of gene drive organisms. *Science* **370:**1417–1419.

Losey JE, **Rayor LS**, **Carter ME.** 1999. Transgenic pollen harms monarch larvae. *Nature* **399:**214.

Melnikova I. 2012. Rare diseases and orphan drugs. *Nat Rev Drug Discov* **11:**267–268.

Moreno-Fierros L, **García N**, **Gutiérrez R**, **López-Revilla R**, **Vázquez-Padrón RI.** 2000. Intranasal, rectal and intraperitoneal immunization with protoxin Cry1Ac from *Bacillus thuringiensis* induces compartmentalized serum, intestinal, vaginal and pulmonary immune responses in Balb/c mice. *Microbes Infect* **2:**885–890.

National Academies of Sciences, Engineering, and Medicine. 2019. *Second International Summit on Human Genome Editing: Continuing the Global Discussion: Proceedings of a Workshop—in Brief*. The National Academies Press, Washington, DC.

National Institutes of Health. 2016. *NIH Guidelines for Research Involving Recombinant or Synthetic Nucleic Acid Molecules (NIH Guidelines)*. Department of Health and Human Services. National Institutes of Health, Washington, DC. [Online: http://osp.od.nih.gov/sites/default/files/NIH_Guidelines.html]

Nawaz MA, **Mesnage R**, **Tsatsakis AM**, **Golokhvast KS**, **Yang SH**, **Antoniou MN**, **Chung G.** 2019. Addressing concerns over the fate of DNA derived from genetically modified food in the human body: a review. *Food Chem Toxicol* **124:**423–430.

Netherwood T, **Martín-Orúe SM**, **O'Donnell AG**, **Gockling S**, **Graham J**, **Mathers JC**, **Gilbert HJ.** 2004. Assessing the survival of transgenic plant DNA in the human gastrointestinal tract. *Nat Biotechnol* **22:**204–209.

Organization for Economic Co-operation and Development. 2015. Risk assessment considerations of genetically modified micro-organisms for releases. In *Biosafety and the Environmental Uses of Micro-Organisms: Conference Proceedings*. OECD Publishing, Paris, France.

Pew Initiative on Food and Biotechnology. 2005. U.S. vs. E.U. An examination of the trade issues surrounding genetically modified food. www.pewtrusts.org.

Rudenko L, **Matheson JC.** 2007. The US FDA and animal cloning: risk and regulatory approach. *Theriogenology* **67:**198–206.

Schmidt SM, **Belisle M**, **Frommer WB.** 2020. The evolving landscape around genome editing in agriculture: many countries have exempted or move to exempt forms of genome editing from GMO regulation of crop plants. *EMBO Rep* **21:**e50680.

Shelton AM, **Sarwer SH**, **Hossain MJ**, **Brookes G**, **Paranjape V.** 2020. Impact of Bt brinjal cultivation in the market value chain in five districts in Bangladesh. *Front Bioeng Biotechnol* **8:**498.

Sherkow JS, **Greely HT.** 2015. The history of patenting genetic material. *Annu Rev Genet* **49:**161–182.

Singh MB, **Bhalla PL.** 2008. Genetic engineering for removing food allergens from plants. *Trends Plant Sci* **13:**257–260.

Smyth SJ. 2017. Genetically modified crops, regulatory delays, and international trade. *Food Energy Secur* **6:**78–86.

Swamy BPM, **Samia M**, **Boncodin R**, **Marundan S**, **Rebong DB**, **Ordonio RL**, **Miranda RT**, **Rebong ATO**, **Alibuyog AY**, **Adeva CC**, **Reinke R**, **MacKenzie DJ.** 2019. Compositional analysis of genetically engineered GR2E "Golden Rice" in comparison to that of conventional rice. *J Agric Food Chem* **67:**7986–7994.

Turnbull C, **Lillemo M**, **Hvoslef-Eide TAK.** 2021. Global regulation of genetically modified crops amid the gene edited crop boom: a review. *Front Plant Sci* **12:**630396.

U.S. Department of Agriculture. 2018. National Bioengineered Food Disclosure Standard. *Fed Regist* **83:**65814–65817.

U.S. Environmental Protection Agency. 1997. Microbial products of biotechnology; final regulation under the Toxic Substances Control Act. *Fed Regist* **62:**17910–17958.

United States Food and Drug Administration. 2009. Guidance for industry no. 187: Regulation of genetically engineered animals containing heritable recombinant DNA constructs. January 15, 2009. (www.fda.gov/cvm, accessed 29/04/09)

Upton HF, **Cowan T.** 2015. *Genetically Engineered Salmon*. Congressional Research Service Report. Congressional Research Service, Washington, DC.

Veterinary Medicine Advisory Committee, FDA. 2010. *VMAC Briefing Packet AquAdvantage Salmon*. Center for Veterinary Medicine, Rockville, MD.

Warwick SI, **Simard M-J**, **Légère A**, **Beckie HJ**, **Braun L**, **Zhu B**, **Mason P**, **Séguin-Swartz G**, **Stewart CN Jr.** 2003. Hybridization between transgenic *Brassica napus* L. and its wild relatives: *Brassica rapa* L., *Raphanus raphanistrum* L., *Sinapis arvensis* L., and *Erucastrum gallicum* (Willd.) O.E. Schulz. *Theor Appl Genet* **107:**528–539.

Wakefield AJ, **Murch SH**, **Anthony A**, **Linnell J**, **Casson DM**, **Malik M**, **Berelowitz M**, **Dhillon AP**, **Thomson MA**, **Harvey P**, **Valentine A**, **Davies SE**, **Walker-Smith JA**, **Davies SE**, **Walker-Smith JA.** 1998. Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. *Lancet* **351:**637–641.

Wong CH, **Siah KW**, **Lo AW.** 2019. Estimation of clinical trial success rates and related parameters. *Biostatistics* **20:**273–286.

Wouters OJ, **McKee M**, **Luyten J.** 2020. Estimated research and development investment needed to bring a new medicine to market, 2009-2018. *JAMA* **323:**844–853.

Wozniak CA, **McClung G**, **Gagliardi J**, **Segal M**, **Matthews K.** 2013. Regulation of genetically engineered microorganisms under FIFRA, FFDCA and TSCA. *In* Wozniak CA, McHughen A. (ed), *Regulation of Agricultural Biotechnology: the United States and Canada*. United States Government, Washington, DC.

review questions

1. Discuss some of the properties of living microorganisms that must be evaluated before a genetically modified microorganism is released into the environment. What factors reduce the likelihood that these will have an adverse effect on the environment?
2. What criteria are used by the FDA to determine if a recombinant protein is acceptable as a food or food additive?
3. What are some of the criteria that regulatory agencies consider in assessing the safety of transgenic crops?
4. Why are plants developed through genome editing technologies currently not subject to regulation by the USDA?
5. Why is it important to evaluate transgenic livestock on a case-by-case basis?
6. What is meant by the term "substantial equivalence," and how is it used to assess food derived from genetically engineered organisms?
7. Why are animal feeding trials important in the evaluation of the safety of genetically modified foods?
8. Discuss some of the reasons why foods derived from genetically engineered organisms might be safer for consumption than those derived through selective breeding. Discuss some of the reasons why genetically modified foods might be less safe.
9. What are some of the benefits of Golden Rice? Why was commercialization delayed?
10. Explain why the risk of incorporation of a transgene in genetically modified food into the genome of the consumer or intestinal bacteria is low.
11. Present an argument for or against the labeling of genetically modified foods.
12. Discuss how agriculture has decreased biodiversity. How might genetically engineered crops contribute to this? What strategies can be employed to reduce the impact of genetically engineered crops on biodiversity?
13. Why is the risk of harm to nontarget insects lower for insect-resistant transgenic crops than for insecticide sprays?
14. Describe some ways in which genetically modified organisms can benefit the environment.
15. What are some of the environmental benefits of cultivating genetically engineered crops?
16. Briefly describe the stages of testing required for a new drug to be considered for approval by the U.S. FDA.
17. Visit the U.S. National Institutes of Health registry and database of human clinical trials at http://www.clinicaltrials.gov. Search for studies using key words such as "stem cells" and "gene therapy." Choose one trial and briefly summarize the purpose of the trial, including the treatment, procedure, or test under investigation, and its potential application.
18. What are biologics and how are they regulated?
19. What are some of the issues regarding genetic and genomic testing to determine disease susceptibility?

20. What is an orphan drug? What are some of the incentives to produce these drugs?

21. What are some of the factors that contribute to the use and availability of illegitimate drugs?

22. What is the benefit of a patent to the patent holder? What is the benefit of a patent to researchers who do not hold the patent?

23. What are the essential requirements for patenting an invention?

24. What are some arguments for and against the patenting of human genes?

25. Visit the USPTO website (http://www.uspto.gov/) and conduct a search for biotechnology patents. Use various combinations of search words, such as "transgenic" and "mouse," "DNA" and "diagnostic," and so on. Summarize the inventions in some recent patents related to molecular biotechnology.

Amino Acids of Proteins and Their Designations

Amino acid	Three-letter designation	Single-letter designation
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Index

A

Aβ (β-amyloid protein), 197–198
Abecma, 831
ABO blood group system, glycosidases, 261–263
Acetate, reducing, in fermentation, 558–561
Acetobacter, 463
Acetobacterium woodii, formate synthesis system, 536
Acetobacter suboxydans, microbial conversion, 464
Acetosyringone, structure of, 595
Acids, metabolically engineered in microbes, 463
Acinetobacter calcoaceticus, 17
Acquired immunodeficiency syndrome (AIDS). *See also* Human immunodeficiency virus (HIV)
  HIV (human immunodeficiency virus) and anti-HIV antibodies, 300–302
  HIV type 1 (HIV-1) and, 359
Acremonium chrysogenum, cephalosporin production, 554
Actin, competition between DNA and, 254
Actinorhodin
  biosynthesis of, 480
  production, 484
  structure of, 485
Acute myocardial infarction, diagnosis of, 191
Acyclovir, herpes simplex virus (HSV), 310
Adalimumab (Humira), 244
Adeno-associated virus (AAV), 362–365
  clinical trials of AAV vectors, 362, 364
  CRISPR-*Staphylococcus aureus* cas9, 364
  known strains of, 362
  promoterless gene therapy with, 379, 381
  schematic of life cycle of, 363
  schematic of surface showing proteins and symmetry, 365
  viral delivery systems, 357, 358
Adenoviruses, 437; *see also* Adeno-associated virus (AAV)
  proteins of, 360
  structure of, 360
  viral delivery systems, 357, 358
Ad Hoc Committee on Recombinant Nucleic Acids, 800
Aedes mosquito, 193
Aedes aegypti mosquitoes
  dengue fever virus-resistant, 791–792
  population of, 811
Aequorea victoria
  fluorescent proteins, 233
  rhizosecretion, 622, 623
Affibody molecules, 315–316
  in clinical testing, 317
  officially available, 317
Affinity purification systems, protein, 111–113
Agency for Toxic Substance and Disease Registration, 496
Agglutination, 185, 186
Agriculture, impact on, 7
Agrobacterium radiobacter, 803
Agrobacterium tumefaciens, 137, 681
  acetosyringone and hydroxyacetosyringone, 595
  biosynthesis of auxin and cytokinin by enzymes, 598
  increasing transformation efficiency, 601–603
  infection of plant with, 595
  plant transformation with Ti plasmid of, 595–597
  schematic of Ti plasmid, 595
  spectinomycin resistance, 643
  Ti plasmid of, 596
  Ti plasmid system of, 633
Agropine, structure of, 599
Ahmadu Bello University, 594
Airlift reactors, 562, 564–565
  double-external-loop design, 566–567
  two-stage fermentation in tandem, 566–568
Alcaligenes eutrophus (*Cupriavidus metallidurans*)
  poly(3-hydroxybutyric acid), 710, 711
  polyhydroxyalkanoates, 489–493
Alcaligenes latus, polyhydroxyalkanoate biosynthesis, 574
Alcohol(s). *See also* Starch and sugars
  commercial production of fructose and, 508–510
  increasing production, 510–517
  industrial production of, 510
  metabolically engineered in microbes, 463
  production by *Zymomonas mobilis*, 531–534
  tolerance, 515–517

Alginate lyase, 254–258
  detection of clone from clone bank, 256
  DNA construct encoding, 257
  pegylation of, 258
  processing recombinant *Flavobacterium* precursor, 256, 257
  structure of portion of, 255
Allele-specific hybridization, 200, 202, 204
Allergen(s)
  potential for introducing into food, 816–819
  reducing, through genetic engineering, 819
Allergen microarray, diagnostic, 196
Alnylam, patisiran, 347
Alzheimer's disease, 77, 194, 331
  amyloid-β (Aβ) peptides, 756–759
  amyloid plaques associated with, 756–757
  DNA construct for modeling, 758
  identification of miRNAs using hybridization, 231–232
  lentiviral vector to knock down BACE1 expression, 758
  mouse models of, 756–759
  neurofibrillary tangles and amyloid plaques for, 197
American Cancer Society, 284
American Civil Liberties Union, 842
American College of Medical Genetics, 204, 226
American Health Foundation, 693
Amino acids, 468–473, 851
  biosynthetic pathway for, 692
  L-citrulline, 471–472
  commercial applications of, 469
  commercial production of, 468–469
  L-cysteine, 470
  designations, 851
  engineering bacterium for, 469–470
  genetic code and codon usage, 99
  metabolically engineered in microbes, 463
  mutant proteins with unusual, 159–160
  plant nutritional content, 690–692
  production of protein containing nonstandard, 160
  random mutations of selected residues, 171
  stability of β-galactosidase, 104
  L-valine, 470–471
Aminocyclopropane-1-carboxylate (ACC) deaminase, 602–603
Ampicillin (AMP), 568
  resistance, 501, 502
  RNA signature of antibiotic susceptibility, 229
AmpliChip CYP450 microarray, 227, 239
α-Amylase inhibitor, plant protection against insects, 645–646
Amyloid precursor protein (APPI), 171–173
  random mutagenesis decreasing protease sensitivity, 172
Amylopectin, 508; *see also* Starch and sugars
  pathway for enzymatic hydrolysis of, 509
Amylose, 508; *see also* Starch and sugars
  pathway for enzymatic hydrolysis of, 509
Amyotrophic lateral sclerosis, 331
Anabaena variabilis, phenylalanine ammonia lyase from, 259
Ancestry determination
  DNA microarrays, 223–225
  estimating by allele frequencies, 224
Animal(s). *See* Food animals; Transgenic animals
Animal bioreactors. *See also* Bioreactors; Food animals
  donor organ production in pigs, 773–774
  human protease inhibitor production in rabbits, 770
  recombinant antithrombin production in goat milk, 768–769
  recombinant therapeutic protein production, 767–774
  therapeutic protein production in chicken eggs, 771–773
Animal species determination, DNA microarrays, 225
Animal virus, schematic of, 393
Anopheles mosquitoes, malaria, 790
Anthrax, *Bacillus anthracis*, 311
Antibacterial therapy, bacteriophages in, 357
Antibiotics, 474–486
  actinorhodin, 484, 485
  biosynthesis genes, 475–477
  detecting RNA signature of, resistance in bacteria, 228–230
  emergence of resistant bacterial strains, 353
  engineering polyketide, 477–480
  erythromycin A assembly line biosynthesis, 479
  erythromycin derivatives, 481
  gene clusters for biosynthesis of aromatic polyketides, 480
  genetic manipulation of *Streptomyces*, 475, 476
  jadomycin, 486
  lapaquistat, 483
  market, 474
  minimal polyketide synthase, 478–480
  overproducing, 484–486
  oxytetracycline, 486, 487
  penicillin, 474–475, 485
  Penicillium chrysogenum, 485
  production by *Streptomyces* spp., 485–486
  resistance, 501, 502
  Staphylococcus aureus staphyloxanthin, 481–482
  undecylprodigiosin, 477, 478
  unique, 481–484
Antibodies, 180–185; *see also* Anticancer antibodies; Recombinant antibodies
  against various diseases, 309–313
  anticancer, 302–309
  anti-HIV, 300–302
  antiobesity, 313–315
  bispecific, against hemophilia, 300
  combinatorial library of full-length, 297
  diagnosis of acute myocardial infarction, 191
  dual-variable-domain, 298, 300
  enhancing half-life, 315, 316
  industrial-scale production in plants, 707–708
  measuring neutralizing, 194, 195
  monoclonal, 182–185
  nonviral delivery, 373
  plants as reactors, 706–709
  polyclonal, 182
  schematic of target antigen, 182
  single-chain, for virus resistance, 663–664
  structure of, 180–182
  transient expression of, 615–616
  Zika virus infection, 193
Antibody fragments, 289–294
  antigen-binding single-protein chains (scFv), 289, 290
  CDRs, 292–293
  colicins, 293–294
  combinatorial libraries of, 294–296
  computer simulations of potential single-chain antibody, 290
  fusion of portion of *Pseudomonas* exotoxin A to scFv, 291–292
  human epidermal growth factor receptor 2 (HER2) and breast cancer, 292, 293
  immunoglobulin G (IgG), 289
  potential, from plants, 593
  schematic of several, 289
  schematic of use of monoclonal, 290
Anticancer antibodies, 302–309
  blockade of immune checkpoints, 304–305
  breast cancer cells and paclitaxel, 307
  derivatized monoclonal antibodies, 308–309
  human hepatocyte growth factor receptor (MET), 307, 308
  immune checkpoint, 304
  immunotherapy types, 305–306
  masking of target, 308–309
  ovarian cancer stem cells, 305–306
  receptor tyrosine kinases, 306–307
  schematic of cancer immunotherapy, 305
  severe side effects with monoclonal antibodies, 308–309
  topoisomerases, 303
Antigen delivery system
  bacteria as, 444–449
  cholera, 445
  Salmonella strains, 444–445
  Staphylococcus aureus, 446–448
  tetanus, 445–446
Antiobesity antibodies, 313–315
Antisense RNA, 327–331; *see also* Targeting mRNA and DNA sequences

chemical structures of modified nucleotides in synthesis of single-stranded, 344
clinical trials for, 329
expression vector to deliver, 327, 329
fomivirsen and cytomegalovirus infections, 330
inhibition of translation of specific mRNAs by, 328
insulin-like growth factor 1 receptor cDNA, 327, 329–330
mipomersen for cholesterol-lowering, 330
oligonucleotides, 327
retrovirus infections, 330–331
spinal muscular atrophy and, 329
Antithrombin
recombinant, in goat milk, 768–769
transgene construct for expression of, 769
transgenic animals, 811
α_1-Antitrypsin, 259–261
schematic of, 261
Antitumor agent, tumor necrosis factor alpha, 251
Aphids, targeting, 657–658
Appearance. *See* Food plant taste and appearance
Aptamers
affinities for proteins, 334
bioavailability, 334
cancer and, 334
comparison of properties to monoclonal antibodies, 332
description of, 331–332
nonviral delivery, 373–374
NOX-E36, 335
oligonucleotide delivery, 338
overview of screening procedure to isolate, 337
pegaptanib, 334–335
prostate cancer and, 335
safety in clinical trials, 335–336
schematic of chimeric RNA molecule of, and siRNA, 374
SELEX (systematic evolution of ligands by exponential enrichment) procedure, 332–333
targeting delivery of RNAi, 337
targeting vascular endothelial growth factor (VEGF), 334–335
Aquaculture, transgenic fish, 812
Arabidopsis, 4
dicamba, 673
glyphosate, 672
Arabidopsis thaliana, 470
creation and testing of transgenic strawberries, 676
CRISPR/Cas modification, 630, 631
edible vaccine for *E. coli*, 715
engineering for salt tolerance, 684
fatty acids in, 695
model organism, 649
poly(3-hydroxybutyric acid), 710–711
protection against fungal pathogens, 676
Arabinose
chemical structure, 532
engineered assimilation of, by *Z. mobilis*, 533–534
L-Arginine, L-citrulline synthesis, 471–472
Artemia salina, toxin-intein infusions, 264
Arthrobacter, 463
Arthrocnemum macrostachyum, DNA extracted from roots of, 34
Arxula adeninivorans, 134–136
Aryloxphenoxypropionates, herbicide resistance, 669
L-Ascorbic acid
commercial synthesis of, 464
Corynebacterium sp., 463–464, 465
2,5-diketo-D-gluconic acid (2,5-DKG) reductase in, 463, 465–466
2-keto-L-gluconic acid (2-KLG), 463–465
Pantoea agglomerans (*Erwinia herbicola*), 463–464, 465
reduced nicotinamide adenine dinucleotide phosphate (NADPH) and reduced nicotinamide adenine dinucleotide (NADH), 465–466
vitamin C, 463–466
Asilomar Conference Center, California, 800
Aspen trees (*Populus tremuloides*), lignin production, 718
Aspergillus, 136, 137
aflatoxin, 825
Aspergillus awamori, glucoamylase from, 511, 512
Aspergillus giganteus, 677
Aspergillus nidulans, alcohol dehydrogenase gene, 614
Aspergillus niger
chymosin, 805
glucoamylase from, 510, 514
Assays. *See* Enzyme-linked immunosorbent assays (ELISAs)
Atlantic salmon, 732, 787; *see also* Salmon
FDA approval, 812
nutritional content of genetically engineered, 814
transgenic, 787
Autism, MMR vaccine and, 830
Autographa californica multiple nuclear polyhedrosis virus (AcMNPV), 138
DNA from, 139, 140
genome, 142
replication, 141
Autoimmune diseases, diagnosis by indirect ELISA, 191–192
Auxin, biosynthesis of, 598
Aventis, 818
Avidin, transgenic corn plants expressing, 649
Azotobacter sp., polyhydroxybutyrate from, 492
Azuki bean weevil (*Callosobruchus chinensis*), α-amylase inhibitor, 645, 646

B

Bacillus sp., 17
degrading starch, 513
glyphosate, 671
Bacillus amyloliquefaciens
α-amylase from, 510, 511
insertion of target DNA into vector, 26
subtilisin gene from, 169
Bacillus anthracis
anthrax by, 311
vaccinia-based vaccine, 435
Bacillus brevis, xylose/glucose isomerase, 518
Bacillus Calmette-Guérin (BCG), tuberculosis vaccine, 442–444
Bacillus stearothermophilus
α-amylase from, 511
α-amylase and glucose isomerase genes, 705
Bacillus subtilis, 92, 170
acetate production, 559
alginate lyase, 257
L-citrulline, 472
engineered microbial enzymes, 529
insertion of target DNA into vector, 26
loop-mediated isothermal amplification (LAMP), 214
secretion into medium, 108
spores of, 219
yield of secreted recombinant protein by, 105
Bacillus thuringiensis (Bt)
approval process for Bt corn, 809
binding targets, 655
chemical insecticide with, toxin, 656
cloning and expression of, 638
cointegrate cloning vector carrying, 643
Cry protein, 817–818
crystal (Cry) proteins, 803
engineering Bt toxin gene, 656
gene stacking, 652–653, 655
increasing expression of protoxin, 642–645
insecticidal activity of, subsp. *kurstaki*, 640
insecticidal toxin, 637–642
insertion of toxin into membrane of insect gut epithelial cell, 641
integration of protoxin gene into chloroplast DNA, 644
limiting expression time of insecticidal toxin, 656
managing insect resistance by gene stacking, 655
mimicking, subsp. *israelensis* behavior, 654–655
parasporal crystal composed of Cry1 protoxin protein, 639, 640, 641
preventing development of resistant insects, 652–656
properties of insecticidal toxins, 639
protoxin and toxin genes, 642
relative expression levels of protoxin and toxin genes, 642
site on chloroplast genome encoding, 644

Bacillus thuringiensis subsp. *israelensis*, mimicking behavior, 654–655, 657
Bacteria
  as antigen delivery system, 444–449
  detecting RNA signatures of antibiotic resistance, 228–230
  factories for production, 538
  hybridization for identifying, 199–200
  nonviral delivery systems, 369–371
  protein secretion in, 106
  resistance to antibiotics, 353
  resistance to infection, 355–356
  solvent tolerance, 493–494
  vaccines directed against, 441–444
Bacteria and therapeutics, 267–279
  cancer, 279
  clinical trials using lactic acid bacteria, 270
  an HIV inhibitor, 274–276
  insulin, 276–279
  interleukin-10, 270–272
  leptin, 272–274
  proteins in *Lactococcus lactis*, 269
Bacteria as antigen delivery system
  cholera, 445
  Helicobacter pylori, 448–449
  Salmonella strains, 445–446
  Staphylococcus aureus, 446–448
  tetanus, 445–446
Bacteriophage(s)
  as adjunct to antibiotic therapy, 356
  antibacterial therapy, 357
  bacterial resistance to antibiotics, 355–356
  cystic fibrosis, 353–354
  development of modified, 355
  engineering, 352–357
  filamentous, M13, 357
  osteomyelitis, 354–355
  schematic of isolation and use of, 353
Bacteriophage λ DNA, 295, 296
Bacteriophage lysin protein, transgenic tobacco plants, 609
Bacteroidales, cancer and, 279
Bacteroides fragilis, blood group, 262
Bacteroidetes, 200
Baculovirus, 572
Baculovirus-insect cell expression systems, 137–145, 174
  construction of recombinant bacmid, 142
  expression vectors, 139–143
  insertion of heterologous gene into baculovirus genome, 139
  mammalian glycosylation and processing of proteins in insect cells, 143–145
  N-glycosylation of proteins in Golgi apparatus, 144
  production of recombinant baculovirus, 141
  production of virus-like particles using, 143
  recombinant products by, 138
Barrangou, Rodolphe, 37
Baseball Player's Union, human growth hormone, 250
BASF, 824
Batch fermentation, fed-batch versus, 569–574
Bayer CropScience, 824
B-cell chronic lymphocytic leukemia, 231
Bean (*Phaseolus vulgaris*), α-amylase inhibitor, 646
Belgian blue cattle, double muscling, 785
Bensulfuron, chemical structure of, 674
Berg, Paul, 2
Berg letter, 800
Bevacizumab (Avastin), 315
Bifidobacteria, milk quality, 781
Bifidobacterium longum, nonviral delivery, 369
Bioavailability, aptamers, 334
Biochemistry, 2, 5
Biodiversity, genetically engineered crops and, 822–824
Bioethanol, United States production, 519
Biofluorescent and bioluminescent systems
  assaying for pollutants with *P. fluorescens*, 236
  fluorescent proteins, 233–234
  luciferase, 234–235
  microbial biosensors, 235–238
  strain of yeast *Saccharomyces cerevisiae*, 237
Biolistics, Ti plasmid DNA delivery system, 603–604
Biological drugs, new, 828–832
Biologically functional bacterial plasmids *in vitro*, construction of, 6
Biological molecules. *See* Small biological molecules
Biologics
  economic issues, 835–837
  FDA-approved, 829
Biomarkers, 238; *see also* Protein biomarker detection
Biomass, 538
  direct conversion to ethanol, 530–531
Biopolymers
  copolymer poly(3-hydroxybutyrate-co–3-hydroxyvalerate), 490, 491
  poly(3-hydroxybutyrate)-based blends, 490
  polyhydroxyalkanoates (PHAs), 489–493
  xanthan gum, 487–489
Bioreactors, 294, 561–565, 723; *see also* Animal bioreactors; Fermentation
  agitation of culture medium, 563–564
  airlift, 564–565
  bubble columns, 561, 562
  configurations, 587
  fundamental classes of, 561–563
  limitation on size of, 563
  plants as, 706–711
  pneumatic, 564
  simplified configurations, 562
  stirred-tank reactors (STRs), 561, 562, 563–564
Biosensors, microbial, 235–238
Biotechnology. *See also* Medical products of biotechnology; Transgenic plants
  drugs, 243–244
  emergence of molecular, 1–3
  industry, 8
  invention examples, 838
  patenting, 838–839
  patenting in different countries, 839–841
  patenting nucleic acid sequences, 841–842
  recombinant DNA technology, 3, 6
  term, 2, 4
Biotinylated antiphosphotyrosine antibody, 77
Biparental conjugation, plasmid transfer, 26, 27
Bispecific antibody
  against HIV envelope protein, 303
  hemophilia and, 300
Blood groups
  ABO system, 261–263
  compatible and incompatible, 262
Bluebird Bio, gene therapy for β-thalassemia, 377
Blumeria graminis, 681
Bohlen, Larry, 818
Boll weevil (*Anthonomus grandis grandis*), cholesterol oxidase, 647
Bovine herpes virus-1
  schematic of DNA construct encoding, 395
  subunit and peptide vaccines, 394–395
Bovine milk, somatotropin, 805–806
Bovine spongiform encephalopathy (BSE), 775–776
Boyer, Herbert, 3, 6, 7, 8, 10, 844
Bradyrhizobium japonicum, 803
BRAF gene, 210–211
  TaqMan PCR detecting alleles, 210
Brassicaceae Family, *Arabidopsis thaliana*, 649
Brassica napus (transgenic canola), 823
Brassica rapa (field mustard), 823
Breast cancer, 194
  genes, 834
  human epidermal growth factor receptor 2 (HER2) monoclonal antibodies, 292, 293
  paclitaxel, 307
  trastuzumab for treatment, 281
Breeding methods, livestock and domesticated animals, 731
Brevibacterium spp., 463
  amino acid production, 468–470
Brinjal (eggplant), genetically modified crops, 594
Bromoxynil
  herbicide resistance, 669
  resistance, 673
Bubble columns, bioreactors, 561, 562, 564
Burkholderia sp.
  degrading xenobiotics, 496, 497

detoxification of nitroaromatic compounds, 504, 505, 506
Bypass surgery, application of antidotes, 336

C

Caldicellulosiruptor bescii, breaking down lignocellulosic biomass, 530–531
Calf alkaline phosphatase, 17
CAM (camphor-degrading) plasmid, microbial degradation, 498–499, 500
Canada, human papillomavirus (HPV) vaccine, 404
Cancer, 7
bacteria and, 279
diagnosis of ovarian, 190–191
DNA vaccines, 419, 422
microarray analysis of miRNA signatures, 231
protein p53 as tumor suppressor, 336–337
Candidatus, 62
Cardiopulmonary bypass surgery
application of antidotes, 336
flow chart for blood clots and bleeding prevention, 336
Cardiovascular disease, 194
liquid bead assay for protein biomarkers, 196
muscle contraction, 762, 763
rabbit models of, 761–763
Cartagena Protocol on Biosafety, 621
Catabolite activator protein (CAP), 95–96
Cattle. *See also* Livestock, disease-resistant; Transgenic livestock
muscle mass increases in, 782–786
normal brain protein *PrpC* in wild-type, 775, 776
prion-resistant, 775–776
protein composition of milk, 732
transgenesis creating tuberculosis-resistant, 777
tuberculosis-resistant, 776–778
Cauliflower mosaic virus, 691
binary cloning vector, 645
B. thuringiensis protoxin gene, 657
ethylene levels, 689
modified *B. thuringiensis* toxin gene, 643
oxygen content, 722
potato plant tubers, 701, 702
synthesis of poly(3-hydroxybutyric acid), 710, 711
transforming plants, 677
viral coat protein, 660
Celiac disease, gluten, 700–701
Cell biology, 5
Cellulase genes, 522–529
designer cellulosomes, 527–529
differential hybridization for isolating eukaryotic, 525
enzymatic biodegradation of cellulose, 523
eukaryotic, 525–526
isolating, 522–527
modifying, 529
prokaryotic endoglucanase genes, 523–524
recombinant *E. coli* close expressing, 524
Cellulomonas fimi, endoglucanase and exoglucanase genes, 526
Cellulose, 521–522
abundance, 521–522
E. coli expressing endoglucanase and exoglucanase genes, 524
enzymatic biodegradation of, 523
schematic of, 520
schematic of stained polyacrylamide gels, 526
Cellulose and hemicellulose
alcohol production by *Zymomonas mobilis*, 531–534
cellulase genes, 522–529
designer cellulosomes, 527–529
digestion of straw by designer cellulosome, 528
direct conversion of biomass to ethanol, 530–531
engineered microbial cellulases and xylanases, 529
eukaryotic cellulase genes, 525–526
isolating cellulase genes, 522–527
lignocellulosics, 519–522
metabolic control of cellulose hydrolysis, 527
modifying cellulase genes, 529
prokaryotic endoglucanase gene, 523–524
schematic of cellulosome protein complex, 522
utilization of, 518–534
Cellulosome(s), designer, 527–529
Cellulosome protein complex, 522
Center for Biologics Evaluation and Research, FDA, 828
Center for Drug Evaluation and Research, 828
Centers for Disease Control and Prevention, 828
Centrifugation, harvesting microbial cells, 574
Cervarix, human papillomavirus (HPV) vaccine, 403
Chain-terminating inhibitors, DNA sequencing with, 51
Chakrabarty, Ananda, 498, 843
Charpentier, Emmanuelle, 37, 840
ChemChina, 824
Chemical compounds, metabolically engineered in microbes, 463
Chemical engineering, 5
Chemical insecticide, *Bacillus thuringiensis* toxin with, 656
Chemotherapeutic agents, nanoparticles delivering, 352
Chestnut blight, *Cryphonectria parasitica* causing, 678–680
Chickens. *See also* Food animals
influenza-resistant, 779–781
mortality of nontransgenic and transgenic decoy, 780
production of therapeutic proteins in eggs, 771–773
Chikungunya virus, 811
Chimeric antibody, 284, 285
Chinese hamster ovary (CHO) cells, 145, 253
HSV-1 vaccines, 394, 395
human monoclonal antibody in, 623
SARS virus, 398
stirred tank bioreactors, 706
zinc finger nucleases, 348
Chinook salmon, 732, 787; *see also* Salmon
growth hormone, 812
Chlamydia
diagnostic monoclonal antibodies, 185
oxytetracycline, 486
Chlamydomonas reinhardtii, photosynthetic electron transport chain, 537
Chlorinated hydrocarbons, insect pests and, 639
Chloroplast engineering
chloroplast DNA, 606–607
genes as selectable markers and reporters, 607
plant cells with chloroplasts and double-stranded circular DNA, 605
plants, 604–608
very-high-level protein expression, 607–608
Chlorosulfuron
chemical structure of, 674
herbicide resistance, 674
Cholera, 437
bacteria as antigen delivery system, 445
edible vaccines, 712–714
engineered attenuated vaccines, 426–428
schematic of toxin protein, 396
strain carrying defective A_1 peptide, 426–428
strategy for deleting part of toxin A_1 peptide, 427
subunit and peptide vaccines, 396
toxin of, 426
VAXCHORA vaccine, 428
Cholesterol
biosynthesis in humans, 482
fatty acids and bile acids in place of, 368
lipid for nonviral delivery, 367
mipomersen for lowering, 330
Cholesterol oxidase, plant protection against insects, 646–647
Chromatin structure, recombinant protein expression and, 154, 155
Chromosomal DNA
one-step homologous recombination, 116–117
removal of selectable markers by site-specific recombination, 118–119
two-step homologous recombination, 117, 118

Chromosomal integration, mammalian cell genome, 152–155
Chymosin, 804–805
Ciprofloxacin (CIP), RNA signature of antibiotic susceptibility, 229
L-Citrulline
  global market, 471
  modifying biosynthesis of, 472
  precursor of L-arginine, 471–472
  whey and, 495
Clinical testing, affibody molecules, 317
Clinical trials, 247
  adeno-associated virus (AAV) vectors, 364
  cancer vaccines, 419
  DNAzymes testing in mice, 341
  human gene therapy, 325–326
  safety of aptamers in, 335–336
  small interfering RNA (siRNA) drugs, 346
Cloning
  electroporation procedure, 23
  eukaryotic genes, 26, 28–30
  PCR products, 42, 43, 44
  preparation of DNA for, 13–18
  recombinational, 30–32
  selecting host cells transformed with pUC19, 24, 25
  transformation and selection of cloned DNA in bacterial host, 21, 23–24, 26
Clostridium tetani, gene expression, 100
Clostridium tetani toxin, 446
Clostridium thermocellum
  biomass breakdown, 530
  designer cellulosome, 528
  engineered microbial enzymes, 529
  rhizosecretion, 622, 623
Clustered regularly interspaced short palindromic repeats (CRISPR), 36, 744; *see also* CRISPR technology
Codex Alimentarius Commission, 811
Codon usage, optimizing, 98–100
Cofactor requirements, modifying, 169–171
Cohen, Stanley, 3, 6, 10, 844
Colicins, 293
Combinatorial libraries
  antibody fragments, 294–296
  construction of single-chain antibodies, 299
  DNA constructs of Fv, 295
  full-length antibodies, 297
  immunological screening of bacteriophage M13, 296
Commercialization, molecular biotechnology, 6–8
Commercially available systems
  affibody molecules, 317
  tRNAs in *E. coli*, 100
Commodities, large-scale production of, 2
Community, Health and Environment Program, Friends of the Earth, 818
Complementary DNAs (cDNAs), 243
  synthesis of double-stranded, 28–30
Contagion, 1
Continuous fermentation, 550–551
Corino de Andrade's disease, patisiran (siRNA drug), 347
Coronavirus
  COVID-19, 413
  MERS (Middle East respiratory syndrome), 413
  outbreaks, 413
Corteva Agriscience, 824
Corynebacterium sp.
  amino acid production, 468–470
  L-ascorbic acid, 463–464, 465
  mutagenesis of gene, 466
Corynebacterium diphtheriae
  inoculating animals with, 450
  treating diphtheria in humans, 280
Corynebacterium glutamicum, 92
  amino acids, 470
  L-citrulline, 472
  genome of strain of, 473
  modified, 472–473
  L-valine, 471
Court of Justice of the European Union, 618
COVID-19 pandemic, 5, 389, 413
  development of RNA vaccines, 586
  mRNA vaccine, 391
  vaccines for, 391, 399–401, 830
Cowpea (*Vigna unguiculata*), 645
  trypsin inhibitor gene, 645
Cowpea weevil (*Callosobruchus maculatus*), α-amylase inhibitor, 645, 646
Cowpox, 387
Cows. *See also* Cattle; Livestock, disease-resistant; Transgenic livestock
  somatotropin in milk, 805–806
CpG dinucleotides, genetic immunization, 417
Creutzfeldt-Jakob disease, 248
CRISPR/Cas9 system
  disruption of ovine *MSTN* gene, 786
  Duchenne muscular dystrophy, 759–760, 761
  fruits and flowers, 690
  fungal and bacterium resistance, 681
  geminiviruses, 665–667
  genomic editing, 744–747, 831
  manipulation of genes in plants, 617–621
  microprojectile bombardment, 619
  modified genomes of plants using, 618
  plants by genome editing, 808
  resisting browning, 702
  reversal drives, 792–793
  schematic of plants with modified genomes, 618
  transgenic animals, 795
  transgenic livestock, 745–747
  virus resistance, 665–667
  zebrafish melanoma model, 765
CRISPR technology
  bacterial CRISPR-Cas system, 36
  CRISPR-Cas-based diagnostic assays, 218–219
  CRISPR-Cas-based lateral flow test strip, 219
  CRISPR-Cas system as diagnostic tool, 218
  directed evolution, 625–626
  discovery of CRISPR-Cas system, 37
  DNA sequence repair, 350–351
  gene repair in cells by CRISPR-Cas system, 349–351
  gene therapy, 376
  genome engineering using, 36–39
  germline correction of mutation using, 366
  patent controversy, 840
  plants by genome editing, 808
Crohn's disease, 268, 835
  interleukin–10, 270–272
Crop plants, 5
  fungal and bacterium resistance, 674–675
  genetically modified, 807–810
  herbicide resistance, 668–669
Cryphonectria parasitica, chestnut blight by, 678–680
Cucumber mosaic virus (CuMV), viral coat protein, 660
Cucumber necrosis virus, inhibition by single-chain antibody, 664
Cucumber vein yellowing virus, CRISPR/Cas9, 667
Cupriavidus metallidurans, polyhydroxyalkanoates, 489–493
Cyanamide, herbicide resistance, 669
Cyanobacteria, lipids from, 534–535
Cyanovirus N, HIV inhibitor, 274–276
Cyclic AMP [cAMP] repressor protein (CRP), 95–96
Cyclohexanediones, herbicide resistance, 669
L-Cysteine, biosynthesis and overproduction, 470
Cystic fibrosis, 7, 202
  allele-specific hybridization screening for, 203
  bacteriophage studies, 353–354
  enzyme DNase I, 253–254
Cystic fibrosis transmembrane conductance regulator (CFTR) gene, 202, 204
  CFTR mutations, 203
Cytochrome P450 oxidases, detection of multiple alleles, 220
Cytokines
  rheumatoid arthritis, 310–311
  targets for monoclonal antibodies, 185
Cytokinin, biosynthesis of, 598
Cytomegalovirus
  fomivirsen and, 330
  ribozyme for, 340
Cytosine, conversion to uracil, 222

D

Dalapon, herbicide resistance, 669
DdeI modification enzyme
  Desulfovibrio desulfuricans DNA, 461

scheme for cloning, 462
DDT (dichlorodiphenyltrichloroethane), 639
Degradative pathways
genetic engineering of, 497–507
manipulation by gene alteration, 499–507
manipulation by transfer of plasmids, 498–499
Deletion mutation, 158
Delivery systems. *See also* Nonviral delivery systems; Viral delivery systems
bacteria as antigen delivery system, 444–449
DNA vaccines, 414–422
nonviral, 365–376
viral, 357–365
Dementia, L-citrulline, 471
Dendrimers
nonviral delivery, 372
structures of commercially available, 372
Dengue fever
vector vaccines, 438, 439
virus-resistant mosquitoes, 791–792
Dengue virus, 422, 811
Dengvaxia, 438
Dental caries, DNA vaccines, 423–424
Desulfovibrio desulfuricans, restriction enzyme DdeI, 461, 462
d'Hérelle, Félix, 352
Diabetes
L-citrulline, 471
insulin, 276
Diabrotica virgifera virgifera LeConte, 649
Diagnostics, 5
Dicamba, herbicide resistance, 672–673
Dideoxynucleotide sequencing, 50–52, 53
Diphtheria, 387, 389
annual cases in Canada, 388
Directed mutagenesis, antibody half-life, 315, 316
Direct injection, nonviral delivery, 365–367
Discosoma Coral, fluorescent proteins, 233
Disease(s). *See* Livestock, disease-resistant; RNA biomarkers of disease
Disulfide bonds, protein stability, 102–103
DNA (deoxyribonucleic acid). *See also* Microarrays, DNA; Targeting mRNA and DNA sequences
amplification by PCR, 39–42
chloroplast DNA, 606–607
chromosomal integration by one-step homologous recombination, 116–117
cleavage by restriction endonuclease, 14
DNA shuffling, 162–163, 164
forensic analysis of microsatellite DNA, 216–218
genomic libraries, 32–36
insertion of target, into plasmid vector, 18–21
integrating into host chromosome, 114–119
large-scale production of plasmid DNA, 583–586
microarrays, 68–70
preparation for cloning, 13–18
recombinant technology, 3, 6
removal of selectable markers by site-specific recombination, 118–119
removing marker genes from chloroplast DNA, 632
removing marker genes from nuclear DNA, 627–630
Saccharomyces cerevisiae expression in, 127
sequence repair by CRISPR-Cas9 system, 350–351
sequencing technologies, 48–56
sequencing with chain-terminating inhibitors, 51
specific enzymatic amplification, 40
DNA-based diagnostics, 238–239; *see also* Polymerase chain reaction (PCR) detection methods
CRISPR-Cas-based diagnostic assays, 218–219
DNA microarrays, 219–225
hybridization probes, 199–207
PCR-based detection methods, 207–218
whole-genome sequencing assessing genetic disease risk, 225–226
DNA Data Bank of Japan, 63
DNA library, randomized 36-amino-acid peptides, 252
DNA microarrays, transcriptomics, 68–70
DNA microparticles, genetic immunization, 416–417
DNA vaccines
advantages over conventional vaccines, 415
cancer, 419, 422
CpG dinucleotides, 417
delivery, 414–422
dental caries, 423–424
DNA microparticles, 416–417
electroporation, 418–419
genetic immunization, 414–424
influenza, 420
methods for introducing, 418
MIDGE (minimalistic immunogenically defined gene expression) vectors, 421–422
mucosal immunity, 417–418
Shigella, 419–420
Zika virus, 422–423
DNAzymes
kinetic parameters of, 340–341
preclinical trials in mice, 341
ribozymes and, 338–341
schematic of cleaved c-*jun* mRNA, 342
Dolly, cloned sheep, 743–744
Donor organs, production in pigs, 773–774
L-DOPA (dopamine), Parkinson's disease, 279, 280
Double muscling, 783; *see also* Muscle mass
Doudna, Jennifer, 37, 840
Dow Chemical, 824
Downstream processing
inclusion bodies, 579–581
magnetic separation of proteins, 582–583
protein production, 578–583
schematic of large-scale protein purification using ultrafiltration membranes, 580
schematic of large-scale ultrafiltration dialysis of protein solution, 579
schematic of recovery of overexpressed target proteins, 581
utilizing immobilized enzyme, 582
Droplet digital PCR, quantifying low-abundance sequences, 213–214
Drought. *See also* Salt and drought stress
delaying senescence, 685–686
Drug monitoring, targets for monoclonal antibodies, 185
Drugs, 5
Duchenne muscular dystrophy
genetic defect in, 366
germline correction of mutation, 366
mouse model of, 759–760, 761
DuPont, 824, 843

E

Ebola virus
development of vaccine, 440–441
flow chart depicting disease, 440
schematic of vesicular stomatitis virus (VSV)-based vaccine against, 441
vector vaccines, 439–441
Economic issues, biologic drugs, 835–837
Edema factor (EF), 311
Edible vaccines, 711–716
cholera, 712–714
construct expressing cholera toxin B subunit in rice, 714
Escherichia coli, 714–716
generating immune response, 712
plant-based oral and oral booster, 713
Shiga toxin A and B subunits, 715
Eisenia andreri, tissue regeneration, 68
Electroporation
DNA vaccines, 418–419
procedure, 23
ELISAs. *See* Enzyme-linked immunosorbent assays (ELISAs)
Elizabethkingia meningoseptica, blood group, 262
Endoplasmic reticulum (ER), protein folding summary, 131
ENGERIX-B, commercial recombinant subunit vaccine, 421
Engineered attenuated vaccines, 424–430; *see also* Vaccines
cholera, 426–428
genotypes of, 429
herpes simplex virus (HSV), 425–426
Leishmania species, 430
Salmonella species, 428–429

Engineered embryonic stem cell method
  determination of transgene integration at targeted site, 742
  establishing transgenic mice with, 739, 740
  gene disruption by targeted homologous recombination, 743
  homologous recombination result, 741
  nonspecific integration result, 741
  positive-negative selection procedure for chromosomal integration, 739, 741
  transgenic animals, 737, 739–743
Engineering. *See* Protein engineering
Engineering mammalian cell hosts, enhanced productivity, 150–152
Enterococcus faecalis, dopamine and Parkinson's, 279, 280
Enterotoxigenic *Escherichia coli*, hybridization to identify, 200
Environmental concerns, genetically modified microorganisms, 802
Environmental contaminants, phytoremediation, 686–688
Environmental Protection Agency (EPA), 496, 803
Enzyme(s). *See also* DNAzymes; Ribozymes
  alginate lyase, 254–258
  α_1-antitrypsin, 259–261
  DNase I, 253–254
  glycosidases, 261–263
  masking nonhuman epitopes, 263–264
  phenylalanine ammonia lyase, 258–259
  replacement therapy, 267
  targeting mitochondria, 265–267
  toxin-intein fusions, 264–265
Enzyme-linked immunosorbent assays (ELISAs), 186–194, 238
  acute myocardial infarction diagnosis, 191
  conversion of colorless substrates to colored products, 188
  detecting infectious diseases by indirect, 192–194
  detecting pregnancy by sandwich, 188–190
  detection of viral infection, 193
  diagnosing autoimmune diseases by indirect ELISA, 191–192
  indirect ELISA, 187
  measuring disease-associated proteins by sandwich, 190–191
  measuring neutralizing antibodies, 194, 195
  ovarian cancer diagnosis, 190–191
  sandwich ELISA, 187
Epigenetic markers, detection of, 221–223
Epigenetic modifications, detection of markers, 221–223
Episomal plasmid vectors, *Saccharomyces cerevisiae* expression, 124–126
Ereky, Karl, 2, 4
Erwinia herbicola, 463
Erwinia uredovora, 815
Erythromycin A. *See also* Antibiotics
  assembly line biosynthesis, 479
  derivatives, 481
  synthesis of, 480, 481
Escherichia coli, 3, 91, 185
  antibacterial therapy, 357
  antibody production in, 561
  bacteriophage P1, 747–748
  bacteriophage T7, 94
  cell wall and membrane, 313
  chymosin, 805
  cloning and expression of *B. thuringiensis*, 638
  cloning experiment, 21, 23–24, 26
  edible vaccine for, 714–716
  engineered host strain, 104
  engineering L-valine production, 470–471
  expressing endoglucanase and exoglucanase genes, 524
  expressing organophosphorus hydrolase, 503, 504
  fermentation process, 551
  formate hydrogen lyase, 535–536
  genetic code and codon usage in humans and, 99
  high-density cell cultures, 552–553
  infection by filamentous bacteriophage, 296
  infections of, 312–313
  insertion of target DNA into vector, 26
  integration and excision of bacteriophage γ DNA, 30
  isolating and expressing foreign genes in, 460
  lac promoter, 94–96
  leptin production in, 273
  lexA3 gene, 356
  lycopene production, 473–474
  mixture of pUC19-target DNA ligation products, 25
  mutated genes improving plasmid DNA production, 584–585
  nonviral delivery, 369, 370, 371
  plasmid cloning vector pUC19, 20, 21, 22
  plasmids, 6, 18, 20
  polyalkanoates, 492
  promoter resembling consensus sequence, 92
  protein production by recombinant, 569
  protein synthesis in oxygen-limited, 546
  quiescent cells for fermentation, 555–556, 557, 558
  resistant strains, 229
  restriction endonuclease from, 14, 15
  spectinomycin resistance, 643
  synthesis of, of polypeptide with leukocyte interferon activity, 245
  T7 expression system, 96
  tac promoter, 94
  Ti plasmid and, 599
  tissue plasminogen activator (tPA), 840
  two-stage airlift reactors for NM989, 566–568
  xanthan gum, 488, 489
  xylose/glucose isomerase, 518
  yield of secreted recombinant protein by, 105
Etanercept (Enbrel), 835
Eteplirsen, 329
Ethanol. *See also* Alcohol(s)
  direct conversion of biomass to, 530–531
Ethylene, plant growth regulator, 688–690
Eukaryotic cells
  baculovirus-insect cell expression systems, 137–145
  heterologous protein production in, 119–155
  mammalian cell expression systems, 145–155
  N-linked glycosylations, 122
  posttranslational modification of proteins, 120–123
  yeast expression systems, 123–137
Eukaryotic expression systems, general features of, 123
Eukaryotic genes
  cloning, 26, 28–30
  purified mRNA from, 28
European Commission, 377, 808
European Medicines Agency, 167, 827
European Nucleotide Archive, 62–63
European Union, 806, 827
  monoclonal antibodies for human use, 282–283
  recombinant proteins for human use, 244
Expression profiling, proteins, 74–77
Expression systems, baculovirus-insect cell, 137–145
Expression vectors
  commonly used pET series, 98
  heterodimeric and -tetrameric products in mammalian cells, 149
  selectable markers for mammalian, 148, 150
Extracellular protein secretion systems, 109–111

F

Fab molecule, structure of, 168
Fabry disease, 267
Faecalibacterium spp., cancer and, 279
Faecalibacterium prausnitzii, 200
Familial amyloid polyneuropathy, patisiran (siRNA drug), 347
Farmers, production of genetically modified foods, 824–825
Fatigue, L-citrulline, 471
Fatty acids, omega-3 and omega-6, 694–695
FDA. *See* U.S. Food and Drug Administration
Fed-batch fermentation, batch versus, 569–574

Federal Food, Drug, and Cosmetic Act, 828, 832
Fermentation. *See also* Bioreactors; Microbial growth
batch, 548–549
contamination of, 563
continuous, 550–551
fed-batch, 549–550
harvesting microbial cells, 574–575
high-density cell cultures, 552–555
improving antibody production in *E. coli*, 561
increasing plasmid stability, 555
large-scale, 546, 547
maximizing efficiency of process, 551–561
protein secretion, 558, 559
quiescent *E. coli* cells, 555–556, 557, 558
reducing acetate, 558–561
two-stage, in tandem airlift reactors, 566–568
Fermentation systems, large-scale, 565–574, 587
batch versus fed-batch, 569–574
flow chart for vaccine development against porcine parvovirus, 573
two-stage, in single stirred-tank reactor (STR), 568–569
two-stage, in tandem airlift reactors, 566–568
Ferritin, iron storage, 699, 700
Fibrobacter succinogenes, engineered microbial enzymes, 529
Figwort mosaic virus (plant virus), cholesterol oxidase, 647
Filamentous fungi, expression in, 136–137
Filtration, harvesting microbial cells, 574–575
Firmicutes, 200
Fisheries. *See also* Salmon
enhancing growth of salmon, 786–787
Flavin mononucleotide (FMN), 483
Flavivirus, Zika virus, 422
Flavobacterium sp.
alginate lyase gene from, 256
organophosphorus hydrolase, 504
pesticide degradation, 503
Flavobacterium okeanokoites, zinc finger nucleases, 348, 349
Flavonifractor plautii, blood group, 263
Flavr Savr tomato, 688
Floral dip, 649
Flowers. *See* Fruits and flowers
Fluorescent dye, quantitative PCR, 44
Fluorescent proteins, 233–234
colored monomeric, from monomeric red, 234
construction of modified monomeric red, 233
Fluorophores, molecular-beacon probes with, 207, 208
Folate, plant nutrition, 697–698
Folding chaperones, 101
Fomivirsen, 329
cytomegalovirus infections, 330
Food, Drug, and Cosmetic Act, 828, 832
Food and Agriculture Organization, 594, 811
Food animals. *See also* Animal(s)
disease-resistant livestock, 774–781
enhancing production traits of, 774–787
influenza-resistant chickens, 779–781
mastitis-resistant cows, 778–779
milk quality improvements, 781–782
muscle mass increases in cattle, 782–786
prion-resistant cattle, 775–776
salmon growth enhancement, 786–787
tuberculosis-resistant cattle, 776–778
Food plant taste and appearance
preventing discoloration, 701–702
starch, 703–705
Food samples, quantitative real-time PCR, 45
Foot-and-mouth disease virus (FMDV)
picornavirus, 405
vaccine for, 404–405
Formate hydrogen lyase, *Escherichia coli*, 535–536
Francisella novicida, CRISPR-Cas9 and, 350
Frenolicin, biosynthesis of, 480
Friends of the Earth, 818
Fructose. *See also* Starch and sugars
commercial production of alcohol and, 508–510
improving production, 517–518
improving production of, 517–518
industrial production of, 510
xylose/glucose isomerase, 517–518
Fruits and flowers
centrality of ACC in synthesis of ethylene, 688–690
CRISPR mutants, 690
Flavr Savr tomato, 688
lowering ethylene levels, 688–690
Full-length antibodies, combinatorial library of, 297
Fungal expression systems, 131–137
expression in filamentous fungi, 136–137
"humanized" *Pichia pastoris*, 132–134
Pichia pastoris expression vectors, 134, 135
yeasts, 134–136
Fungus and bacterium resistance, 674–681
chestnut blight, 678–680
fusarium disease, 677–678
RNAi and CRISPR/Cas, 681
transgenic plants, 675–681
wine disease, 680–681
Fusarium diseases, resistance to, 677–678
Fusarium graminearum, 681
chickens immunized with, 678
Fusarium oxysporum, resistance to, 678
Fusion proteins, 101–102
affinity tag, 113
immobilized enzyme for, 582
immunoaffinity chromatography, 113
proteolytic cleavage of, 114
schematic for construction of, 582

G

Galacturonosyltransferase 4 (GAUT4), 721, 722
Gammaretroviruses
schematic of, 358
viral delivery systems, 357, 358
Ganciclovir, prodrug activation therapy, 378, 380
Gardasil, human papillomavirus (HPV) vaccine, 403, 404
Gastroenteritis, 200
Gaucher disease, 267
Geminiviruses, CRISPR/Cas9, 665–667
GenBank, 62
Gene(s), 243; *see also* Gene therapy
assembling oligonucleotides into, 46, 48
assembling PCR products into, 46–47, 49
assembly, 86–87
augmenting for treating genetic diseases, 377
chemical synthesis of, 46–47
cloning, 6
conditional modification with Cre-*loxP* recombination system, 747–749, 750, 751
Gibson assembly, 46, 49
Gibson cloning, 50
repair in cells by CRISPR-Cas system, 349–351
silencing, and interfering RNA, 341–343
suppression for treating genetic diseases, 377
use as therapeutic agents, 325–327
Gene drives. *See also* Transgenic animals
dengue fever virus-resistant mosquitoes, 791–792
eradicating vector-transmitted diseases, 787–794
genetic modification with, 788
inheritance of, 789
malaria vector population suppression, 789–791
reversal drives, 792–793
self-eliminating, 794
split, 793
Gene expression. *See also* Manipulation of gene expression in plants
commercially available systems, 100
increases from codon usage, 100
polycistronic, 626, 627
regulation of target, 96, 97
transient, 611–616
virus resistance, 663–668
Gene knockdown, by RNA interference (RNAi), 754–756
Genentech, 6, 8, 839–840
General Electric Corporation, 843
Gene stacking
B. thuringiensis resistance, 652–653, 655
transgenic plants, 624–625

Gene therapy, 7, 376–381; *see also* Gene(s)
approval for β-thalassemia, 377
clinical trials for human, 325–326
gene augmentation, 377
gene suppression, 377
genome editing, 377
large-scale production of plasmid DNA, 583–586
Luxturna, 829
mitochondrial diseases, 378, 379
NIH Guidelines for human, 801
overview of approaches for genetic diseases, 377
prodrug activation therapy, 378–379, 380
promoterless gene targeting, 379, 381
schematic of *in vivo*, 326
somatic cell, 326
Genetic(s), 2, 5
information, 10
regulation and safety of testing, 832–834
Genetically engineered antibodies, 284, 285, 286
Genetically engineered livestock, 810–812
Genetically engineered microorganisms. *See* Genetically modified microorganisms
Genetically modified crops, 807–810
debate on, 594
environmental benefits of, 825–826
Genetically modified foods
alteration of nutritional content, 812–815
animal feeding studies, 817
benefits from production of, 824–825
controversy about labeling, 820–822
crops, 807–810, 825–826
Golden Rice, 814–815
Golden Rice 2, 815, 816
impact of genetically engineered crops on biodiversity, 822–824
impact on trade, 827
ingredients by genetically engineered microorganisms, 804–807
livestock, 810–812
nutritional content of genetically engineered Atlantic salmon, 814
nutritional content of grain from Golden Rice 2 and rice, 816
potential for toxin or allergen introduction into food, 816–819
potential for transferring transgenes to humans or intestinal microorganisms, 819–820
societal concerns about, 812–827
transgenic corn grain, 813
Genetically modified microorganisms
environmental concerns, 802
food ingredients by, 804–807
regulations, 803–804
Genetic code and codon usage, *E. coli* and humans, 99
Genetic diseases
approaches for gene therapy, 377
nucleic acid-based tests, 834
whole-genome sequencing assessing risk of, 225–226
Genetic engineering, 3
biodegradative pathways, 497–507
crop plants producing functional insecticides, 641–642
manipulation by gene alteration, 499–507
manipulation by transfer of plasmids, 498–499
Genetic immunization
advantages over conventional vaccines, 415
cancer, 419, 422
CpG dinucleotides, 417
delivery, 414–422
dental caries, 423–424
DNA microparticles, 416–417
DNA vaccines, 414–424
electroporation, 418–419
influenza, 420
MIDGE (minimalistic immunogenically defined gene expression) vectors, 421–422
mucosal immunity, 417–418
Shigella, 419–420
survival of DNA-immunized mice, 416
Zika virus, 422–423
Genome(s)
CRISPR technology for engineering, 36–39
Genome 10K Project, 65
high-throughput next-generation sequencing, 59–60
preparation of genomic DNA sequencing libraries, 57–58
sequence assembly, 60–61
sequencing whole, 56–62
synthetic, 47
Genome editing
CRISPR-Cas system, 744–747
discovery of CRISPR-Cas system for, 37, 38
treating genetic diseases, 377
tuberculosis-resistant cattle, 777
Genomic(s)
proteomics, 70–84
transcriptomics, 65–70
Genomic libraries, 32–36
construction of, 33
isolation of genes from DNA library of halophilic bacteria, 34
preparation of sequencing, 57–58
screening of DNA library, 35
Genomics England, 63
Genomic testing, regulation and safety, 832–834
Gentamicin (GENT), RNA signature of antibiotic susceptibility, 229
Geobacillus stearothermophilus
engineered microbial enzymes, 529
loop-mediated isothermal amplification (LAMP), 214
German measles, 389
Germ line editing, human, 832
Ghrelin
antiobesity antibodies, 314–315
structure of, 314
Gibson, Daniel, 46
Gibson assembly, 46, 49
Gibson cloning, 50
Gilbert, Walter, 51
Global Alliance for Vaccines and Immunization, 388
Gluconobacter, 463
β-D-Glucuronidase (GUS)
activity in plant, 630, 631
transformed plant tissues, 611
Glufosinate
chemical structure of, 674
herbicide resistance, 669
Gluten, plant nutritional content, 700–701
GlycoDelete, 623
Glycoproteins, differential processing of, 133
Glycosidases, ABO blood group system, 261–263
Glycosylation, 121
Glyphosate
herbicide resistance, 669–672
tolerant transgenic corn, 813
Goats, recombinant antithrombin in milk, 768–769
Golden Rice, 4, 5, 814–815
vitamin A, 696–697
Golden Rice 2, 815, 816
Gonorrhea, 200
Gonsalves, Dennis, 663
Gossypol, inactivating, 650–651
Gram-negative bacteria
antibiotics, 474
cell walls and membranes, 576, 577
protein secretion in, 106
restriction endonucleases, 460
Gram-positive bacteria
antibiotics, 474
cell walls and membranes, 576, 577
protein secretion in, 106
Green fluorescent protein (GFP), 154, 155
marker for transgenic plants, 611
Greenpeace, 698, 817
Green Revolution, 726
Griseusin, biosynthesis of, 480
Growth hormone. *See also* Salmon
Chinook salmon, 732, 787
derivatization of, 249
levels in rat blood, 250
treatment of deficiency, 248–250

H

Haemophilus influenzae, 4
annual cases in Canada, 388
sequenced genome of, 56–57
Half-life, extending protein, 252–253
Halophilic bacteria, isolation of genes, 34
Hansenula polymorpha, 134–136
Harvard University, 362

Hawaiian papaya industry, 663
Health Canada, 806
Hedén, Carl Göran, 2
Helicobacter pylori
  as antigen delivery system, 448–449
  development of recombinant *L. acidophilus* vaccine against, 449
Hemagglutination test, blood type, 186
Hemagglutinin
  development of llama antibodies, 451
  influenza virus, 420
  neutralizing antibodies against influenza virus, 450
Hemicellulose. *See also* Cellulose and hemicellulose
  schematic of, 520
Hemochromatosis, 833
Hemoglobin, protein synthesis in oxygen-limited *E. coli*, 546
Hemophilia, bispecific antibodies against, 300
Heparin, 336
Hepatitis B, diagnostic monoclonal antibodies, 185
Hepatitis B virus, 422
Hepatitis C virus infection, 247–248
Herbicide resistance, 668–674
  biological manipulations, 668–669
  biosynthesis of branched-chain amino acids, 674
  bromoxynil (3,5-dibromo-4-hydroxy benzonitrile), 673
  chlorosulfuron, 674
  dicamba, 672–673
  examples of gene-based, 669
  glyphosate, 669–672
  transgenic maize varieties, 673–674
Hereditary transthyretin amyloidosis, patisiran (siRNA drug), 347
Herpes, diagnostic monoclonal antibodies, 185
Herpes simplex viruses (HSV)
  acyclovir for HSV-1 and HSV-2, 310
  deletion of virtual glycoprotein D from HSV-2, 425
  engineered attenuated vaccines, 425–426
  insertion of target DNA into vector, 26
  organization and structure of HSV-1, 361
  ribozyme for, 340
  vaccines for HSV-1 and HSV-2, 393–394
  viral delivery systems of HSV-1, 357, 358, 361–362
Herpes thymidine kinase, somatic expression in mice, 732
Herpesvirus, ribozyme for, 340
Heterocyclic compounds, *Burkholderia* spp. degrading, 497
Hexahydro-1,3,5-trinitro-1,3,5-triazine, 803
High-performance liquid chromatography (HPLC), 467
High-pressure homogenization, microbial cell disruption, 577, 578
High-throughput next-generation sequencing, 59–60
Hirudo medicinalis, 129
HIV. *See* Human immunodeficiency virus (HIV)
Horvath, Philippe, 37
HPV. *See* Human papillomaviruses (HPV)
HSV. *See* Herpes simplex viruses (HSV)
Human(s)
  genetic code and codon usage in *E. coli* and, 99
  production of recombinant proteins, 92
  recombinant proteins approved for, 244
  transferring transgenes from food to, 819–820
Human blood coagulation cascade, 301
Human cell line editing, 832
Human chorionic gonadotrophin (hCG) protein, pregnancy detection, 188–190
Human cytomegalovirus (CMV), 422
Human diseases
  clinical trials of siRNA-based drugs, 346
  mouse model of Duchenne muscular dystrophy, 759–760, 761
  mouse models of Alzheimer's disease, 756–759
  nonhuman primate models of neurodevelopmental disorders, 766–767
  rabbit models of cardiovascular disease, 761–763
  therapeutic systems for disorders, 326–327
  transgenic animal models of, 756–767
  zebrafish melanoma model, 763–765, 766
Human embryos, modified CRISPR-Cas system and, 350
Human enzyme deoxyribonuclease I (DNase I), 253–254
  digestion of DNA, 254
Human gene therapy. *See* Gene therapy
Human genetic diseases. *See* Genetic diseases; Human diseases
Human growth hormone, 248–250; *see also* Growth hormone
  International Olympic Committee, 250
  levels in rat blood, 250
Human immunodeficiency virus (HIV)
  AIDS and, 359, 388
  anti-HIV antibodies, 300–302
  diagnostic monoclonal antibodies, 185
  droplet digital PCR, 213
  human germ line editing, 832
  inhibitor, 274–276
  monoclonal antibody against, 591
  type 1 glycoprotein gp160, 261
Human interferons, 244–248
  construction of hybrid interferon-α (IFN-α) genes, 246
  hepatitis C virus infection, 247–248
  strategies for constructing hybrid IFN genes, 246
  synthesis of *E. coli* of polypeptide with leukocyte activity, 245
Human interleukin-7
  flow chart for large-scale production of, 572
  immunotherapy agent, 571–572
Humanized antibody, 284, 285
Humanized monoclonal antibody, anthrax, 312
Human lysosomal acid lipase, transgenic chicken eggs, 771
Human monoclonal antibodies, 287–289
Human papillomaviruses (HPV)
  immunization approval, 828
  vaccines for, 402–404
Human protease inhibitor, production in rabbits, 770
Hunter syndrome, 267
  zinc finger nucleases, 349
Huntington's disease, 197, 331, 834
  development of mouse model for, 752–753
  nonhuman primate models of, 766–767
  single-stranded short interfering RNA (siRNA), 345
Hybridization probes. *See also* DNA-based diagnostics
  allele-specific, 200, 202, 204
  allele-specific, screening for cystic fibrosis, 203
  Alzheimer's disease using, 231–232
  detection of sickle-cell β^5-globin allele, 202
  DNA-based diagnostics, 199–207
  dysbiosis of intestinal microbiota, 201
  identifying bacteria, 199–200
  molecular beacons, 206–207
  oligonucleotide ligation assay (OLA), 204–205
  padlock probes, 205–206
  quantifying RNA, 232
Hybridoma therapy, recombinant antibodies, 280–281
Hybrid plants
  clonal seeds from, 619–621
  male sterility and, 621
Hydrogen, production, 535–538
Hydroxyacetosyringone, structure of, 595
Hyperglycemia, 278
Hypocrea jecorina, engineered microbial enzymes, 529
Hypoglycemia, 278
Hypoxanthine, aminopterin, and thymidine (HAT medium), 183, 184
Hypoxanthine-guanine phosphoribosyltransferase (HGPRT), 183, 184

I

Ibritumomab, non-Hodgkin's lymphoma, 284
Ideonella sakaiensis, plastic biodegradation, 507–508
Illegitimate drugs, 837

Illumina sequencing, 53–55
Illumina technology, 61
Imidazolinones, herbicide resistance, 669
Immobilized fusion protein, 582
Immunoglobulins, 180; *see also* Antibodies
Immunological assay
  protein conformation-specific disorders, 197–198
  screening of genomic DNA library using, 35
Immunology, 5
Immunotherapy agent, human interleukin-7, 571–572
Imperial Chemical Industries, Ltd., 565
Impingement, microbial cell disruption, 577, 578
India, transgenic cotton, 594
Indigo, 467–468
  biosynthesis from tryptophan, 468
  synthesis of, 467
Indigofera tinctoria, 467
Induced mutagenesis and selection, 2
Industrial chemicals, 5
Infectious diseases, 7; *see also* Livestock, disease-resistant
  detecting by indirect ELISA, 192–194
  nucleic acid-based tests, 834
  real-time PCR to diagnose, 211–213
  targets for monoclonal antibodies, 185
  toxin-intein infusions for bacterial, 264–265
Influenza, 437
  complement inhibitor C4 binding protein (OVX313), 397
  DNA vaccines, 420
  passive immunity to virus, 450–452
  schematic of virus, 397, 420
  subunit and peptide vaccines, 396–397
Influenza-resistant chickens, 779–781
Inotersen, 329
Insect(s)
  N-linked oligosaccharides, 122
  O-linked oligosaccharides, 121
Insecticidal toxins
  limiting expression time of, 656
  properties of, 639
Insect resistance, 637–658
  α-amylase inhibitor, 645–646
  Bacillus thuringiensis insecticidal toxin, 637, 639–642
  cholesterol oxidase, 646–647
  gene stacking, 652–653
  increasing binding targets, 655
  increasing expression of *B. thuringiensis* protoxin, 642–645
  lectins, 648
  limiting expression time of insecticidal toxin, 656
  mimic *B. thuringiensis* subsp. *israelensis* behavior, 654–655
  preventing development of *B. thuringiensis*-resistant insects, 652–656
  protease inhibitors, 645
  proteins protecting plants, 648–649
  refugia-like strategy, 652
  RNA interference, 650–652
  spatial refugia, 652
  strategies for protecting plants, 645–652
  targeting aphids, 657–658
  toxins, 653
  use chemical insecticide with *B. thuringiensis* toxin, 656
  vegetative insecticidal toxins, 647–648
Insertion mutations, large and small, 158
Insulin
  batch versus fed-batch fermentation for B peptide, 570
  cleavage of inactive preproinsulin, 277
  genetic regulation of expression in *Lactococcus lactis*, 277
  hyperglycemia, 278
  hypoglycemia, 278
  preproinsulin to, 120, 121
  recombinant human, 244
  schematic of glucose-responsive, 278
  type I diabetes, 276
Insulin-like growth factor 1 receptor cDNA, 329–330
Integrating vectors, *Saccharomyces cerevisiae* expression, 126, 127
Interfering RNA (ribonucleic acid), 341–347
  approval of first RNAi drug, 347
  clinical trials of siRNA-based drugs, 346
  gene silencing, 341–343
  Huntington's disease, 345
  phenomenon of RNA interference (RNAi), 342, 343–344
  research tool in gene expression, 347
  schematic of silencing expression of gene by, 343
  small (siRNA) or short hairpin RNA (shRNA), 343, 344–347
  therapeutic use of siRNA, 345
Interferons (IFNs). *See also* Human interferons
  human, 244–248
  production of IFN-γ in *E. coli*, 570
  vaccinia virus and, 435, 437
Interleukin-10, ulcerative colitis and Crohn's disease, 270–272
International Olympic Committee, human growth hormone, 250
International Sequence Database Collaboration, 63
Intestinal microbiota, diagnostic test for dysbiosis of, 201
Intestinal microorganisms, transferring transgenes from food to, 819–820
Ion-exchange chromatography, 586
Iron
  ferritin, 699, 700
  plant nutritional content, 698–700
Ishino, Yoshizumi, 37
Isoptericola variabilis epsps Gene, rice plants, 670

J

Jadomycin, *Streptomyces venezuelae*, 486, 487
Japanese encephalitis, 422
J. Craig Venter Institute, 46
Jenner, Edward, 387
Joint Food and Agricultural/World Health Organization Expert Committee on Food Additives, 806
Jost, A., 4

K

Kadcyla (ado-trastuzumab emtansine), 828
Kanamycin, 568
  resistance, 501, 502
Karembu, Margaret, 594
Kearns-Sayre syndrome, 266
KHARON1, *Leishmania* vaccines, 430
Klebsiella oxytoca, 534
Klebsiella ozaenae, gene for nitrilase, 673
Klenow polymerase, 15, 17
Kluyveromyces lactis, chymosin, 805
Krebs cycle, 560
Kymriah, 831

L

Labeling, genetically modified foods, 820–822
Lachnospiraceae, Cas12 from, 218
Lactate production, tricarboxylic acid (TCA) cycle, 151
Lactic acid bacteria, 268
  clinical trials for delivering recombinant proteins, 270
  delivery of vaccines, 411–412
  overview for vaccine antigen delivery, 412
Lactobacillus
  HIV inhibitor, 274
  insulin and *L. lactis*, 277
  lactic acid bacteria, 268
  milk quality, 781
Lactobacillus jensenii, HIV inhibitor, 274–276
Lactococcus, lactic acid bacteria, 268
Lactococcus lactis, 92
  interleukin-10, 270–272
  lactic acid bacteria, 268
  therapeutic proteins expressed in, 269
Lactose intolerance, 582
Land area, cultivation of transgenic crops, 807
Lapaquistat
  chemical structure of, 483
  squalene synthase, 482
Leader peptide, 107
Leber hereditary optic neuropathy, 266
Lectins, plant protection against insects, 648
Legionella, diagnostic monoclonal antibodies, 185
Legionella pneumophila, hybridization to identify, 199
Leishmania species
  engineered attenuated vaccines, 430

immunity of virulent *L. major*, 431
schematic of auxotrophic cell line, 431
Lemna minor
protein glycosylation, 623–624
structures of glycans from native and modified, 624
Lentiviruses
anti-HIV antibodies and, 301
human immunodeficiency virus type 1 (HIV-1) and AIDS, 359
schematic of virion, 359
viral delivery systems, 358, 359–360
Leptin
construct facilitating secretion from *Lactococcus lactis*, 273
produced in *E. coli*, 273
strategy to release from *L. reuteri*, 274
weight reduction and, 272–274
Lethal factor (LF), 311
Leucinodes orbonalis, 594
Leuconostoc, lactic acid bacteria, 268
LexA protein, 154, 155
Libraries. *See also* Combinatorial libraries; DNA library; Genomic libraries
combinatorial, of antibody fragments, 294–296
Ligand binding affinity, increasing, 167–169
Lignin, 519
biosynthesis, 718
genetically modified in trees, 719
plant yield content, 717–720
Populus tremuloides (aspen trees), 718
schematic of, 520
Lignocellulose, 538
Lignocellulosics, 519–522; *see also* Cellulose and hemicellulose
biomass composition by source, 520
Lipids
chemical structures of common, 693
cholesterol, 367
conjugating or encapsulating siRNAs in, 368
from cyanobacteria, 534–535
dietary fats in various oils, 694
nonviral delivery system, 367–368
plant nutritional content, 692–695
schematic of lipopeptide nanoparticle, 369
Lipoamide dehydrogenase (LAD), 267
Lipopeptide nanoparticles, modular design of, 368, 369
Liposome, schematic of nucleic acid in, 339
Liquid bead assay, protein biomarker detection, 196
Listeria monocytogenes, nonviral delivery, 369
Listeria seeligeri, Cas13 endonuclease, 219
Liu, David, 362
Liver disease, hepatitis C virus infection, 247–248
Livestock, 5
genetically engineered, 810–812
Livestock, disease-resistant, 774–781
influenza-resistant chickens, 779–781
mastitis-resistant cows, 778–779
prion-resistant cattle, 775–776
tuberculosis-resistant cattle, 776–778
Living organisms, patenting, 842–844
Longfellow, Henry Wadsworth, 679
Loop-mediated isothermal amplification (LAMP). *See also* Polymerase chain reaction (PCR) detection methods
assay, 215
point-of-care diagnostics, 214, 216
Luciferase, 234–235
Lupine, 691
Luxturna, 831
Lycopene, 473–474
chemical structure of, 474
overview of production of, 474
Lysine, 691, 692

M

Macaca fascicularis, growth hormone, 249
Madagascar periwinkle plants (*Catharanthus roseus*), proteins and, 649
Magnoporthe oryzae, 681
Maize (*Zea mays*), herbicide resistance, 673–674
Major histocompatibility complex (MHC-II), T-cell epitopes and, 263–264
Major League Baseball, human growth hormone, 250
Malaria, 388, 795
difficulties in developing vaccines, 410–411
schematic of RTS,S hybrid vaccine particle in treatment of, 411
vaccines for, 408–411
vector population suppression, 789–791
Male sterility, hybrid plants conferring, 621
Mammalian cell expression system, 145–155
chromosomal integration and environment, 152–155
engineering hosts for enhanced productivity, 150–152
expression vectors for heterodimeric or -tetrameric proteins, 149
increasing host cell viability, 150–152
increasing protein secretion, 152
selectable markers for vectors, 148, 150
strategies to increase expression of recombinant proteins in, 155
strategy to increase yields of secreted recombinant proteins, 153
transcription signals, 146, 147
translation signals, 147–148
vector design, 145–148
Mammals
N-linked oligosaccharides, 122
O-linked oligosaccharides, 121
MammaPrint microarray, 227–228, 239
Mammary glands
milk production and recombinant protein yields, 769
recombinant therapeutic proteins in, 768
transgenic animals, 795
Manipulation of gene expression in plants
Agrobacterium tumefaciens and lettuce, 613–614
alcohol dehydrogenase gene promoter from *Aspergillus nidulans*, 614
clonal seeds from hybrid rice, 619–621
conferring male sterility, 621
CRISPR-based directed evolution, 625–626
CRISPR/Cas9 system, 617, 618–620
ethanol for activation, 614
facilitating protein purification, 621–622
first-generation viral vectors, 612
genes in plants, 617–621
gene stacking, 624–625
oleosins, 621–622
plant promoters, 616, 617
plant RNA transcription complex, 617
polycistronic gene expression, 626, 627
protein glycosylation, 623–624
rhizosecretion, 622
second-generation transient-expression vector system, 614
second-generation viral vectors, 612
transformation of head of lettuce, 613–614
transient expression of antibodies, 615–616
transient gene expression, 611–616
Marraffini, Luciano, 37
Masking, derivatized monoclonal antibodies, 308–309
Massachusetts Institute of Technology, 845
Massive parallelization, 57
Mastitis-resistant cows, 778–779
Maxam, Allan, 51
Measles, 387, 388
Measles, mumps, and rubella (MMR) vaccine, 830
Medical products of biotechnology
antivaccine movement, 830
approved gene therapies, 831
economic issues, 835–837
FDA-approved nucleic acid-based tests, 834
genetic and genomic testing, 832–834
illegitimate drugs, 837
new biological drugs, 828–832
orphan drugs by FDA, 836
regulation and safety of, 827–837
Melanoma, 210
zebrafish model of, 763–765, 766
MELAS syndrome (mitochondrial myopathy, encephalopathy, lactic acidosis and stroke), 266
Membrane filtration, harvesting microbial cells, 574–575
Membrane permeabilization, protein secretion, 109, 110

Messenger RNA (mRNA), 12–20
codons in, by site-directed mutagenesis, 160
large-scale production of, 586
Metabolomics, 84–85
pulmonary arterial hypertension (PAH), 85
Metagenomes
construction of libraries, 62
sequencing, 61–62
Methanococcus jannaschii, 159
Methicillin-resistant *Staphylococcus aureus* (MRSA)
deaths from infection by, 481
hospital-acquired infection, 401
real-time PCR to diagnose, 211–213
Mice. *See* Mouse
Microarrays, DNA, 68, 219–225
ancestry determination, 223–225
animal species determination, 225
detection of epigenetic markers, 221–223
detection of multiple alleles, 220
drugs metabolized by cytochrome P450, 220
hypermethylation of DNA, 221
methylation of DNA, 221, 223
Microbes, compounds metabolically engineered in, 463
Microbial biofilm, schematic of, 255
Microbial biosensors, 235–238
Microbial cells
disrupting, 576–578
harvesting, 574–575
methods of mechanical cell disruption, 577
schematic of methods for disruption, 577
Microbial degradation of xenobiotics
biodegradation of organic contaminants, 496, 499
CAM (camphor-degrading) plasmid, 498–499, 500
degradative pathways of *Pseudomonas* plasmids, 498
genetic engineering of biodegradative pathways, 497–507
manipulation by gene alteration, 499–507
manipulation by transfer of plasmids, 498–499
OCT (octane-degrading) plasmids, 498–499, 500
plastics, 507–508
toluene- and xylene-degrading pathway of plasmid pWWO, 499–503
Microbial growth, 545–546; *see also* Fermentation
batch fermentation, 548–549
continuous fermentation, 550–551
fed-batch fermentation, 549–550
principles of, 547–551
Microbiology, 5
Microinjection of DNA
efficiency of transgenesis process after, 735
genetic transformation of mouse embryos, 732
transgenic animals, 733–735
transgenic strains in chicken eggs, 771, 772
Microorganisms. *See also* Genetically modified microorganisms
restriction endonucleases, 460–461
Microprojectile bombardment
biolistics, 603–604
CRISPR/Cas9 system into wheat plants, 619
DNA bound to microprojectile particle, 603
procedure increasing transformation of monocots, 604, 605
schematic of, 604
MicroRNAs (miRNAs)
Alzheimer's disease using hybridization probes, 231–232
cancer miRNA signature, 231
detection of miRNA signatures of disease, 230–232
production and activity of, 230
virus resistance, 664–665
Microsatellites, forensic analysis using PCR-amplified, 216–218
MIDGE (minimalistic immunogenically defined gene expression) vectors, DNA vaccines, 421–422
Milk
collection of, 732–733
improving quality of, 781–782
protein composition from cattle and sheep, 732
recombinant antithrombin in goats, 768–769
Mipomersen, 329
cholesterol-lowering, 330
Mitochondria, enzymes targeting, 265–267
Mitochondrial diseases
gene therapy for, 378, 379
schematic for development of three-parent baby, 379
Mitochondrial replacement therapy, 378
Mojica, Francisco, 37
Molecular beacons, 206–207, 208
Molecular biology, 2, 5
Molecular biotechnology, 86–87
advances, 799, 845–846
benefiting humanity, 8–9
commercialization of, 6–8
developments in history of, 4–5
emergence of, 1–3
microbial growth, 545–546
positive aspects of advances, 9
research restriction endonucleases, 459–461
scientific disciplines contributing, 5
Molecular chaperones, 101
Molecular cloning
cloning eukaryotic genes, 26, 28–30
genomic libraries, 32–36
insertion of target DNA into plasmid vector, 18–21
preparation of DNA for, 13–18
recombinational cloning, 30–32
transformation and selection of cloned DNA in bacterial host, 21, 23–24, 26
Molecular diagnostics, 179–180
immunological approaches to protein biomarker detection, 180–198
Molecules, small biological. *See* Small biological molecules
Moloney murine leukemia virus, 358
Monarch butterfly, 823–824
Monoaromatic hydrocarbons, *Burkholderia* spp. degrading, 497
Monoclonal antibodies, 182–185, 243–244
approved for human use in US or EU, 282–283
comparing properties of aptamers and, 332
creation of hybridoma cell line, 184
diagnosis of acute myocardial infarction, 191
engineered to bind to cancer, 309
first human, against human immunodeficiency virus (HIV), 302
HIV, 591
human, 287–289
hybrid human-mouse, 284, 286–287
masking and derivatized, 308–309
non-Hodgkin's lymphoma treatment, 284
nonviral delivery, 373
passive immunity, 449–452
synthesis in plants, 708–709
targeting tumor cells for destruction, 304
targets for diagnostic, 185
trastuzumab for breast cancer treatment, 281
Mononucleosis, 437
Monsanto, 824
Monsanto Corporation, glyphosate, 669–670
Mosquitoes, dengue fever virus-resistant, 791–792
Mouse
DNA microinjection method, 733–735
genetic transformation of embryos by DNA microinjection, 732
mammary tumor virus, 843
model for Huntington's disease, 752–753
model of Duchenne muscular dystrophy, 759–760, 761
models of Alzheimer's disease, 756–759
transgenesis, 732
mRNA sequences. *See* Messenger RNA (mRNA); Targeting mRNA and DNA sequences
Mucosal immunity, DNA vaccines, 417–418

Mullis, Kary, 40
Multiple sclerosis, 7
Mumps, 387, 388
Mung bean nuclease, 15, 17
Muscle mass
  Belgian blue cattle, 785
  double muscling, 783
  increasing in cattle, 782–786
  transgenic mice with, 784
Muscle weakness, L-citrulline, 471
Mutagenesis. *See also* Protein engineering
  Corynebacterium sp. gene, 466
  directed, 156–160
  DNA shuffling, 162–163, 164
  error-prone PCR, 161
  mutant proteins with unusual amino acids, 159–160
  random, 161–163
  random, with degenerate oligonucleotide primers, 161–162
  site-directed, by inverse PCR, 158–159
  site-directed, by overlap extension PCR, 156–158
Mutant proteins, unusual amino acids, 159–160
Mycale hentscheli
  antiviral and antitumor agents, 63
  microbiome of, 61
Mycobacterium spp., engineering bacteriophage, 354
Mycobacterium bovis, tuberculosis-resistant cattle, 776–778
Mycobacterium tuberculosis
  antibiotic sensitivities, 228
  droplet digital PCR, 213
  drug-resistance strains, 836
  tuberculosis-resistant cattle, 777
  vaccine against, 441–444
Mycoplasma, oxytetracycline, 486
Mycoplasma capricolum, 47
Mycoplasma genitalium, genome of bacterium, 47
Mycoplasma mycoides, genome of, 47
Myostatin, 783
Myriad Genetics, 842

N

Nanoparticles
  chemotherapeutic agent delivery, 352
  commercially available dendrimers, 372
Nanopatch, vaccine delivery, 412, 413
Nanopore sequencing, 56
Nanozymes
  cleaving cellular mRNAs, 351–352
  schematic of, 352
National Association of Boards of Pharmacies, 837
National Bioengineered Food Disclosure Law, 5, 821
Naturally occurring organic compounds, metabolically engineered in microbes, 463
Natural phenomena, 842
Nature (journal), 823
Neisseria gonorrhoeae, hybridization to identify, 200
Neurodevelopmental disorders, nonhuman primate models of, 766–767
Neurogenerative diseases, 331
New biological drugs, 828–832
New York Stock Exchange, 6
New Zealand Environmental Protection Agency, 618
Nicotiana benthamiana
  inhibition by single-chain antibody, 664
  producing humanized monoclonal antibodies in, 615
  transient expression of antibodies, 615–616
Nicotine, biosynthesis of, 723
Nigeria, transgenic cowpea, 594
Nocardia argentinensis, 17
Nocardia corallina, 17
Non-Hodgkin's lymphoma
  obinutuzumab, 286–287
  therapeutic monoclonal antibodies for, 284
Nonhuman epitopes
  masking, 263–264
  schematic of modification, 263
Nonviral delivery systems. *See also* Viral delivery systems
  antibodies, 373
  aptamers, 373–374
  bacteria, 369–371
  dendrimers, 372
  direct injection, 365–367
  lipids, 367–368, 369
  transposons, 374–376
Nopaline, structure of, 599
Northwestern University, 37
Nostoc ellipsosporum, HIV inhibitor, 274–275
Novel and Exceptional Technology and Research Advisory Committee (NExTRAC), 801
Nucleic acid
  patenting sequences, 841–842
  schematic of liposome containing, 339
Nusinersen, 329
Nutritional content, alteration in genetically engineered foods, 812–815

O

Obese (ob) gene, leptin, 272–274
Obesity, antiobesity antibodies, 313–315
Obinutuzumab
  non-Hodgkin's lymphoma, 284, 286–287
  schematic of mode of action of, 287
OCT (octane-degrading) plasmid, microbial degradation, 498–499, 500
Octopine, structure of, 599
Octopine synthase, regeneration of plants expressing, 596
Oidium neolycopersici, 681
Oleispira antarctica, molecular chaperone, 101
Oleosins
  protein purification, 621–622
  schematic of fusion protein, 622
Oligonucleotide ligation assay (OLA), 204–205
Oligonucleotides
  antisense RNA, 327–328
  assembling into genes, 46, 48
  chemical synthesis of primers, 162
  detection of sickle-cell β^5-globin allele by hybridization, 202
Oligosaccharides
  N-linked, 122
  O-linked, 121
Oncomouse, 844
100,000 Genomes Project, 63
One-step homologous recombination, chromosomal integration by, 116–117
Organic contaminants/pollutants
  biodegradation of, 499
  phytoremediation, 686–688
Organ transplantation, donor organs in pigs, 773–774
Orphan Drug
  definition, 835
  FDA-approved, 836
  patisiran, 347
Orthomyxoviridae Family, influenza viruses, 397
Osteomyelitis, engineering bacteriophage, 354–355
Ovarian cancer, diagnosis of, 190–191
Oxygen content, plant yield, 722–723
Oxytetracycline
  discovery and commercialization, 486
  engineering rapid synthesis of, 487
  structure of, 486

P

Pacific Biosciences (PacBio), 55
Padlock probes, 205–206
  function of, 205
Panicum virgatum (switchgrass), 530
Pantoea agglomerans (*Erwinia herbicola*)
  L-ascorbic acid, 463–464, 465
  lycopene production, 473, 474
Papaya (*Carica papaya*), 663
Papaya industry, saving Hawaiian, 663
Papaya ringspot mosaic virus, CRISPR/Cas9, 667
Papaya ringspot virus (PRSV), 663
  viral coat protein, 662
Parkinson's disease, 168, 197, 331
  L-citrulline, 471
  Enterococcus faecalis and dopamine, 279, 280, 318
Passive immunity
  influenza virus, 450–452
  monoclonal antibody, 449–452
Pasteur, Louis, 842

Patent Appeals Board, 839
Patenting, 838–839
  biotechnology, 837–845
  in different countries, 839–841
  fundamental research and, 844–845
  living organisms, 842–844
  nucleic acid sequences, 841–842
  types, 838
Patisiran, first RNAi drug, 347
PCR. *See* Polymerase chain reaction (PCR)
Pea plants (*Pisum sativum*), α-amylase inhibitor, 646
Pectin content, plant yield, 720–722
Pediococcus, lactic acid bacteria, 268
Pegaptanib, aptamer as therapeutic agent, 334–335
Pegylated therapeutic protein, 248
Pegylation, enhancing antibody half-life, 315
Penicillin, 352
Penicillium, 136
Penicillium chrysogenum, penicillin, 485
Penicillium oxalicum, raw starch-degrading enzyme (RSDE), 513
Pentachlorophenol, 496
Peptide mass fingerprinting, 72, 73
Peptide vaccines, 392–414, 407; *see also* Subunit vaccines; Vaccines
  malaria, 408–411
  usefulness of, 414
Pesticides, poisoning, 503
Pharmaceuticals, 244–253
  clinical trials for, 247
  extending protein half-life, 252–253
  human growth hormone, 248–250
  human interferons, 244–248
  plants as reactors, 709–710
  plant-synthesized examples, 709
  protein production in plants, 592
  tumor necrosis factor alpha (TNF-α), 251
Pharming, 709, 732
Phenoxycarboxylic acids, herbicide resistance, 669
Phenylalanine ammonia lyase, 258–259
  fate of dietary phenylalanine, 260
Phenylketonuria, 258
  phenylalanine ammonia lyase and, 258–259
Phipps, James, 387
Photorhabdus luminescens, protein toxin A, 649
Phred quality score (Q), 60
Phytophthora capsici, 681
Phytoremediation
  phytodegradation, 686, 687
  phytoextraction, 687
  phytotransformation, 687
  phytovolatization, 687
  rhizodegradation, 687
  rhizofiltration, 687
  schematic overview, 687
Pichia pastoris, 174
  codons optimized for, 171
  differential processing of glycoproteins, 133
  expression vectors, 134
  "humanized", 132–134
  integration of DNA into, 135
  interleukin-7, 572
  vector system, 136
Picornavirus, foot-and-mouth virus, 405
Pierce disease, wine disease, 680–681
PiggyBac, gene therapy delivery vector, 375
Pigs, production of donor organs, 773–774
Pink bollworm (*Pectinophora gossypiella*), resistance of, 656
Plant(s). *See also* Manipulation of gene expression in plants; Plants as bioreactors; Transgenic plants
  chloroplast engineering, 604–608
  debate over genetically modified crops, 594
  genetic engineering of, 633
  microprojectile bombardment, 603–604
  overview of systemic acquired resistance, 675
  pharmaceutical proteins produced in, 592
  potential vaccine antigens, 593
  reporter gene use in transformed plant cells, 610–611
  strategies for protecting against insects, 645–652
  therapeutic antibodies and antibody fragments, 593
  Ti plasmid-delivery vector systems, 597–603
  transformation with Ti plasmid of *A. tumefaciens*, 595–597
Plant nutritional content
  amino acids, 690–692
  gluten, 700–701
  iron, 698–700
  lipids, 692–695
  modification, 690–701
  vitamins, 695–698
Plants as bioreactors
  antibodies, 706–709
  comparing recombinant protein production, 706
  pharmaceuticals, 709–710
  poly(3-hydroxybutyric acid), 710–711
  vaccines, 709–710
Plant yield
  decreasing lignin content, 717–720
  decreasing pectin content, 720–722
  increasing grain yield, 716
  increasing harvest index, 716–717
  increasing oxygen content, 722–723
  lignin genetically modified in trees, 719
  native and transgenic plant cell walls, 720
  overexpression of switchgrass *MYB4* gene, 721
  role of 4-coujarate:coenzyme A (CoA) ligase in lignin biosynthesis, 718
  structure of pectin molecule, 722
Plasmid DNA (deoxyribonucleic acid)
  E. coli genes mutating to improve production of, 585
  large-scale production, 583–586
  schematic of process for continuous thermal lysis of *E. coli* cells, 584
  schematic of reduction of chromosome size of host *E. coli* cell, 586
Plasmids, 6, 18
  biparental conjugation, 26, 27
  insertion of target DNA into vector, 18–21
  stability in fermentation, 555
  triparental conjugation, 26, 27
Plasmin, 839
Plasmodium spp.
  difficulties in developing malaria vaccines, 410–411
  drug-resistance strains, 836
  infection of individual with *P. falciparum*, 409
  malaria, 408–411
  malaria vector population suppression, 789–790
  schematic of P. *falciparum*, 410
  vaccine for gene expression, 100
Plastics, microbial degradation of, 507–508
Pneumonia, 437
Point mutation, 158
Polio, annual cases in Canada, 388
Poliomyelitis, 387, 389
Poliovirus, 437
Poly(3-hydroxybutyrate), fed-batch fermentation, 573–574
Poly(3-hydroxybutyrate-co-3-hydroxyvalerate), fed-batch fermentation, 573–574
Poly(3-hydroxybutyric acid), plants as reactors, 710–711
Polycistronic gene expression, 626, 627
Polyclonal antibodies, 182
Polycyclic aromatic hydrocarbons, *Burkholderia* spp. degrading, 497
Polyethylene glycol, structure of, 248
Polygenic diseases, protein arrays to detect, 194–197
Polyhydroxyalkanoates (PHAs), 489–493
  Azotobacter sp. genes for, 492
  bacterial synthesis of, 490, 491
  chemical backbone of, 489
  copolymer poly(3-hydroxybutyrate-co-3-hydroxyvalerate), 490, 491
  poly(3-hydroxybutyrate), 490, 491
  schematic of types, 493
Polyketide
  engineering, antibiotics, 477–480
  minimal, synthase, 478–480
  term, 477
Polymerase chain reaction (PCR)
  addition of restriction enzyme recognition sites to target DNA, 43
  amplification of DNA by, 39–42

annealing, 39–40
assembling products into genes, 46–47, 49
cloning PCR products, 42, 43, 44
cycles of, 41
denaturation, 39
error-prone PCR, 161, 162
extension, 40
quantitative PCR, 44–45
TaqMan PCR protocol, 209–211
Polymerase chain reaction (PCR) detection methods
allele-specific PCR, 207–209, 239
droplet digital PCR quantifying low-abundance sequences, 213–214
forensic analysis using PCR-amplified microsatellites, 216, 217, 218
loop-mediated isothermal amplification for point-of-care diagnostics, 214, 216
real-time PCR diagnosing infectious diseases, 211–213
TaqMan PCR for SNP genotyping, 209–211
Polypeptide hormones, targets for monoclonal antibodies, 185
Polyplexes, 372
Polysaccharide. *See also* Starch and sugars
starch, 508
Pompe disease, 267
Porcine endogenous retrovirus (PERV), genome-integrated pig pathogens, 774
Porcine parvovirus, flow chart for vaccine development against, 573
Posilac, 806
Posttranslational modification,
antibody microarrays, 75, 77
eukaryotic proteins, 120–123
Potato virus X, expression system, 707, 708
Potyvirus, ribonuclease, 667
Pregnancy test, sandwich ELISA, 188–190
Preproinsulin, insulin from, 120, 121
Prevotella, 200
Prime editing
Cas9 enzyme, 351
schematic of, 351
Prion diseases, 197
Prion-resistant cattle, 775–776
Prodrug activation therapy
conversion of ganciclovir to ganciclovir monophosphate, 378, 380
gene therapy for, 378–379
overview of, 380
Prokaryotic hosts
affinity purification systems, 111–113
facilitating protein purification, 111–114
factors increasing metabolic burden, 93
increasing protein secretion, 105–111
increasing protein stability, 100–105
increasing translation efficiency, 96–100
integrating DNA into host chromosome, 114–119
promoters for expression of cloned genes, 93
protein production in, 91–119
proteolytic removal of purification tags, 113, 114
regulating transcription, 92–96
secretion into medium, 108–111
secretion into periplasm, 105–108
self-cleaving purification tags, 114
Promoterless gene targeting, 379, 381
Promoters
inducers of *lac*, 95
T7 expression system, 96, 97
tac promoter from *trp* and *lac*, 94
transcription from *lac*, 95–96
Prostate cancer
aptamers and, 335
insulin-like growth factor 1 receptor, 327, 329
Prostate-specific membrane antigens (PSMA), 338
Protease inhibitors, plant protection against insects, 645
Protease sensitivity, decreasing, 171–173
Protective antigen (PA), 311
Protein(s)
amino acids of, 851
expression profiling, 74, 74–77
glycosylation, 623–624
identification of, 71–74
large-scale screens for protein interactions, 82
modifying specificity of, 167
peptide mass fingerprinting, 72, 73
plant protection against insects, 648–649
protein-protein interactions, 78–84
self-splicing, 115
tandem affinity purification for multiprotein complexes, 83
2D PAGE for separation of, 71, 72
two-hybrid assay for detecting pairwise interactions, 80, 81
Protein biomarker detection, 238; *see also* Antibodies
agglutination, 185, 186
antibodies, 180–185
detecting infectious diseases by indirect ELISA, 192–194
detecting pregnancy by sandwich ELISA, 188–190
diagnosing autoimmune diseases by indirect ELISA, 191–192
enzyme-linked immunosorbent assays (ELISA), 186–194
immunoassays for protein conformation-specific disorders, 197–198
immunological approaches to, 180–198
measuring disease-associated proteins by sandwich ELISA, 190–191
protein arrays detecting polygenic diseases, 194–197
Protein conformation-specific disorders, immunoassays for, 197–198
Protein engineering, 156–173
decreasing protease sensitivity, 171–173
directed mutagenesis, 156–160
examples of, 164–173
increasing ligand binding affinity, 167–169
increasing protein stability, 164–166
modifying cofactor requirements, 169–171
modifying protein specificity, 167
random mutagenesis, 161–163
Protein folding
disulfide bonds, 102–103
endoplasmic reticulum in eukaryotes, 129–130
facilitating, 101–103
yeast endoplasmic reticulum cells, 131
Protein fusion, purification of foreign proteins in *E. coli*, 113
Protein half-life, extending, 252–253
Protein p53, cancers and mutation of, 336–337
Protein production, 173–174, 587; *see also* Downstream processing; Fermentation systems, large-scale
comparing recombinant, in plants and systems, 706
downstream processing, 578–583
flow chart of purification of target recombinant protein, 583
heterologous, in eukaryotic cells, 119–155
inclusion bodies, 579–581
induced recombinant with batch and fed-batch *E. coli* cells, 571
magnetic separation of proteins, 582–583
schematic of large-scale protein purification using ultrafiltration membranes, 580
schematic of large-scale ultrafiltration dialysis, 579
Protein purification
affinity purification systems, 111–113
control of minimal type III secretion system (T3SS), 112
facilitating, 111–114, 621–622
oleosins, 621–622
proteolitic removal of purification tags, 113, 114
rhizosecretion, 622
self-cleaving tags, 114, 115
self-splicing proteins, 115
Protein secretion
in bacteria, 106
extracellular systems, 109–111
fermentation, 558, 559
increasing, 105–111
membrane permeabilization, 109, 110
pathway in eukaryotes, 129
secretion into medium, 108, 109
secretion into the periplasm, 105–108
secretion via Sec pathway, 105, 107–108
secretion via the Tat pathway, 108, 109

Protein secretion (*continued*)
strategy for recombinant protein to periplasm, 107
strategy to increase yields of mammalian cells, 152
strategy to increase yields of secreted recombinant proteins from mammalian cells, 153
type III secretion system, 109–111
type II secretion system, 109
yields of secreted recombinant proteins, 105
Protein stability
decreasing protein degradation, 103–105
disulfide bonds, 102–103
facilitating protein folding, 101–103
fusion proteins, 101–102
increasing, 100–105, 164–166
molecular chaperones, 101
solubility of recombinant protein, 101
Protein therapeutics, 243–244, 318; *see also* Enzymes; Pharmaceuticals; Recombinant antibodies
Proteobacteria, 200
Proteomics, 70–84
identification of proteins, 71–74
protein expression profiling, 74–77
protein-protein interactions, 78–84
Pseudomonas spp.. 681
degradative pathways and sizes of plasmids, 498
degrading xenobiotics, 496
exotoxin, 291–292
indigo, 467
Pseudomonas aeruginosa
alginate excretion, 255
alginate lyase, 257
antibacterial therapy, 357
bacteriophages and, 353–354
Pseudomonas alcaligenes, 17
Pseudomonas chlororaphis, protein IPDO72Aa, 649
Pseudomonas diminuta MG, pesticide degradation, 503
Pseudomonas fluorescens, 91
chromosome of, 236
Pseudomonas maltophilia, dicamba, 673
Pseudomonas oleovorans, polyhydroxyalkanoates, 492
Pseudomonas putida, L-citrulline, 472
Public health officials
antivaccine campaigns and, 830
vaccination, 387–389
Public Health Service Act, 828, 830, 832
Public Patent Foundation, 842
Pyruvate carboxylase, *Rhizobium etli*, 560, 561

Q

Quantitative polymerase chain reaction (PCR), 44–45

R

Rabbits
human protease inhibitor production, 770
models for cardiovascular disease, 761–763
Rabies, 437
Ralstonia eutropha, yield of secreted recombinant protein by, 105
Random mutagenesis, 161–163; *see also* Mutagenesis
affinity of antibody for, 168
decreasing protease sensitivity to protease inhibitor, 172
Raphanus sativus, 677
Raw starch-degrading enzymes (RSDEs), 513, 514
Receptor-associated protein (RAP), 165
binding affinity of LRP1 for mutant, 166
directed mutagenesis of, 166
Recognition sequences, restriction endonucleases, 14
Recombinant antibodies, 280–315; *see also* Antibodies
antibodies against various diseases, 309–313
antibody fragments, 289–294
anticancer antibodies, 302–309
anti-HIV antibodies, 300–302
antiobesity, 313–315
bispecific antibodies against hemophilia, 300
chimeric, 285
combinatorial libraries of antibody fragments, 294–296
combinatorial libraries of full-length antibodies, 297
dual-variable-domain antibodies, 298, 300
enhanced antibody half-life, 315
generation of XenoMouse, 288
genetically engineered antibodies, 285
humanized antibody, 285
human monoclonal antibodies, 287–289
hybrid human-mouse monoclonal antibodies, 284, 286–287
hybridoma technology, 280–281
monoclonal antibodies for human use in US or EU, 282–283
shuffling CDR sequences, 298, 299
trastuzumab for breast cancer treatment, 281
Recombinant DNA Molecule Program Advisory Committee (RAC), NIH, 801
Recombinant DNA research, development of guidelines for, 800–801
Recombinant DNA technology, 3, 6, 10, 243
cleavage of DNA, 14
potential biohazards of, 800
regulating foodstuffs, 806–807
restriction endonucleases, 459–461
scientific foundation, 6
small biological molecules, 461–495
types of, 799
Recombinant human insulin, 244
Recombinant human proteins, production of, 92
Recombinant mammalian cells, strategy to increase yields of, 151
Recombinant products, baculovirus-insect cell expression, 138
Recombinant proteins, 91; *see also* Prokaryotic hosts
correct folding of, 101
fungal expression systems, 131–137
integrating DNA into host chromosome, 114–119
production in prokaryotic hosts, 91–119
Recombinant therapeutic proteins
antithrombin in goat milk, 768–769
donor organs in pigs, 773–774
human protease inhibitor in rabbits, 770
mammary glands of transgenic animals, 768
production in chicken eggs, 771–773
Recombinant vaccines. *See also* Vaccines
limited commercialization of, 392
Recombinational cloning, 30–32
Recombination system (Cre-*loxP*), conditional gene modification with, 747–749, 750, 751
Red carotenoid pigment, lycopene production, 473
Refugia-like strategy, *B. thuringiensis* resistance, 652, 653, 654
Regulations, genetically modified microorganisms, 803–804
Reovirus mRNA, 340
Reporter genes, use in transformed plant cells, 610–611
Restriction endonucleases, 459–461
isolation of, 17
preparation of DNA for cloning, 13–18
protection of DNA from digestion, 460
recognition sequences of, 14
strategy for isolating genes for, 461
type II, 13, 16
Retroviral RNA, reverse transcription of, 331
Retroviral vector method
production of packaged retroviral vector RNA, 737, 738
transgene introduction into animal cells, 737
transgenic animals, 736–737
Retroviridae
gammaretrovirus of, 358
lentiviruses of, 359–360
Retrovirus infections, antisense RNA and, 330–331
Reversible chain terminators, sequencing using, 53–55

Rheumatoid arthritis, 7, 244
cytokines for, 310–311
diagnosing by indirect ELISA, 192
Rhizobium etli, pyruvate carboxylase, 560, 561
Rhizoctonia solani, resistance to, 677
Rhizopus, 136
Rhizosecretion, schematic of plant in hydroponic culture, 622
Rhodosporidium toruloides, phenylalanine ammonia lyase from, 259
Ribocil, chemical structure, 483
Riboflavin (vitamin B_2), 483–484
chemical structure, 483
modulation of, 484
Ribonuclease, virus resistance, 667–668
Ribosome-binding site, translation, 97–98
Ribozymes
delivery of, 339
DNAzymes and, 338–341
hairpin, 339
hammerhead, 339
inhibiting expression of viral genes, 340
schematic of, binding to mRNA, 340
two-dimensional representation of, 339
Rice
clonal seeds from hybrid, 619–621
fixation of heterozygosity by editing multiple genes, 620
genetically modified crops, 594
glyphosate, 670
vitamin A, 696–697
Rickettsia rickettsii, vaccines, 441
Rituximab (Rituxan), 244
non-Hodgkin's lymphoma, 284
RNA (ribonucleic acid). *See also* Antisense RNA; Targeting mRNA and DNA sequences
development of RNA vaccines, 586
interfering RNA, 341–347
RNA biomarkers of disease, 226–232
antibiotic resistance of bacteria, 228–230
breast cancer, 227–228
detecting associated changes in gene expression, 227–228
detection of microRNA (miRNA) signatures, 230–232
prostate cancer, 227
RNA interference (RNAi), 342
fungal and bacterium resistance, 681
gene knockdown by, 754–756
plant protection against insects, 650–652
protecting plants from insects, 650–652
RNA sequencing
high-throughput, 66, 67
single-cell, 68, 69
transcriptomics, 65–68
Rocky Mountain spotted fever, vaccines, 441
Roundup, glyphosate, 669–672
Roundup Ready corn, 813
Rubella, 387
annual cases in Canada, 388
diagnostic monoclonal antibodies, 185
Ruminococcus flavefaciens, designer cellulosome, 528
Ruminococcus gnavus, 200

S

Saccharomyces cerevisiae, 4, 79
alcohol production vs. *Z. mobilis*, 531
cellulase enzymes, 527
codons optimized for, 171
endoglucanase and exoglucanase genes, 526–527
engineering tolerance of glucose and ethanol, 516
genetically engineered, 236
genome sequence of, 47
glucoamylase, 514, 515
glucoamylase cDNA in, 511, 512
growth of, 174
lycopene production, 473, 474
PCR cloning and expression in, 169
protein folding, 129–130
protein trafficking, 130–131
secretion of heterologous proteins by, 127–131
signal peptides, 127–129
spores of, 219
strain to detect low levels of estrogenic substances, 237
transmembrane SNARE proteins, 132
Saccharomyces cerevisiae expression vectors, 124–127
episomal vectors, 124–126
integrating vectors, 126
integration of DNA into, 127
promoters for, 125
recombinant proteins by, 124
yeast artificial chromosomes (YAC), 126, 128
Saccharomyces diastaticus, amylolytic yeast, 512
Saccharomycopsis fibuligera, β-glucosidase gene, 527
Saccharopolyspora erythraea
erythromycin, 480
erythromycin production, 554
Safety
aptamers in clinical trials, 335–336
genetic testing, 832–834
medical products of biotechnology, 827–837
Salmon. *See also* Atlantic salmon; Chinook salmon
Atlantic, 732, 787
Chinook, 732, 787
enhancing growth of, 786–787
genetically engineered for human consumption, 732
Salmonella strains
bacteria as antigen delivery systems, 444–446
deleted genes and functions for development of attenuated, 429
engineered attenuated vaccine for, 428–429
genotypes of genetically attenuated, 429
nonviral delivery, 369, 370
Salmonella enterica
quantitative real-time PCR, 45
type III secretion system, 110
Salmonella enterica serovar Typhi
DNA vaccines, 418
hybridization to identify, 199
Salt and drought stress, 682–686
delaying drought-induced senescence, 685–686
genetic engineering of plants, 682
increasing trehalose production, 683–684
sequestering sodium ions, 684–685
Sanger, Frederick, 49, 51
Sanger sequencing, 50–52, 53
SARS (severe acute respiratory syndrome)
SARS coronavirus virus (SARS-CoV), 398
SARS-CoV virus-like particle, 143
schematic of human body affected by SARS-CoV–2, 400
spikes of SARS-CoV, 398, 399
subunit and peptide vaccines, 397–399
SARS-CoV-2, 56
development of RNA vaccines, 586
ELISA quantifying, neutralizing antibodies, 194, 195
RNA virus, 216
spike protein, 710
Schizosaccharomyces pombe, ribonuclease, 667
Sclerotinia sclerotiorum, 681
Selectable markers, mammalian expression vectors, 148, 150
Self-elimination gene drive, 794
Self-splicing proteins, 115
Sequencing technologies
assembly of genome, 60–61
comparison of, 54
dideoxynucleotide, 50–52, 53
DNA (deoxyribonucleic acid), 48–56
metagenomes, 61–62
nanopore, 56
preparation of genomic DNA sequencing libraries, 57–58
single-molecule real-time sequencing, 55–56
using reversible chain terminators, 53–55
whole genomes, 56–62
Serratia fonticola, 17
Sheep
clone Dolly, 743–744
one-cell ovine embryos, 786
protein composition of milk, 732
somatic cell nuclear transfer, 745
Shigella, DNA vaccines, 419–420
Shigella/*Escherichia*, 200
Shotgun cloning approach, 57

Shuffling CDR sequences, 298, 299
Signal peptides, proteins of *Saccharomyces cerevisiae*, 127–129
Signal sequence, 107
Single-chain antibodies, virus resistance, 663–664
Single-chain antibody fragment (scFv)
  cell growth and protein secretion, 558
  synthesis in quiescent *E. coli* cells, 557
Single-chain Fab (scFab) fragment, 373
Single-molecule real-time sequencing, 55–56
Single nucleotide polymorphism (SNP), 199
  allele-specific PCR detecting, 208
  ancestry determination, 223–225
  animal species determination, 225
  competitive oligopriming to detect, 209
  PCR amplification refractory mutation system, 207–209
  TaqMan PCR for genotyping, 209–211
Sinorhizobium meliloti, 803
Site-directed mutagenesis. *See also* Mutagenesis
  by inverse PCR, 158–159
  by overlap extension PCR, 156–158
Site-specific recombination, removal of selectable markers by, 118–119
Sleeping Beauty, gene therapy delivery vector, 375–376
Small biological molecules, 461–495; *see also* Amino acids; Antibiotics; Biopolymers
  amino acids, 468–473
  antibiotics, 474–486
  L-ascorbic acid, 463–466
  biopolymers, 487–493
  compounds metabolically engineered in microbes, 463
  indigo, 467–468
  lycopene, 473–474
  solvent tolerance, 493–494
  systems metabolic engineering optimizing product yield, 494–495
Smallpox, 387, 389
Snowdrop plant (*Galanthus nivalis*), lectins, 648, 658
Sodium bisulfite, converting cytosine to uracil, 222
Sodium ions, sequestering, for salt tolerance, 684–685
Solvent(s)
  industrial uses, 494
  tolerance of bacteria, 493–494
Somatic cell gene therapy, 326, 376, 382; *see also* Gene therapy
Somatic cell nuclear transfer, for transgenic livestock, 743–744
Somatotropin
  bovine milk, 805–806
  human growth hormone, 248–250
Sontheimer, Erik, 37
Southern University of Science and Technology, 832
Soybean(s), genetically modified oil, 694
Soybean mosaic virus (SMV), ribonuclease, 667–668
Spatial refugia, *B. thuringiensis* resistance, 652
Spike glycoprotein, 143
Spinal muscular atrophy, antisense RNA and, 329
Spodoptera frugiperda, 138, 143, 144
Spodoptera littoralis (African cotton leafworm), resistance, 658
Stabilizing and antirepressor (STAR) elements, 154
Stanford University, 2, 3, 845
Staphylococcus aureus
  antibodies, 297
  bacteria as antigen delivery system, 446–448
  detection of methicillin-resistant, (MRSA), 211–213
  development of vaccine against, 401
  fermentation in single stirred-tank reactor, 568
  nonviral delivery, 371
  SARS-CoV spikes, 398
  single guide RNA (sgRNA)-Cas system, 350
  transgenic cows, 778, 779
  vaccines for, 401–402
Staphylococcus carnosus, yield of secreted recombinant protein by, 105
Staphylococcus simulans, transgenic cows, 778–779
Staphyloxanthin
  biosynthesis of, 482
  Staphylococcus aureus, 481
Starch
  food plant taste and appearance, 703–705
  reactions in biosynthesis of, 703
Starch and sugars
  alcohol tolerance, 515–517
  amylopectin hydrolysis, 509
  amylose hydrolysis, 509
  commercial production of fructose and alcohol, 508–510
  fermentation of soluble starch, 512
  improving fructose production, 517–518
  increasing alcohol production, 510–517
  industrial production of fructose and alcohol from starch, 510
  schematic of protocol optimizing raw starch-degrading enzyme (RSDE), 513, 514
  utilization of, 508–518
  xylose/glucose isomerase, 517–518
StarLink Cry9C gene, 818
Stem cell method. *See* Engineered embryonic stem cell method
Stirred-tank reactors (STRs), 561, 562, 563–564
  two-stage fermentation in single, 568–569
Streptococcus infections
  flow chart for identification of group A, 407
  group A, 406, 407
  group B, 405–406
  lactic acid bacteria, 268
  vaccines for, 405–406, 407
Streptococcus sp. strain G148, fermentation in single stirred-tank reactor, 568
Streptococcus agalactiae
  bacteriophage lysin protein, 609
  infection, 607
  vaccines for group B, 405–406
Streptococcus mutans, dental caries, 423–424
Streptococcus pyogenes
  CRISPR-Cas as diagnostic, 218
  CRISPR-Cas9 system, 349, 350
  endonuclease Cas9, 37
  infection, 608
  vaccines for group A, 406, 407
Streptococcus sobrinus, dental caries, 423–424
Streptomyces
  antibiotic biosynthesis genes, 475–477
  antibiotics, 474
  genetic manipulation of, 475
  schematic of DNA transformation and selection of transformants of, 476
Streptomyces coelicolor A3, 477, 478
Streptomyces lividans, yield of secreted recombinant protein by, 105
Streptomyces rimosus, oxytetracycline, 486
Streptomyces venezuelae, oxytetracycline, 486, 487
Streptomycin, 352
Strimvelis, 831
Strokes, 7
Substantial equivalence, 812
Subtilisin
  development of modified, 169
  genetic engineering of calcium-independent, 170
Subunit vaccines, 392; *see also* Vaccines
  bovine herpes virus-1, 394–395
  cholera, 396
  COVID-19, 399–401
  foot-and-mouth disease virus (FMDV), 404–405
  herpes simplex virus (HSV), 393–394, 395
  human papillomavirus (HPV), 402–404
  influenza, 396–397
  lactic acid bacteria for testing in mice, 412
  Nanopatch, 412, 413
  SARS (severe acute respiratory syndrome), 397–399
  Staphylococcus aureus, 401–402
  Streptococcus infections, 405–406, 407
  usefulness of, 414
Sugars. *See* Starch and sugars
Suicide gene therapy, 378
Sulfonylureas, herbicide resistance, 669

Superbug, 498
Swanson, Robert, 7
Switchgrass (*Panicum virgatum*), 530
  lignin content, 720, 721
Synechocystis sp., flow chart of fatty acids from, 534
Syngenta, 824
Synthetic genomes, 47
α-Synuclein, 168, 168–169
Systems metabolic engineering
  flow chart of, 495
  optimizing product yield, 494–495

T

T4 polynucleotide kinase, 17
T7 expression system, regulation, 96, 97
TaqMan PCR protocol, 209–211
Targeting mRNA and DNA sequences. *See also* Antisense RNA; Aptamers; Interfering RNA (ribonucleic acid)
  antisense RNA, 327–331
  aptamers, 331–338
  clinical trials for antisense RNAs, 329
  comparing aptamers and monoclonal antibodies, 332
  CRISPR-Cas system, 349–351
  engineering bacteriophages, 352–357
  interfering RNA, 341–347
  nanoparticles, 352
  nanozymes, 351–352
  ribozymes and DNAzymes, 338–341
  SELEX (systematic evolution of ligands by exponential enrichment) procedure, 332–333
  spinal muscular atrophy and antisense RNA, 329
  zinc finger nucleases, 348–349
Taste. *See* Food plant taste and appearance
Taxane, 828
Tectaria macrodonta, proteins and, 648
Tenecteplase, triple-mutant tPA, 167
Tetanus, 389
  bacteria as antigen delivery system, 445–446
Tetracenomycin, biosynthesis of, 480
Tetracycline, resistance, 501, 502
Tetracycline-inducible system
  tet-off and tet-on systems, 750–752
  transgene expression control, 749–753
Tetrahydrofolate
  structure of, 699
  synthesis pathway, 699
  vitamin B9, 697–698
β-Thalassemia, first approved gene therapy for, 377
Therapeutic agents. *See also* Bacteria and therapeutics
  bacteria and, 267–279
  engineering plants synthesizing, 709
  gene use as, 325–327
  nucleic acids, 382
Therapeutic antibodies, potential, from plants, 593
Therapeutic drug development, clinical trials, 247
Therapeutic monoclonal antibodies
  approved for human use in US or EU, 282–283
  treating non-Hodgkin's lymphoma, 284
Therapeutic proteins
  production in chicken eggs, 771–773
  transgenic animals, 811
Thermoascus aurantiacus, engineered microbial enzymes, 529
Thermobifida fusca
  designer cellulosome, 528
  engineered microbial enzymes, 529
Thermus aquaticus, denaturation, 39
Thermus thermophilus
  α-amylase and glucose isomerase genes, 705
  xylose/glucose isomerase, 518
Ti plasmid
  Agrobacterium tumefaciens, 595–597, 633
  binary cloning vector, 600
  binary cloning vectors for viral coat protein, 660
  biosynthesis of auxin and cytokinin, 598
  cointegrate cloning vector, 600
  increasing transformation efficiency, 601–603
  plant transformation with, of *A. tumefaciens*, 595–597
  schematic of, 595
  transferred DNA (T-DNA), 595–597
  vector systems, 597–603
Tissue plasminogen activator (tPA), 167, 839
Tobacco budworm (*Heliothis virescens*), 645
Tobacco crown galls, octopine synthase from, 596
Tobacco hornworms (*Manduca sexta*), 643
Tobacco mosaic virus
  expression system, 707, 708
  inhibition by single-chain antibody, 663
Tobacco plants, transgenic, 608, 609
Tol2, gene therapy delivery vector, 375
Toluene- and xylene-degrading pathway of plasmid pWWO, microbial degradation, 499–503
Tomato bush stunt virus, inhibition by single-chain antibody, 664
Tomato fruitworms (*Heliothis zea*), 643
Tomato pinworms (*Keiferia lycopersicella*), 643
Toxic Substances Control Act (TSCA), Chemical Substance Inventory, 803
Toxin-intein fusions, 264–265
Toxins, potential for introducing into food, 816–819
Trametes multicolor, engineered microbial enzymes, 529
Transcription
  glucose and lactose concentration in growth medium, 95
  from *lac* promoter, 95–96
  promoters for expression of cloned genes, 93
  regulating, 92–96
Transcription activator-like effector nucleases (TALENs), 785, 808
Transcriptional activator of transcription (TAT) protein, 267
Transcription signals, mammalian cell expression systems, 146, 147
Transcriptomics, 65–70
  DNA microarrays, 68–70
  RNA sequencing, 65–68, 69
Transformed plants cells, use of reporter genes in, 610–611
Transgene (cargo gene), 788
Transgene expression, control with tetracycline-inducible system, 749–753
Transgenic animals, 731–733, 795; *see also* Animal bioreactors; Food animals; Gene drives
  animal bioreactors for production of recombinant therapeutic proteins, 767–774
  conditional gene modification with Cre-*loxP* recombination system, 747–749, 750, 751
  control of transgene expression with tetracycline-inducible system, 749–753
  DNA microinjection method, 733–735
  engineered embryonic stem cell method, 737, 739–743
  gene knockdown by RNAS interference, 754–756
  genome editing with CRISPR-Cas system, 744–747
  methodologies, 733–756
  milk quality improvements, 781–782
  models of human diseases, 756–767
  mouse model of Duchenne muscular dystrophy, 759–760, 761
  mouse models of Alzheimer's disease, 756–759
  nonhuman primate models of neurodevelopmental disorders, 766–767
  rabbit models of cardiovascular disease, 761–763
  recombinant therapeutic proteins in mammary glands of, 768
  retroviral vector method, 736–737, 738
  somatic cell nuclear transfer for transgenic livestock, 743–744, 745
  steps in development of transgenic livestock, 743
  zebrafish melanoma model, 763–765, 766
Transgenic canola (*Brassica napus*), 823
Transgenic chickens, production of therapeutic proteins in eggs, 771–773
Transgenic corn, glyphosate-tolerant, 813
Transgenic cotton plants, *B. thuringiensis* insecticidal protoxin, 654

Transgenic livestock. *See also* Transgenic animals
cloned sheep Dolly, 743–744
CRISPR-Cas system, 745–747
somatic cell nuclear transfer for, 743–744, 745
steps in development of, 743
Transgenic mice. *See also* Mouse; Transgenic animals
Cre-*loxP* recombination system, 747–749, 750, 751
DNA microinjection method, 733, 734
establishing with genetically engineered embryonic stem cells, 739, 740
generating increased muscle mass, 784
model for Huntington's disease, 752–753
models of human diseases, 795
production of packaged retroviral vector RNA, 737, 738
retroviral transfer vector for introducing transgenes, 737
retroviral vectors for establishing, 736
Transgenic mosquitoes, dengue fever virus-resistant, 791–792
Transgenic nonhuman mammals, 843
Transgenic pigs, production of donor organs, 773–774
Transgenic plants,633; *See also* Food plant taste and appearance; Insect resistance; Plant(s); Plant nutritional content
antibody and antibody fragments, 593
commercially available, 638
development, 591–592, 594–595
enhancing nutritional value, 815
fruits and flowers, 688–690
fungus and bacterium resistance, 674–681
gene stacking, 624–625
genetically modified crops, 594
herbicide resistance, 668–674
insect resistance, 637–658
marker gene excision system for transformed chloroplasts, 632
modification of food plant taste and appearance, 701–705
modification of plant nutritional content, 690–701
Nobel laureates supporting, 698
phytoremediation, 686–688
potential vaccine antigens, 593
production of marker-free, 626–632
regulation of, 809–810
removing marker genes from chloroplast DNA, 632
removing marker genes from nuclear DNA, 627–630
salt and drought stress, 682–686
schematic of DNA construct for generation marker-free, 629
schematic of T-DNA-based selectable marker gene excision system, 629
tobacco, 608, 609
viral coat protein-mediated protection, 659
virus resistance, 658–668
Transgenic zebrafish, melanoma model, 763–765, 766
Translation, expression vector pET, 98
Translation efficiency
genetic code and codon usage in humans and *E. coli*, 99
increasing, 96–100
optimizing codon usage, 98–100
ribosome-binding site, 97–98
Translation signals, mammalian cell expression systems, 147–148
Transposons
gene therapy delivery vectors, 375–376
movement of target gene into chromosomal DNA, 375
nonviral delivery system, 374–376
relative efficiencies of transposition systems, 375
Sleeping Beauty, 375–376
Trastuzumab, 828
first monoclonal antibody for breast cancer treatment, 281
Herceptin, 244
Trehalose
salt and drought stress, 683–684
structure of, 683
Triazines, herbicide resistance, 669
Tricarboxylic acid (TCA) cycle, 151, 494, 560
Trichoderma, 136
Trichoderma reesei
engineered microbial enzymes, 529
β-glucosidase gene, 527
Triparental conjugation, plasmid transfer, 26, 27
Tripartite fusion protein
schematic representation of, 253
two-stage fermentation in single STR, 568–569
Triple-mutant tPA, tenecteplase, 167
Trispecific antibody, 309
Tryptophan, indigo biosynthesis from, 468
Tuberculosis, 388
bacillus Calmette-Guérin (BCG) vaccine, 442–444
cattle resistant to, 776–778
Mycobacterium bovis, 442
Mycobacterium tuberculosis, 441–444
schematic of development of subunit vaccine for, 443
vaccines directed against bacteria, 441–444
Tumor markers, targets for monoclonal antibodies, 185
Tumor necrosis factor alpha (TNF-α), three-dimensional representation of, 251
Turnip crinkle virus, inhibition by single-chain antibody, 664
Turnip mosaic virus, micro-RNAs, 665
Turnip yellow mosaic virus, micro-RNAs, 665
Tuskegee Study, 830
Twin-arginine translocase (Tat) pathway, secretion, 108, 109
Two-dimensional polyacrylamide gel electrophoresis (2D PAGE), identification of proteins, 71, 72
Twort, Frederick, 352
Two-step homologous recombination, chromosomal integration by, 117, 118
Type I diabetes, 194, 276
Type III secretion system (T3SS), schematic on surface of Gram-negative bacteria, 447
Typhoid fever, 437
Tyrosine phosphorylation, detection of, 75, 77

U

Ulcerative colitis, 268, 835
interleukin-10, 270–272
Undecylprodigiosin
biosynthesis of, 477
chemical structure of, 478
United Nations Food and Agriculture Organization, 594, 811
United Nations Food and Agriculture Program, 697, 698
United States
human papillomavirus (HPV) vaccine, 404
monoclonal antibodies for human use, 282–283
recombinant proteins for human use, 244
University of Alicante, 37
University of California, 840, 845
University of California at Berkeley, 37
University of California at San Francisco, 3, 8
University of Vienna, 37
Uromyces appendiculatus, 681
U.S. Centers for Disease Control and Prevention (CDC), 389
human papillomavirus (HPV) vaccine, 404
U.S. Department of Agriculture (USDA), 808
U.S. Environmental Protection Agency, 496, 803
U.S. Food and Drug Administration, 167, 244, 284, 688, 732
antithrombin in goat's milk, 768
Center for Veterinary Medicine, 810, 814
human esterase inhibitor protein, 770
human gene therapy, 801
new animal drugs by, 811
papillomavirus types, 403
U.S. Genetic Information Nondiscrimination Act, 834
U.S. National Institutes of Health (NIH), 800, 801
U.S. Office of Technology Assessment, 9
U.S. Orphan Drug Act, 836
U.S. Patent and Trademark Office (USPTO), 697, 839, 841

U.S. Patent Statute, 843
U.S. Plant Patent Act, 844
U.S. Supreme Court, 839, 842, 843, 846

V

Vaccination, 387–389
Vaccines, 5, 452; *see also* Edible vaccines; Engineered attenuated vaccines; Subunit vaccines; Vector vaccines
  annual cases of diseases in Canada before and after, 388
  antivaccine movement, 830
  current and future, 389–392
  delivery, 411–414
  development of RNA, 586
  edible, 711–716
  engineered attenuated, 424–430
  herpes simplex virus 1 and 2 (HSV-1 and HSV-2), 393–394
  illegitimate drugs, 837
  inactivated forms of, 390
  injection techniques, 412–413
  limitations of traditional mode of production, 389–390
  limited commercialization of recombinant, 392
  live-attenuated, 390
  Nanopatch, 412, 413
  novel strategies for development, 390–391
  peptides, 407
  plant-based oral, and oral booster, 713
  plants as reactors, 709–710
  potential antigens from plants, 593
  schematic of, against human or animal viruses, 391
  schematic of types, 391
  subunit and peptide, 392–414
  vaccination with, 387–389
  vector, 430–441
Vaccinia virus
  antibiotic resistance genes in, 432–433
  basis of resistance of, 436
  characterization of, 432
  construction of K3L-negative mutant of, 436
  live recombinant viral vaccine, 435
  method for integration into, 433
  modified vaccinia Ankara virus (MAV), 438
  portion of transfer vector, 434
  portion of wild-type, 434
  sensitivity to interferon, 435, 437
  use of vector vaccines, 434–435
  vector vaccine for, 432–438
  veterinary vaccines, 436–437
L-Valine
  engineering *E. coli* to overproduce, 470–471
  overview of production of, 471
Varicella-zoster virus, 437
Vascular endothelial growth factor (VEGF)
  aptamer targeting, 334–335
  schematic of, 335
VAXCHORA vaccine, cholera, 428
Vector design, mammalian cell expression systems, 145–148
Vector-transmitted diseases
  dengue fever virus-resistant mosquitoes, 791–792
  gene drives eradicating, 787–794
  malaria population suppression, 789–791
  reversal drives, 792–793
Vector vaccines. *See also* Vaccines
  bacteria as antigen delivery system, 444–449
  dengue virus, 438, 439
  Ebola virus, 439–441
  modified vaccinia Ankara virus (MAV), 438
  tuberculosis, 441–444
  vaccines directed against bacteria, 441–444
  vaccines directed against viruses, 430, 432–441
  vaccinia virus, 432–438
Vegetative insecticidal toxins, plant protection against insects, 647–648
Verticillium dahliae, 681
Veterinary Medicine Advisory Committee, 806
Vibrio cholerae
  antibacterial therapy, 357
  DNA vaccines, 418
  edible vaccine for, 712–714
  engineered attenuated vaccine for, 426–428
  toxin-intein infusions, 264, 265
  vaccines for, 396
Vibrio fischeri, bioluminescent bacterium, 236
"Village Blacksmith" (Longfellow), 679
Viral delivery systems, 357–365; *see also* Nonviral delivery systems
  adeno-associated virus (AAV), 362–365
  adenoviruses, 360
  clinical trials involving AAV vectors, 364
  delivering foreign DNA into human cells, 358
  herpes simplex virus 1 (HSV-1), 361–362
  lentiviruses, 359–360
  schematic of life cycle of AAV, 363
  schematic of mature gammaretrovirus virion, 358
Viral infection, detection of, 193
Virus resistance
  CRISPR/Cas9 technology, 665–667
  engineered transgenic plants, 659
  micro-RNAs, 664–665
  protection by expression of genes, 663–668
  ribonuclease, 667–668
  single-chain antibodies, 663–664
  viral coat protein-mediated protection, 658–662
Vitamin(s)
  biosynthesis of β-carotene in rice and vitamin A in humans, 696
  folate (tetrahydrofolate, vitamin B9), 697–698
  plant nutritional content, 695–698
  vitamin A, 696–697
Vitamin A
  Golden Rice, 814–815
  plant nutrition, 696–697
Vitamin C. *See* L-Ascorbic acid
Vitreoscilla sp.
  bioproduct synthesis by, 554
  expression of hemoglobin gene, 553–554
  gene encoding hemoglobin-like molecule, 546
  oxygen and plant yield, 722–723

W

Wambugu, Florence, 594
Waste utilization, 5
Watermelon mosaic virus 2, viral coat protein, 661, 662
Weight reduction, leptin and, 272–274
Western corn rootworm (*Diabrotica virgifera virgifera*), protecting plants from insects, 649, 650
West Nile virus, 414, 422
Wet milling, disrupting microbial cells, 577–578
Whey
  carbon source for microorganisms, 489
  L-citrulline and, 495
  waste of cheese-making process, 488–489
Whitefly (*Bemisia tabaci*), proteins and, 648
Whole-genome sequencing, assessing genetic disease risk, 225–226
Whooping cough, 388, 389
Wine disease, Pierce disease, 680–681
World Health Organization, 353, 811, 830
  cholera cases, 396
  dengue fever, 438
  iron deficiency, 698
  malaria, 408
  pesticide poisoning, 503
World Intellectual Property Organization, 839
World Trade Organization, genetically engineered products, 827

X

Xanthan gum, 487–489
  production of, 489
  structure of, 488
Xanthomonas, 681
Xanthomonas campestris, xanthan gum, 487–489
Xanthomonas citri subsp. *citri*, 681
Xanthomonas oryzae pv. *oryzae*, 681
Xenobiotics. *See also* Microbial degradation of xenobiotics
  microbial degradation of, 496–508
  plastics, 507–508

XenoMouse, generation of, 288
Xenos, 496
XTEN fusion proteins, 252, 253, 315
Xylans, schematic of structure of, 521
Xylella fastidiosa, Pierce disease, 680–681
D-Xylose, chemical structure, 532
Xylose, engineered assimilation of, by *Z. mobilis*, 533–534
Xylose/glucose isomerase, fructose production, 517–518

Y

Yarrowia lipolytica, 134, 136
  lycopene production, 473
Yeast(s)
  fermentation of soluble starch, 512
  N-linked oligosaccharides, 122
  O-linked oligosaccharides, 121
  Pichia pastoris, 132–134
  protein folding in endoplasmic reticulum of, 131
Yeast artificial chromosomes (YAC), 126, 128, 289
  foreign DNA cloned in, 604, 633
Yeast expression systems, 123–137
  Saccharomyces cerevisiae expression vectors, 124–127
Yellow fever, 422
Yellow fever virus, 811
Yescarta, 831

Z

Zebrafish (*Danio rerio*)
  melanoma model, 763–765, 766
  transgenesis, 764
Zeocin, 154
Zika virus, 616, 811
  DNA vaccines, 422–423
  monoclonal antibody against, 615–616
  schematic of surface protein, 422
  transmission of, 193
Zinc finger nucleases, 348–349, 617
  genetic repair, 348–349
  β-lactoglobulin gene, 782
  plants, 808
Zolgensma, 829, 830, 831, 835
Zucchini yellow mosaic virus
  CRISPR/Cas9, 667
  viral coat protein, 661, 662
Zygosaccharomyces rouxii, marker gene expression, 630
Zymomonas-E. coli shuttle vector, alcohol production, 532
Zymomonas mobilis
  alcohol production by, 531–534, 538
  schematic of engineered assimilation, 533, 534
Zynteglo, 377, 831

WILEY END USER LICENSE AGREEMENT

Go to www.wiley.com/go/eula to access Wiley's ebook EULA.