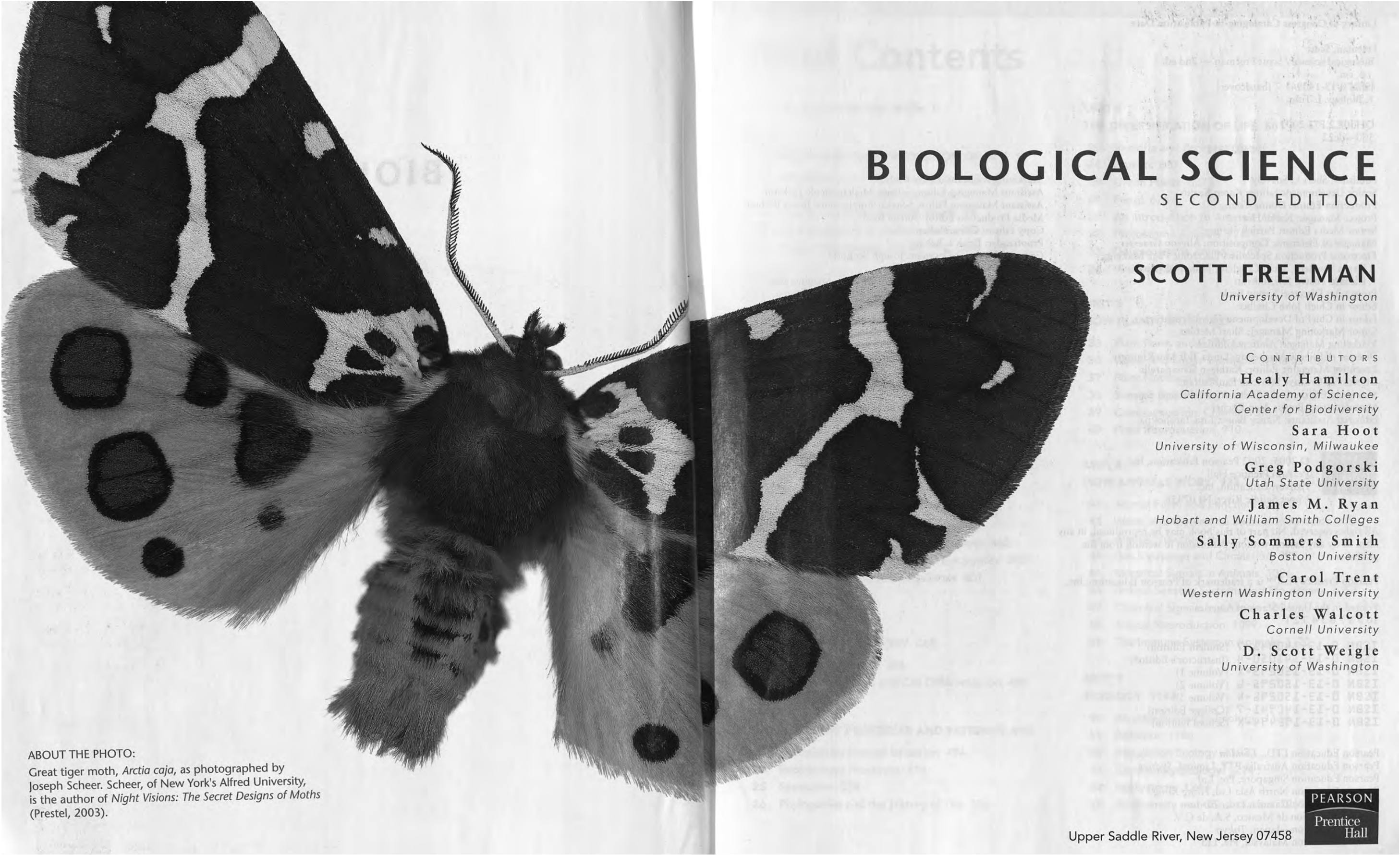


Ex. 1043 Freeman 2005



ABOUT THE PHOTO:
Great tiger moth, *Arctia caja*, as photographed by Joseph Scheer. Scheer, of New York's Alfred University, is the author of *Night Visions: The Secret Designs of Moths* (Prestel, 2003).

Library of Congress Cataloging-in-Publication Data

Freeman, Scott
Biological science / Scott Freeman.— 2nd ed.
p. cm.
ISBN 0-13-140941-7 (hardcover)
1. Biology. I. Title.

QH308.2.F73 2005
570—dc22

2004027312

Publisher: Sheri L. Snavelly
Senior Development Editor: Karen Karlin
Production Editor: Donna Young
Project Manager: Karen Horton
Senior Media Editor: Patrick Shriner
Manager of Electronic Composition: Allyson Graesser
Electronic Production Specialists/Electronic Page Makeup:
Karen Stephens, Vicki Croghan, Julita Nazario,
Richard Foster, Jim Sullivan
Project Art Director: Kenny Beck
Editor in Chief: John Challice
Editor in Chief of Development: Carol Trueheart
Senior Marketing Manager: Shari Meffert
Marketing Manager: Andrew Gilfillan
Director of Science Marketing: Linda Taft MacKinnon
Executive Managing Editor: Kathleen Schiaparelli
Director of Creative Services: Paul Belfanti
Managing Editor, Audio/Video Assets: Patricia Burns
Manufacturing Buyer: Alan Fischer
Editorial Assistants: Nancy Bauer, Lisa Tarabokija

Marketing Assistant: Laura Rath
Assistant Managing Editor, Science Media: Nicole Jackson
Assistant Managing Editor, Science Supplements: Becca Richter
Media Production Editor: Aaron Reid
Copy Editor: Chris Thillen
Proofreader: Brian I. Baker
Cover and Interior Designer: Joseph Sengotta
AV Editor: Connie Long
Illustrators: Pearson Artworks, Quade Paul, Imagineering
National Sales Director for Key Markets: David Theisen
Photo Researcher: Yvonne Gerin
Director, Image Resource Center: Melinda Reo
Manager, Rights and Permissions: Zina Arabia
Manager, Visual Research: Beth Boyd-Brenzel
Manager, Cover Visual Research and Permissions: Karen
Sanatar
Image Permission Coordinator: LaShonda Morris
Cover photo: © Joseph Scheer
Other image credits appear in the backmatter.



© 2005, 2002 Pearson Education, Inc.
Pearson Prentice Hall
Pearson Education, Inc.
Upper Saddle River, NJ 07458

All rights reserved. No part of this book may be reproduced, in any form or by any means, without permission in writing from the publisher.

Pearson Prentice Hall™ is a trademark of Pearson Education, Inc.

Printed in the United States of America
10 9 8 7 6 5 4 3 2 1

ISBN 0-13-140941-7 (Student Edition)
ISBN 0-13-141050-4 (Instructor's Edition)
ISBN 0-13-150293-X (Volume 1)
ISBN 0-13-150295-6 (Volume 2)
ISBN 0-13-150296-4 (Volume 3)
ISBN 0-13-140941-7 (College Edition)
ISBN 0-13-192494-X (School Edition)

Pearson Education LTD., London
Pearson Education Australia PTY, Limited, Sydney
Pearson Education Singapore, Pte. Ltd
Pearson Education North Asia Ltd, Hong Kong
Pearson Education Canada, Ltd., Toronto
Pearson Educación de México, S.A. de C.V.
Pearson Education—Japan, Tokyo
Pearson Education Malaysia, Pte. Ltd

Brief Contents

1 Biology and the Tree of Life 1

UNIT 1

THE ORIGIN AND EARLY EVOLUTION OF LIFE 19

- 2 The Atoms and Molecules of Ancient Earth 20
- 3 Protein Structure and Function 46
- 4 Nucleic Acids and the RNA World 74
- 5 An Introduction to Carbohydrates 90
- 6 Lipids, Membranes, and the First Cells 103

UNIT 2

CELL STRUCTURE AND FUNCTION 127

- 7 Inside the Cell 128
- 8 Cell-Cell Interactions 159
- 9 Cellular Respiration and Fermentation 177
- 10 Photosynthesis 202
- 11 The Cell Cycle 227

UNIT 3

GENE STRUCTURE AND EXPRESSION 247

- 12 Meiosis 248
- 13 Mendel and the Gene 269
- 14 DNA Synthesis 303
- 15 How Genes Work 325
- 16 Transcription and Translation 338
- 17 Control of Gene Expression in Bacteria 362
- 18 Control of Gene Expression in Eukaryotes 382
- 19 Analyzing and Engineering Genes 401
- 20 Genomics 426

UNIT 4

DEVELOPMENTAL BIOLOGY 445

- 21 Early Development 446
- 22 Pattern Formation and Cell Differentiation 469

UNIT 5

EVOLUTIONARY PROCESSES AND PATTERNS 493

- 23 Evolution by Natural Selection 494
- 24 Evolutionary Processes 516
- 25 Speciation 538
- 26 Phylogenies and the History of Life 556

UNIT 6

THE DIVERSIFICATION OF LIFE 581

- 27 Bacteria and Archaea 582
- 28 Protists 607
- 29 Green Plants 637
- 30 Fungi 674
- 31 An Introduction to Animals 698
- 32 Protostome Animals 724
- 33 Deuterostome Animals 749
- 34 Viruses 780

UNIT 7

HOW PLANTS WORK 803

- 35 Plant Form and Function 804
- 36 Water and Sugar Transport in Plants 828
- 37 Plant Nutrition 852
- 38 Sensory Systems in Plants 871
- 39 Communication: Chemical Signals 888
- 40 Plant Reproduction 910

UNIT 8

HOW ANIMALS WORK 933

- 41 Animal Form and Function 934
- 42 Water and Electrolyte Balance in Animals 955
- 43 Animal Nutrition 977
- 44 Gas Exchange and Circulation 999
- 45 Electrical Signals in Animals 1026
- 46 Animal Sensory Systems and Movement 1052
- 47 Chemical Signals in Animals 1076
- 48 Animal Reproduction 1099
- 49 The Immune System in Animals 1120

UNIT 9

ECOLOGY 1143

- 50 An Introduction to Ecology 1144
- 51 Behavior 1166
- 52 Population Ecology 1192
- 53 Community Ecology 1214
- 54 Ecosystems 1243
- 55 Biodiversity and Conservation 1265

Contents

Chapter 1

Biology and the Tree of Life 1

- 1.1 The Cell Theory 2
 - Are All Organisms Made of Cells? 2
 - Where Do Cells Come From? 2
 - 1.2 The Theory of Evolution by Natural Selection 4
 - What Is Evolution? 4
 - What Is Natural Selection? 4
 - 1.3 The Tree of Life 5
 - Linnaean Taxonomy 6
 - BOX 1.1 Scientific Names and Terms 6
 - Using Molecules to Understand the Tree of Life 8
 - 1.4 Doing Biology 10
 - Why Do Giraffes Have Long Necks? An Introduction to Hypothesis Testing 10
 - Why Are Chili Peppers Hot? An Introduction to Experimental Design 12
- ESSAY Where Do Humans Fit on the Tree of Life? 15
- CHAPTER REVIEW 16

UNIT 1

The Origin and Early Evolution of Life 19

Chapter 2

The Atoms and Molecules of Ancient Earth 20

- 2.1 The Ancient Earth 20
 - Studying the Formation of Planets 21
 - When Did Chemical Evolution Take Place? 21
- 2.2 The Building Blocks of Chemical Evolution 24
 - What Atoms Are Found in Organisms? 24
 - How Does Covalent Bonding Hold Molecules Together? 24
 - How Does Ionic Bonding Hold Molecules Together? 26
 - Some Simple Molecules Formed from H, C, N, and O 27
- 2.3 Chemical Reactions, Chemical Evolution, and Chemical Energy 29
 - How Do Chemical Reactions Happen? 30
 - What Is Energy? 30
 - Chemical Evolution: A Model System 32
 - How Did Chemical Energy Change during Chemical Evolution? 34
- 2.4 The Composition of the Early Atmosphere: Redox Reactions and the Importance of Carbon 35

What Is a Redox Reaction? 35

BOX 2.1 Some Other Approaches to Understanding Redox Reactions 36

What Happens When Carbon Is Reduced? 36

- 2.5 The Early Oceans and the Properties of Water 39
 - Why Is Water Such an Efficient Solvent? 39
 - How Does Water's Structure Correlate with Its Properties? 40
 - Acid-Base Reactions and pH 41
 - What Was Water's Role in Chemical Evolution? 42

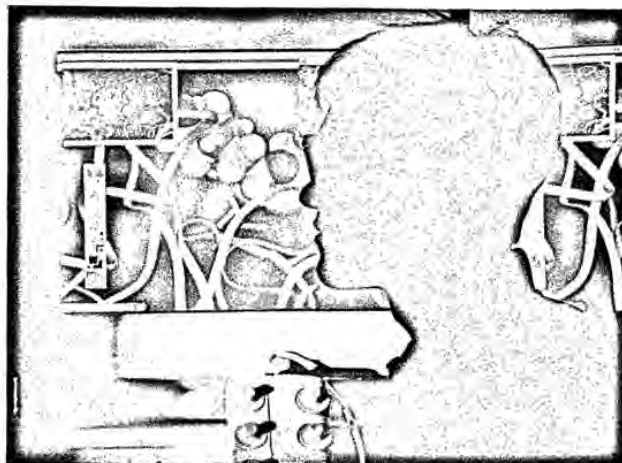
ESSAY The Search for Extraterrestrial Life 43

CHAPTER REVIEW 44

Chapter 3

Protein Structure and Function 46

- 3.1 Early Origin-of-Life Experiments 47
 - BOX 3.1 Was Miller Correct about Conditions on Ancient Earth? 48
- 3.2 Amino Acids and Polymerization 48
 - The Structure of Amino Acids 49
 - How Do Amino Acids Link to Form Proteins? 53
- 3.3 What Do Proteins Look Like? 56
 - Primary Structure 57
 - Secondary Structure 57
 - Tertiary Structure 58
 - Quaternary Structure 59
 - Folding and Function 60
 - BOX 3.2 Prions 62



xvi Contents

- 3.4 What Do Proteins Do? 62**
Proteins Have Diverse Functions in Cells 62
An Introduction to Catalysis 63
How Do Enzymes Work? 64
Was the First Living Entity a Protein? 70

ESSAY Molecular Handedness and the Thalidomide Tragedy 70

CHAPTER REVIEW 71

Chapter 4

Nucleic Acids and the RNA World 74

- 4.1 What Is a Nucleic Acid? 75**
Could Chemical Evolution Result in the Production of Nucleotides? 76
How Do Nucleotides Polymerize to Form Nucleic Acids? 76
BOX 4.1 Gel Electrophoresis and Autoradiography 78
- 4.2 DNA Structure and Function 79**
BOX 4.2 An Introduction to X-Ray Crystallography 80
DNA Is an Information-Containing Molecule 82
Is DNA a Catalytic Molecule? 82
- 4.3 RNA Structure and Function 83**
RNA as an Information-Containing Molecule 84
Is RNA a Catalytic Molecule? 85
- 4.4 The First Life-Form 85**

ESSAY The Human Side of Research 87

CHAPTER REVIEW 88

Chapter 5

An Introduction to Carbohydrates 90

- 5.1 Sugars as Monomers 91**
- 5.2 The Structure of Polysaccharides 92**
BOX 5.1 Lactose Intolerance and Galactosemia 93
Starch: A Storage Polysaccharide in Plants 94
Glycogen: A Highly Branched Storage Polysaccharide in Animals 95
Cellulose: A Structural Polysaccharide in Plants 95
Chitin: A Structural Polysaccharide in Animals 96
Peptidoglycan: A Structural Polysaccharide in Bacteria 96
Polysaccharides and Chemical Evolution 96
BOX 5.2 How Do the Penicillins and Cephalosporins Kill Bacteria? 96
- 5.3 What Do Carbohydrates Do? 97**
The Role of Carbohydrates in Cell Identity 97
The Role of Carbohydrates in Energy Production and Storage 98
Carbohydrates as Structural Molecules 99

ESSAY Why Do We Have a Sweet Tooth? 101

CHAPTER REVIEW 101

Chapter 6

Lipids, Membranes, and the First Cells 103

- 6.1 Lipids 104**
BOX 6.1 Electron Microscopy 104
What Is a Lipid? 106
A Look at Three Types of Lipids Found in Cells 106
The Structures of Membrane Lipids 108
- 6.2 Phospholipid Bilayers 109**
Artificial Membranes as an Experimental System 109
Selective Permeability of Lipid Bilayers 110
Does the Type of Lipid in a Membrane Affect Its Permeability? 111
Why Does Temperature Affect the Fluidity and Permeability of Membranes? 112
- 6.3 Why Molecules Move across Lipid Bilayers: Diffusion and Osmosis 113**
- 6.4 Membrane Proteins 116**
Systems for Studying Membrane Proteins 118
How Do Membrane Proteins Affect Ions and Molecules? 118

ESSAY The Molecular Basis of Cystic Fibrosis 123

CHAPTER REVIEW 125

UNIT 2

Cell Structure and Function 127

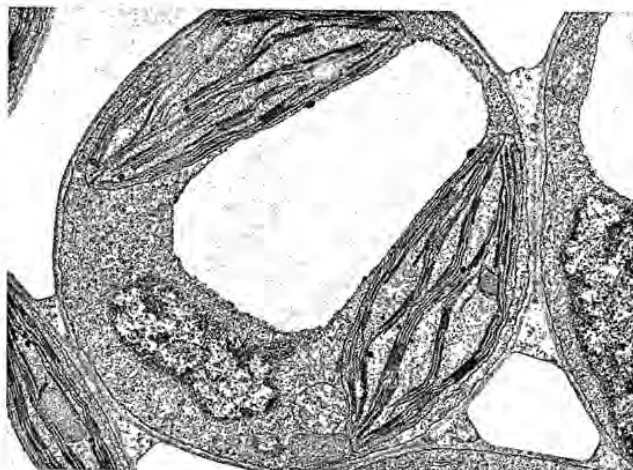
Chapter 7

Inside the Cell 128

- 7.1 What's Inside the Cell? 129**
Prokaryotic Cells 129
Eukaryotic Cells 131
How Does Cell Structure Correlate with Function? 138
The Dynamic Cell 139
BOX 7.1 How Does a Centrifuge Work? 140
BOX 7.2 Techniques for Studying the Dynamic Cell 141
- 7.2 The Nuclear Envelope: Transport Into and Out of the Nucleus 142**
How Are Molecules Imported into the Nucleus? 143
How Are Molecules Exported from the Nucleus? 144
- 7.3 The Endomembrane System: Manufacturing and Shipping Proteins 145**
Entering the Endomembrane System: The Signal Hypothesis 146
Getting from the ER to the Golgi 147
What Happens Inside the Golgi Apparatus? 147
How Are Products Shipped from the Golgi? 148
- 7.4 The Dynamic Cytoskeleton 149**
Actin Filaments 150
Intermediate Filaments 151
Microtubules 152
Cilia and Flagella: Moving the Entire Cell 154

ESSAY Organelles and Human Disease 156

CHAPTER REVIEW 156



Chapter 8

Cell-Cell Interactions 159

- 8.1 The Cell Surface 160
 - The Structure and Function of an Extracellular Layer 160
 - The Plant Cell Wall 160
 - The Extracellular Matrix in Animals 162
 - BOX 8.1 What Happens When the Extracellular Matrix Is Defective? 163
- 8.2 How Do Adjacent Cells Connect and Communicate? 163
 - Cell-Cell Attachments 165
 - Cell-Cell Gaps 169
- 8.3 How Do Distant Cells Communicate? 170
 - Signal Reception 170
 - Signal Processing 171
 - Signal Response 172
 - Signal Deactivation 174

ESSAY How Do Viagra®, Levitra®, and Cialis® Work? 175

CHAPTER REVIEW 174

Chapter 9

Cellular Respiration and Fermentation 177

- 9.1 An Overview of Cellular Respiration 178
 - The Nature of Chemical Energy and Redox Reactions 179
 - Using Redox Reactions to Produce ATP 180
 - Processing Glucose: Glycolysis 180
 - The Krebs Cycle 180
 - Electron Transport 182
 - Methods of Producing ATP 182
- 9.2 Glycolysis 183
 - A Closer Look at the Glycolytic Reactions 183
 - How Is Glycolysis Regulated? 184
- 9.3 The Krebs Cycle 186
 - Converting Pyruvate to Acetyl CoA 187
 - How Is the Krebs Cycle Regulated? 188
 - What Happens to the NADH and FADH₂? 190

- 9.4 Electron Transport and Chemiosmosis 190
 - Components of the Electron Transport Chain 191
 - The Chemiosmotic Hypothesis 192
 - How Is the Electron Transport Chain Organized? 193
 - The Discovery of ATP Synthase 194
 - Oxidative Phosphorylation 194
- 9.5 Fermentation 196
- 9.6 How Does Cellular Respiration Interact with Other Metabolic Pathways? 197
 - Processing Proteins and Fats as Fuel 197
 - Anabolic Pathways Synthesize Key Molecules 198

ESSAY ATP Production during Exercise 199

CHAPTER REVIEW 199

Chapter 10

Photosynthesis 202

- 10.1 An Overview of Photosynthesis 203
 - Photosynthesis: Two Distinct Sets of Reactions 203
 - The Structure of the Chloroplast 204
 - BOX 10.1 Types of Plastids 205
- 10.2 How Does Chlorophyll Capture Light Energy? 206
 - Photosynthetic Pigments Absorb Light 206
 - BOX 10.2 How Do Researchers Measure Absorption Spectra? 207
 - When Light Is Absorbed, Electrons Enter an Excited State 209
- 10.3 The Discovery of Photosystems I and II 211
 - How Does Photosystem II Work? 212
 - How Does Photosystem I Work? 214
 - The Z Scheme: Photosystems I and II Work Together 215
- 10.4 How Is Carbon Dioxide Reduced to Produce Glucose? 217
 - The Calvin Cycle 217
 - The Discovery of Rubisco 219
 - How Is Carbon Dioxide Delivered to Rubisco? 220
 - What Happens to the Sugar That Is Produced by Photosynthesis? 222

ESSAY Are Rising CO₂ Levels in the Atmosphere Affecting the Rate of Photosynthesis? 224

CHAPTER REVIEW 225

Chapter 11

The Cell Cycle 227

- 11.1 Mitosis and the Cell Cycle 228
 - The Cell Cycle 229
 - When Does Chromosome Replication Occur? 229
 - Discovery of the Gap Phases 230
 - BOX 11.1 Cell-Culture Methods 231
 - BOX 11.2 How Do Bacteria Divide? 232
- 11.2 How Does Mitosis Take Place? 232
 - Events in Mitosis 232
 - Cytokinesis 235

xviii Contents

- How Do Chromosomes Move during Mitosis? 237
- 11.3 Control of the Cell Cycle 238
 - The Discovery of Cell-Cycle Regulatory Molecules 238
 - Cell-Cycle Checkpoints 240
- 11.4 Cancer: Out-of-Control Cell Division 241
 - Properties of Cancer Cells 242
 - Cancer Involves Loss of Cell-Cycle Control 242
- ESSAY Cancer Chemotherapy 244
- CHAPTER REVIEW 245

UNIT 3

Gene Structure and Expression 247

Chapter 12

Meiosis 248

- 12.1 How Does Meiosis Occur? 249
 - BOX 12.1 Karyotyping Techniques 250
 - An Overview of Meiosis 252
 - The Phases of Meiosis I 253
 - The Phases of Meiosis II 255
 - A Closer Look at Key Events in Prophase of Meiosis I 257
- 12.2 The Consequences of Meiosis 258
 - Chromosomes and Heredity 258
 - How Does the Separation and Distribution of Homologous Chromosomes Produce Genetic Variation? 258
 - The Role of Crossing Over 259
 - How Does Fertilization Affect Genetic Variation? 259
- 12.3 Why Does Meiosis Exist? Why Sex? 260
 - The Paradox of Sex 261
 - When Do Meiosis and Fertilization Occur during the Life of an Organism? 263
- 12.4 Mistakes in Meiosis 264
 - How Do Mistakes Occur? 264
 - Why Do Mistakes Occur? 265
- ESSAY Seedless Fruits 266
- CHAPTER REVIEW 267

Chapter 13

Mendel and the Gene 269

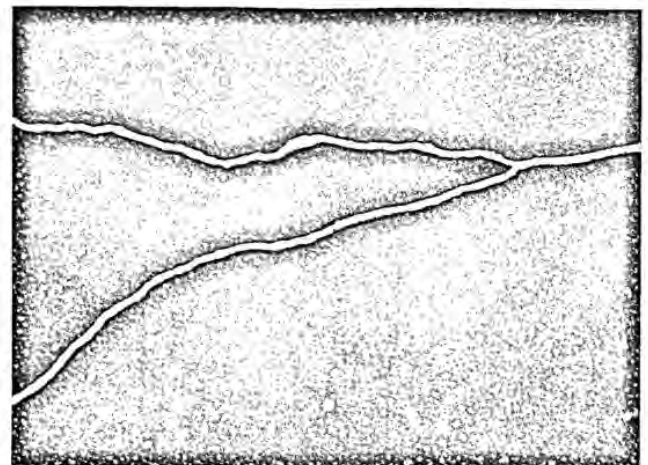
- 13.1 Mendel's Experiments with a Single Trait 270
 - What Questions Was Mendel Trying to Answer? 270
 - Garden Peas Serve as the First Model Organism in Genetics 270
 - Inheritance of a Single Trait 272
 - The Nature and Behavior of the Hereditary Determinants 273
 - BOX 13.1 Why Do Punnett Squares Work? 275
 - Testing the Model 275
- 13.2 Mendel's Experiments with Two Traits 276
 - BOX 13.2 Sample Size and Chance Fluctuations 278

- Using a Testcross to Confirm Predictions 278
- Mendel's Contributions to the Study of Heredity 279
- 13.3 The Chromosome Theory of Inheritance 279
- 13.4 Testing and Extending the Chromosome Theory 281
 - The Discovery of Sex-Linked Traits 281
 - The Discovery of Sex Chromosomes 281
 - X-Linked Inheritance and the Chromosome Theory 282
 - What Happens When Genes Are Located on the Same Chromosome? 283
- 13.5 Extending Mendel's Rules 287
 - Incomplete Dominance and Codominance 288
 - Multiple Alleles and Polymorphic Traits 289
 - Pleiotropy 289
 - Genes Are Affected by the Physical Environment and Genetic Environment 290
 - Organelle Genomes 291
 - Quantitative Traits 291
- 13.6 Applying Mendel's Rules to Humans 294
 - Are Alleles Recessive or Dominant? 294
 - Is the Trait Autosomal or Sex-Linked? 295
- ESSAY Does "Genetic Determinism" Exist? 297
- CHAPTER REVIEW 298

Chapter 14

DNA Synthesis 303

- 14.1 DNA as the Hereditary Material 304
 - Is DNA the Genetic Material? 305
- 14.2 Testing Early Hypotheses about DNA Replication 307
 - The Meselson-Stahl Experiment 308
 - The Discovery of DNA Polymerase 310
- 14.3 A Comprehensive Model for DNA Synthesis 311
 - Opening the Helix 312
 - Synthesis of the Leading Strand 313
 - Synthesis of the Lagging Strand 313
- 14.4 Replicating the Ends of Linear Chromosomes 316



- 14.5 Repairing Mistakes in DNA Synthesis 318
 - How Is DNA Polymerase Proofread? 318
 - Mismatch Repair 319
- 14.6 Repairing Damaged DNA 320
 - DNA Nucleotide Excision Repair: An Overview 320
 - Xeroderma Pigmentosum: A Case Study 321
- ESSAY DNA Mismatch Repair and Cancer 322
- CHAPTER REVIEW 323

Chapter 15

How Genes Work 325

- 15.1 What Do Genes Do? 326
 - The Molecular Basis of Hereditary Diseases 326
 - Genes and Enzymes 326
- 15.2 The Genetic Code 328
 - RNA as the Intermediary between Genes and Proteins 329
 - How Long Is a Word in the Genetic Code? 330
 - How Did Researchers Crack the Code? 332
 - BOX 15.1 The Evolution of the Genetic Code 333
- 15.3 The Central Dogma of Molecular Biology 334
- ESSAY Exceptions to the Central Dogma 335
- CHAPTER REVIEW 336

Chapter 16

Transcription and Translation 338

- 16.1 Transcription in Bacteria 339
 - RNA Polymerase Structure and Function 340
 - Initiation: How Does Transcription Begin in Bacteria? 340
 - Elongation and Termination 341
- 16.2 Transcription in Eukaryotes 342
 - The Startling Discovery of Eukaryotic Genes in Pieces 342
 - BOX 16.1 Toxins and Transcription 343
 - Exons, Introns, and RNA Splicing 344
 - Other Aspects of Transcript Processing: Caps and Tails 345
- 16.3 An Introduction to Translation 345
 - Ribosomes Are the Site of Protein Synthesis 346
 - How Does an mRNA Triplet Specify an Amino Acid? 347
- 16.4 The Role of Transfer RNA 347
 - Connecting Structure with Function 348
 - How Many tRNAs Are There? 349
- 16.5 Ribosomes and the Mechanism of Translation 350
 - Initiation 351
 - Elongation 352
 - Termination 353
- 16.6 The Molecular Basis of Mutation 355
- ESSAY Antibiotics That Poison the Ribosome 358
- CHAPTER REVIEW 359

Chapter 17

Control of Gene Expression in Bacteria 362

- 17.1 Gene Regulation and Information Flow 363
 - Mechanisms of Regulation—an Overview 363
 - Metabolizing Lactose—A Model System 364
- 17.2 Identifying the Genes Involved in Lactose Metabolism 365
 - Screening Mutants—Replica Plating and Indicator Plates 365
 - Different Classes of Lactose Metabolism Mutants 366
 - Several Genes Are Involved in Metabolizing Lactose 367
- 17.3 The Discovery of the Repressor 368
 - The *lac* Operon 369
 - Testing the *lac* Operon Model 369
 - BOX 17.1 Almost Parallel Worlds: Similarities and Contrasts in Control of the *trp* and *lac* Operons 370
 - The Impact of the *lac* Operon Model 370
- 17.4 Catabolite Repression and Positive Control 371
 - How Does Glucose Influence Formation of the CAP-cAMP Complex? 372
- 17.5 The Operator and the Repressor—an Introduction to DNA-Binding Proteins 374
 - Finding the Operator 374
 - BOX 17.2 DNA Footprinting 374
 - DNA Binding via the Helix-Turn-Helix Motif 376
 - How Does the Inducer Change the Repressor's Affinity for DNA? 378
 - Future Directions 378
- ESSAY Control of Gene Expression in Bacterial Pathogens 378
- CHAPTER REVIEW 380

Chapter 18

Control of Gene Expression in Eukaryotes 382

- 18.1 Mechanisms of Gene Regulation—an Overview 383
- 18.2 Eukaryotic DNA and the Regulation of Gene Expression 383
 - Chromatin Structure 384
 - Evidence that Chromatin Structure Is Altered in Active Genes 384
 - How Is Chromatin Altered? 385
- 18.3 Regulatory Sequences and Regulatory Proteins 386
 - Promoter-Proximal Elements 386
 - Enhancers 386
 - How Do Enhancers Work? 388
- 18.4 Transcription Initiation 390
- 18.5 Post-Transcriptional Control 392
 - Alternative Splicing of mRNAs 392
 - Translational Control 393
 - Post-Translational Control 393
 - A 50-Year Retrospective: How Does Gene Expression in Bacteria and in Eukaryotes Compare? 394

xx Contents

- 18.6 Linking Cancer with Defects in Gene Regulation 395
 p53—Guardian of the Genome 395
 STAT Mutations in Cancer 397
ESSAY Fly Eyes and Other Strange Tales of Gene Regulation 398
CHAPTER REVIEW 399

Chapter 19

Analyzing and Engineering Genes 401

- 19.1 Using Recombinant DNA Techniques to Manufacture Proteins: The Effort to Cure Pituitary Dwarfism 402
 Why Did Early Efforts to Treat the Disease Fail? 402
 Using Recombinant DNA Technology to Produce a Safe Supply of Growth Hormone 402
 BOX 19.1 How Do Plasmids Move between Cells Naturally? 404
 Ethical Concerns over Recombinant Growth Hormone 407
19.2 Analyzing DNA: Did Our Ancestors Mate with Neanderthals? 407
 The Polymerase Chain Reaction (PCR) 408
 Dideoxy Sequencing 409
19.3 Gene Hunting Based on Pedigree Analysis: The Huntington's Disease Story 411
 How Was the Huntington's Disease Gene Found? 411
 BOX 19.2 Southern Blotting 412
 Genetic Testing 416
 Ethical Concerns over Genetic Testing 417
19.4 Can Gene Therapy Cure Inherited Diseases in Humans? Research on Severe Immune Disorders 417
 How Can Novel Alleles Be Introduced into Human Cells? 417
 Using Gene Therapy to Treat X-Linked Immune Deficiency 418
 Ethical Concerns over Gene Therapy 420
19.5 Biotechnology in Agriculture: The Development of Golden Rice 420
 Rice as a Target Crop 421
 Synthesizing β -Carotene in Rice 421
 The *Agrobacterium* Transformation System 421
 Golden Rice 422
ESSAY Controversies over Genetically Modified Foods 423
CHAPTER REVIEW 424

Chapter 20

Genomics 426

- 20.1 Whole-Genome Sequencing 426
 Recent Technological Advances 427
 Which Sequences Are Genes? 428
 BOX 20.1 Shotgun Sequencing and Bioinformatics 428
20.2 Bacterial and Archaeal Genomes 431
 The Natural History of Prokaryotic Genomes 431

Evidence for Lateral Gene Transfer 432

- 20.3 Eukaryotic Genomes 433
 Natural History: Types of Sequences 433
 What Is a Gene Family? 436
20.4 Comparative Genomics 437
 Understanding Virulence 437
 Studying Adaptation 437
 Gene Number and Alternative Splicing 438
 Sequence Similarity and the Importance of Regulatory Sequences 439
20.5 Future Prospects 439
 Functional Genomics and Proteomics 439
 Medical Implications 441
ESSAY Genomics and Genetic Discrimination 442
CHAPTER REVIEW 443

UNIT 4

Developmental Biology 445

Chapter 21

Early Development 446

- 21.1 Gametogenesis 447
 Sperm Structure and Function 447
 Egg Structure and Function 448
21.2 Fertilization 449
 Fertilization in Sea Urchins 449
 Fertilization in Mammals 452
 Fertilization in Flowering Plants 452
21.3 Cleavage and Gastrulation 453
 Controlling Cleavage Patterns 454
 Activating the Zygotic Genome 456
 BOX 21.1 Visualizing mRNAs by in Situ Hybridization 457
 Gastrulation 458
 Organizing the Major Body Axes 458
 The Spemann-Mangold Experiments 459
21.4 Patterns of Development 461
 Embryonic Development in *Arabidopsis thaliana* 462
 An Invertebrate Model: *Drosophila melanogaster* 463
 A Vertebrate Model: The Frog 464
 Early Development in Humans 464
 Making Sense of Developmental Variation and Similarity 465
ESSAY Treating Human Infertility 466
CHAPTER REVIEW 467

Chapter 22

Pattern Formation and Cell Differentiation 469

- 22.1 Pattern Formation in *Drosophila* 470
 The Discovery of *bicoid* 470
 BOX 22.1 Maternal Effect Inheritance 471
 The Discovery of Segmentation Genes 473

- The Discovery of Homeotic Genes 476
- Drosophila* as a Model Organism 477
- 22.2 Pattern Formation in *Arabidopsis* 478
 - The Root-to-Shoot Axis of Embryos 478
 - Flower Development in Adults 479
- 22.3 Does the Genetic Makeup of Cells Change as Development Proceeds? 481
 - Are Differentiated Animal Cells Genetically Equivalent? 481
 - BOX 22.2 Regulatory Evolution 482
 - BOX 22.3 Human Cloning? 484
- 22.4 Differentiation: Becoming a Specialized Cell 484
 - Organizing Mesoderm into Somites 485
 - BOX 22.4 Apoptosis: Programmed Cell Death 486
 - Determination of Muscle Cells 488
- ESSAY Human Stem Cells 489
- CHAPTER REVIEW 490

UNIT 5

Evolutionary Processes and Patterns 493

Chapter 23

Evolution by Natural Selection 494

- BOX 23.1 Why Darwin Gets Most of the Credit 495
- 23.1 The Pattern of Evolution: Have Species Changed through Time? 496
 - Evidence for Change through Time 496
 - Evidence That Species Are Related 498
 - How Do Biologists Distinguish Homology from Analogy? 501
 - Darwinism and the Pattern Component of Evolution 502
- 23.2 The Process of Evolution: How Does Natural Selection Work? 503
 - BOX 23.2 Evolutionary Theory before Darwin 503



- 23.3 Evolution in Action: Recent Research on Natural Selection 504
 - How Did *Mycobacterium tuberculosis* Become Resistant to Antibiotics? 504
 - Why Are Beak Size, Beak Shape, and Body Size Changing in Galápagos Finches? 506
 - Can Natural Selection Be Studied Experimentally? 508
- 23.4 The Nature of Natural Selection and Adaptation 510
 - Selection Acts on Individuals, but Evolutionary Change Occurs in Populations 510
 - Evolution Is Not Progressive 510
 - Not All Traits Are Adaptive 511
- ESSAY The Debate over "Scientific Creationism" 512
- CHAPTER REVIEW 513

Chapter 24

Evolutionary Processes 516

- 24.1 Analyzing Change in Allele Frequencies: The Hardy-Weinberg Principle 517
 - The Hardy-Weinberg Model Makes Important Assumptions 518
 - How Does the Hardy-Weinberg Principle Serve as a Null Hypothesis? 518
- 24.2 Natural Selection and Sexual Selection 520
 - Directional Selection 521
 - Stabilizing Selection 522
 - Disruptive Selection 522
 - Sexual Selection 523
- 24.3 Genetic Drift 527
 - Experimental Studies of Genetic Drift 528
 - Genetic Drift in Natural Populations 529
- 24.4 Gene Flow 530
- 24.5 Mutation 531
 - Mutation as an Evolutionary Mechanism 531
 - What Role Does Mutation Play in Evolutionary Change? 532
- 24.6 Inbreeding 533
- ESSAY Evolutionary Theory and Human Health 535
- CHAPTER REVIEW 536

Chapter 25

Speciation 538

- 25.1 Defining and Identifying Species 539
 - The Biological Species Concept 539
 - The Morphospecies Concept 539
 - The Phylogenetic Species Concept 540
 - Species Definitions in Action: The Case of the Dusky Seaside Sparrow 541
- 25.2 Isolation and Divergence in Allopatry 543
 - Dispersal and Colonization Isolate Populations 543
 - Vicariance Isolates Populations 544

xxii Contents

- 25.3 Isolation and Divergence in Sympatry 545
 - Can Natural Selection Cause Speciation Even When Gene Flow Is Possible? 546
 - How Can Polyploidy Lead to Speciation? 547
- 25.4 What Happens When Isolated Populations Come into Contact? 549
 - Reinforcement 549
 - Hybrid Zones 550
 - New Species through Hybridization 550
- ESSAY Human Races 553
- CHAPTER REVIEW 554

Chapter 26

Phylogenies and the History of Life 556

- 26.1 Tools for Studying History: Phylogenies and the Fossil Record 556
 - Using Phylogenies 557
 - Using the Fossil Record 561
 - BOX 26.1 The Molecular Clock 566
- 26.2 The Cambrian Explosion 566
 - Cambrian Fossils: An Overview 567
 - The Doushantuo Microfossils 568
 - The Ediacaran Faunas 569
 - The Burgess Shale Faunas 569
- 26.3 The Genetic Mechanisms of Change 569
 - Gene Duplications and the Cambrian Explosion 569
 - Changes in Gene Expression: The Origin of the Foot 571
- 26.4 Adaptive Radiations 573
 - Colonization Events as a Trigger 574
 - The Role of Morphological Innovation 575
- 26.5 Mass Extinctions 575
 - How Do Background and Mass Extinctions Differ? 576
 - What Killed the Dinosaurs? 576
- ESSAY Is a Mass Extinction Event Under Way Now? 578
- CHAPTER REVIEW 579

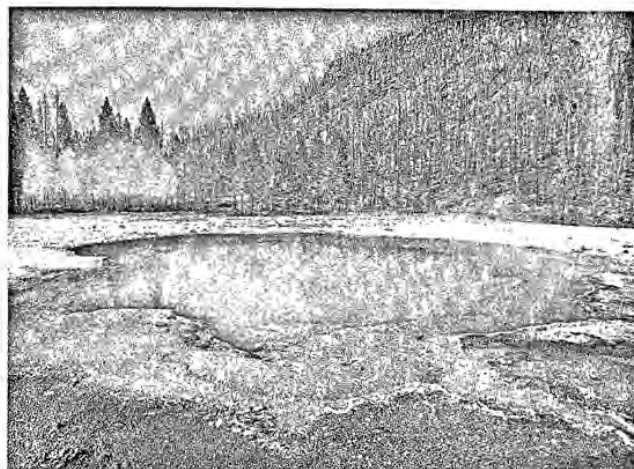
UNIT 6

The Diversification of Life 581

Chapter 27

Bacteria and Archaea 582

- 27.1 Why Do Biologists Study Bacteria and Archaea? 583
 - Bacterial Diseases 583
 - Bioremediation 584
 - Extremophiles 585
 - Global Change 586
- 27.2 How Do Biologists Study Bacteria and Archaea? 588
 - BOX 27.1 A Model Organism: *Escherichia coli* 588



- Using Enrichment Cultures 589
- Using Direct Sequencing 590
- Evaluating Molecular Phylogenies 590
- 27.3 What Themes Occur in the Diversification of Bacteria and Archaea? 592
 - Morphological Diversity 592
 - Metabolic Diversity 594
- 27.4 Key Lineages of Bacteria and Archaea 598
 - Bacteria 599
 - Spirochaetes (Spirochetes) 599
 - Chlamydiales 599
 - High-GC Gram Positives 600
 - Cyanobacteria 600
 - Low-GC Gram Positives 601
 - Proteobacteria 601
 - BOX 27.2 Lateral Gene Transfer and Metabolic Diversity in Bacteria 602
 - Archaea 602
 - Crenarchaeota 603
 - Euryarchaeota 603
- ESSAY Antibiotics and the Evolution of Drug Resistance 604
- CHAPTER REVIEW 605

Chapter 28

Protists 607

- 28.1 Why Do Biologists Study Protists? 608
 - Impacts on Human Health and Welfare 608
 - Ecological Importance of Protists 610
- 28.2 How Do Biologists Study Protists? 612
 - Microscopy: Studying Cell Structure 612
 - Evaluating Molecular Phylogenies 612
 - BOX 28.1 How Should We Name the Major Branches on the Tree of Life? 614
 - Combining Data from Microscopy and Phylogenies: Understanding the Origin of Mitochondria and Chloroplasts 615
 - Discovering New Lineages via Direct Sequencing 617
- 28.3 What Themes Occur in the Diversification of Protists? 617

BOX 28.2 A Model Organism: *Dictyostelium discoideum* 618

Morphological Diversity 619

How Do Protists Find Food? 621

How Do Protists Move? 623

How Do Protists Reproduce? 624

28.4 Key Lineages of Protists 627

Excavata 627

■ Excavata ► Diplomonadida 627

■ Excavata ► Parabasalida 628

Discristata 628

■ Discristata ► Euglenida 628

Alveolata 629

■ Alveolata ► Ciliata 629

■ Alveolata ► Dinoflagellata 629

■ Alveolata ► Apicomplexa 630

Stramenopila (Heterokonta) 630

■ Stramenopila ► Oomycota 630

■ Stramenopila ► Diatoms 631

■ Stramenopila ► Phaeophyta (Brown Algae) 631

Cercozoa 632

■ Cercozoa ► Foraminifera 632

Plantae 632

■ Rhodophyta (Red Algae) 633

Amoebozoa 633

■ Myxogastrea (Plasmodial Slime Molds) 633

ESSAY Revolutions in Science 634

CHAPTER REVIEW 634

Chapter 29

Green Plants 637

29.1 Why Do Biologists Study the Green Plants? 638

Plants Provide Ecosystem Services 638

Agricultural Research: Domestication and Selective Breeding 639

Plant-Based Fuels and Fibers 640

Bioprospecting 642

29.2 How Do Biologists Study Green Plants? 643

BOX 29.1 A Model Organism: *Arabidopsis thaliana* 643

Using the Fossil Record 645

Evaluating Molecular Phylogenies 647

29.3 What Themes Occur in the Diversification of Green Plants? 648

The Transition to Land, I: How Did Plants Adapt to Dry Conditions? 648

The Transition to Land, II: How Do Plants Reproduce in Dry Conditions? 651

The Angiosperm Radiation 658

29.4 Key Lineages of Green Plants 661

Green Algae 661

■ Ulvobionta 662

■ Coleochaetales 662

■ Charales (Stoneworts) 663

Nonvascular Plants (Bryophytes) 663

■ Hepaticophyta (Liverworts) 663

■ Anthocerophyta (Hornworts) 664

■ Bryophyta (Mosses) 664

Seedless Vascular Plants 665

■ Lycophyta (Lycophytes, or Club Mosses) 665

■ Psilotophyta (Whisk Ferns) 666

■ Sphenophyta (or Equisetophyta) (Horsetails) 666

■ Pteridophyta (Ferns) 667

Seed Plants 667

■ Gnetophyta (Gnetophytes) 668

■ Cycadophyta (Cycads) 668

■ Ginkgophyta (Ginkgoes) 669

■ Coniferophyta (Conifers) 669

■ Anthophyta (Angiosperms) 670

ESSAY Genetic Diversity in Crop Plants 671

CHAPTER REVIEW 672

Chapter 30

Fungi 674

30.1 Why Do Biologists Study Fungi? 675

Fungi Feed Land Plants 675

Fungi Speed the Carbon Cycle on Land 675

Fungi Have Important Economic Impacts 676

30.2 How Do Biologists Study Fungi? 677

Analyzing Morphological Traits 678

BOX 30.1 A Model Organism: *Saccharomyces cerevisiae* 679

Evaluating Molecular Phylogenies 680

Experimental Studies of Mutualism 682

30.3 What Themes Occur in the Diversification of Fungi? 684

Fungi Participate in Several Types of Mutualisms 684

Adaptations That Make Fungi Effective

Decomposers 686

Variation in Life Cycles 687

30.4 Key Lineages of Fungi 690

■ Chytridiomycota (Chytrids) 691

■ Zygomycota ► Mucorales and Other Basal Lineages 691

■ Glomeromycota (AMF) 692

■ Basidiomycota (Club Fungi) 693

Ascomycota (Sac Fungi) 693

■ Ascomycota ► Lichen-Formers 693

■ Ascomycota ► Non-Lichen-Formers 694

ESSAY Why Are Frogs Dying? 695

CHAPTER REVIEW 695

Chapter 31

An Introduction to Animals 698

31.1 Why Do Biologists Study Animals? 699

31.2 How Do Biologists Study Animals? 700

Analyzing Comparative Morphology 700

BOX 31.1 A Model Organism: *Hydra* 702

Using the Fossil Record 708

Evaluating Molecular Phylogenies 708

31.3 What Themes Occur in the Diversification of Animals? 710

Feeding 710
Movement 713
Reproduction and Life Cycles 715

- 31.4 Key Lineages of Animals 718
- Choanoflagellates (Collar Flagellates) 718
 - Porifera (Sponges) 719
 - Cnidaria (Jellyfish, Corals, Anemones, Hydroids, Sea Fans) 719
 - Ctenophora (Comb Jellies) 720
 - Acoelomorpha 721

ESSAY Coral Bleaching 721

CHAPTER REVIEW 722

Chapter 32

Protostome Animals 724

- 32.1 Why Do Biologists Study Protostomes? 725
- BOX 32.1 Model Organisms: *Caenorhabditis* and *Drosophila* 725
- Crustaceans and Mollusks Are Important Animals in Marine Ecosystems 726
- Insects, Spiders, and Mites Are Important Animals in Terrestrial Ecosystems 726
- 32.2 How Do Biologists Study Protostomes? 728
- Analyzing Morphological Traits 728
- Using the Fossil Record 730
- Evaluating Molecular Phylogenies 730
- 32.3 What Themes Occur in the Diversification of Protostomes? 731
- Feeding 732
- Movement 733
- Reproduction and Life Cycles 734
- 32.4 Key Lineages: Lophotrochozoans 734
- Rotifera (Rotifers) 735
 - Platyhelminthes (Flatworms) 736
 - Annelida (Segmented Worms) 737
 - Mollusca (Mollusks) 738
 - Mollusca ► Bivalvia (Clams, Mussels, Scallops, Oysters) 738
 - Mollusca ► Gastropoda (Snails, Slugs, Nudibranchs) 739
 - Mollusca ► Polyplacophora (Chitons) 739
 - Mollusca ► Cephalopoda (Squid, Nautilus, Octopuses) 740
- 32.5 Key Lineages: Ecdysozoans 740
- Nematoda (Roundworms) 741
 - Arthropoda (Arthropods) 742
 - Arthropoda ► Myriapods (Millipedes, Centipedes) 742
 - Arthropoda ► Chelicerata (Spiders, Ticks, Mites, Horseshoe Crabs, Daddy Longlegs, Scorpions) 743
 - Arthropoda ► Insecta (Insects) 744
 - Arthropoda ► Crustaceans (Shrimp, Lobster, Crabs, Barnacles, Isopods, Copepods) 744

ESSAY The Role of Natural History in Biological Science 745

CHAPTER REVIEW 746

Chapter 33

Deuterostome Animals 749

- 33.1 Why Do Biologists Study Deuterostome Animals? 750
- 33.2 How Do Biologists Study Deuterostomes? 751
- Analyzing Morphological Traits 751
- Using the Fossil Record 753
- Evaluating Molecular Phylogenies 755
- 33.3 What Themes Occur in the Diversification of Deuterostomes? 756
- Feeding 756
- BOX 33.1 Model Organisms: Sea Urchins, Zebrafish, and Mice 757
- Movement 758
- BOX 33.2 The Evolution of Flight in Birds 760
- Reproduction 761
- 33.4 Key Lineages: Echinodermata 763
- Echinodermata ► Asteroidea (Sea Stars) 764
 - Echinodermata ► Echinoidea (Sea Urchins and Sand Dollars) 764
- 33.5 Key Lineages: Chordata 765
- Chordata ► Myxinoidea (Hagfish) and Petromyzontoidea (Lampreys) 765
 - Chordata ► Chondrichthyes (Sharks, Rays, Skates) 766
 - Chordata ► Actinopterygii (Ray-Finned Fishes) 766
 - Chordata ► Actinistia (Coelacanth) and Dipnoi (Lungfish) 767
 - Chordata ► Amphibia (Frogs, Salamanders, Caecilians) 768
 - Chordata: Mammalia (Mammals) 768
 - Chordata ► Mammalia ► Monotremata (Platypuses, Echidnas) 769
 - Chordata ► Mammalia ► Marsupialia (Marsupials) 769
 - Chordata ► Mammalia ► Eutheria (Placental Mammals) 770
 - Chordata: Reptilia (Turtles, Snakes and Lizards, Crocodiles, Birds) 770
 - Chordata ► Reptilia ► Testudinia (Turtles, Tortoises) 771
 - Chordata ► Reptilia ► Lepidosauria (Lizards, Snakes) 771
 - Chordata ► Reptilia ► Crocodilia (Crocodiles, Alligators) 772
 - Chordata ► Reptilia ► Aves (Birds) 772
- 33.6 Key Lineages: The Hominin Radiation 773
- The Primates 773
- Fossil Humans 774
- The Out-of-Africa Hypothesis 776

ESSAY So Human an Animal 777

CHAPTER REVIEW 778

Chapter 34

Viruses 780

- 34.1 Why Do Biologists Study Viruses? 781
- Recent Viral Epidemics in Humans 781
- Current Viral Epidemics in Humans: HIV 782

- 34.2 How Do Biologists Study Viruses? 783**
 Analyzing Morphological Traits 784
 BOX 34.1 How Are Vaccines Developed? 785
 Analyzing the Phases of the Lytic Cycle 788
 BOX 34.2 HIV Drug Cocktails and the Evolution of Drug Resistance 792
- 34.3 What Themes Occur in the Diversification of Viruses? 793**
 The Nature of the Viral Genetic Material 793
 Where Did Viruses Come From? 794
 Emerging Viruses, Emerging Diseases 794
 BOX 34.3 Where Did HIV Originate? 795
- 34.4 Key Lineages of Viruses 796**
 ■ Double-Stranded DNA (dsDNA) Viruses 797
 ■ RNA Reverse-Transcribing Viruses (Retroviruses) 797
 ■ Double-Stranded RNA (dsRNA) Viruses 798
 ■ Negative-Sense Single-Stranded RNA ([−]ssRNA) Viruses 798
 ■ Positive-Sense Single-Stranded RNA ([+]ssRNA) Viruses 799
- ESSAY The SARS Outbreak 799
 CHAPTER REVIEW 800

UNIT 7

How Plants Work 803

Chapter 35

Plant Form and Function 804

- 35.1 The Diversity of Plant Form 805**
 The Diversity of Roots: North American Prairie Plants 806
 The Diversity of Shoots: Hawaiian Silverswords 807
- 35.2 Plant Cells and Tissues 811**
 The Diversity of Plant Cells 812
 BOX 35.1 Plant Poisons and the Cost of Defense 813
 The Diversity of Plant Tissues 815
- 35.3 The Anatomy of Plant Growth 818**
 Primary Growth: The Root System 819
 Primary Growth: The Shoot System 820
 Secondary Growth 821
 BOX 35.2 Tree-Ring Studies 824
- ESSAY Wood as a Structural Material 825
 CHAPTER REVIEW 825

Chapter 36

Water and Sugar Transport in Plants 828

- 36.1 Water Potential and Cell-to-Cell Movement 829**
 What Factors Affect Water Potential? 829
 Calculating Water Potential 830
 Water Potentials in Soils, Plants, and the Atmosphere 830
 BOX 36.1 How Do Plants Adapt to Dry Soils? 832
 BOX 36.2 How Do Plants Adapt to Salty Soils? 833



- 36.2 Root Pressure and Short-Distance Transport 834**
- 36.3 Transpiration and Long-Distance Water Transport 835**
 The Cohesion-Tension Theory 836
 Water Absorption and Water Loss 841
- 36.4 Translocation 842**
 The Anatomy of Phloem 843
 The Pressure-Flow Hypothesis 844
 Phloem Loading 845
 BOX 36.3 How Are Solutes Transported across Membranes? 845
 Phloem Unloading 848
- ESSAY Irrigated Agriculture 849
 CHAPTER REVIEW 849

Chapter 37

Plant Nutrition 852

- 37.1 Nutritional Requirements 853**
 Essential Nutrients 853
 Nutrient Deficiencies 855
- 37.2 Soil 856**
 Soil Conservation 856
 Nutrient Availability 858
- 37.3 Nutrient Uptake 859**
 Mechanisms of Nutrient Uptake 859
 Mechanisms of Ion Exclusion 861
 BOX 37.1 Can Phytoremediation Help Clean Up Contaminated Soils? 862
- 37.4 Nitrogen Fixation 863**
 How Do Nitrogen-Fixing Bacteria Colonize Plant Roots? 863
 How Do Host Plants Respond to Contact from a Symbiotic Bacterium? 863
- 37.5 Nutritional Adaptations of Plants 864**
 Epiphytic Plants 865
 Parasitic Plants 865

Carnivorous Plants 867
BOX 37.2 Why Are Bogs So Nutrient Poor? 867

ESSAY Tropical Soils 868

CHAPTER REVIEW 869

Chapter 38

Sensory Systems in Plants 871

38.1 Sensing Light 872

What Do Plants See? 872
How Do Plants See? 875
From Perception to Response: Signal Transduction 877

38.2 Gravity Perception 878

The Statolith Hypothesis 878
Is the Gravity Sensor a Transmembrane Protein? 879

38.3 How Do Plants Respond to Wind and Touch? 879

An Introduction to Electrical Signaling 880
Action Potentials 881
How Does the Venus Flytrap Close? 881

38.4 How Do Plants Sense Attacks by Parasites? 882

The Gene-for-Gene Hypothesis 882
Why Do So Many Resistance Genes and Alleles Exist? 883

ESSAY Can Plants Tell Time? 885

CHAPTER REVIEW 886

Chapter 39

Communication: Chemical Signals 888

39.1 Plant Hormones—An Overview 888

39.2 Auxin and Phototropism 891

The Cholodny-Went Hypothesis 891
Isolating and Characterizing Auxin 892
How Does Auxin Produce the Phototropic Response? 893

39.3 Auxin and Apical Dominance 894

Polar Transport of Auxin 895
An Overview of Auxin Action 896

39.4 Cytokinins and Cell Division 896

BOX 39.1 The Use of Hormones to Grow Plant Cells in Culture 897

39.5 Gibberellins and ABA: Growth and Dormancy 898

Gibberellins Stimulate Shoot Elongation 898
Gibberellins and ABA Interact during Seed Dormancy and Germination 899
ABA Closes Guard Cells in Stomata 901

39.6 Ethylene and Senescence 902

39.7 Salicylic Acid, Systemin, and Pheromones: Defense Signals 904

Salicylic Acid Extends the Hypersensitive Response to Pathogens 904
Systemin Leads to the Production of Insecticides 905
Pheromones Released from Wounds Recruit Help from Wasps 906

ESSAY Hormones as Herbicides 907

CHAPTER REVIEW 907

Chapter 40

Plant Reproduction 910

40.1 An Introduction to Plant Reproduction 911

Sexual Reproduction 911
Plant Life Cycles 912
Asexual Reproduction 912

40.2 Reproductive Structures 914

When Does Flowering Occur? 915
BOX 40.1 Is There a Flowering Hormone? 915
The General Structure of the Flower 916
Producing the Female Gametophyte 917
Producing the Male Gametophyte 918

40.3 Pollination and Fertilization 919

Pollination 919
BOX 40.2 Selfing, Outcrossing, and Self-Incompatibility 920
Fertilization 922

40.4 The Seed 923

Embryogenesis 923
The Role of Drying in Seed Maturation 924
BOX 40.3 Pure and Applied Science 924
Fruit Development and Seed Dispersal 925
Seed Dormancy 925
Seed Germination 927

ESSAY Why Do Wasps Try to Mate with Hammer Orchids? 929

CHAPTER REVIEW 930

UNIT 8

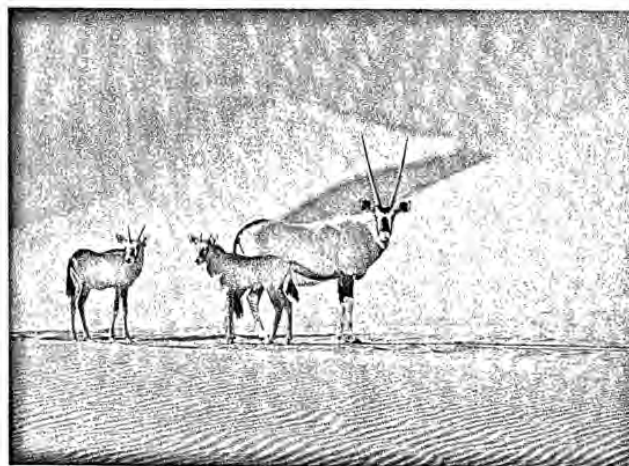
How Animals Work 933

Chapter 41

Animal Form and Function 934

41.1 Form, Function, and Adaptation 935

Trade-Offs 935
Adaptation and Acclimatization 937



- 41.2 Tissues, Organs, and Systems: How Does Structure Correlate with Function? 937
 - Tissues 937
 - Organs and Systems 939
- 41.3 How Does Body Size Affect Animal Physiology? 941
 - Surface Area/Volume Relationships 941
 - Do All Aspects of an Animal's Body Increase in Size Proportionately? 944
- 41.4 Homeostasis 945
 - Regulation and Feedback 946
- 41.5 How Do Animals Regulate Body Temperature? 946
 - Gaining and Losing Heat 947
 - BOX 41.1 A Closer Look at Torpor and Hibernation 949
 - Ectothermy Versus Endothermy 951
- ESSAY Is Fever Adaptive? 952
- CHAPTER REVIEW 952

Chapter 42

Water and Electrolyte Balance in Animals 955

- 42.1 Osmoregulation and Osmotic Stress 956
- 42.2 Water and Electrolyte Balance in Aquatic Environments 958
 - How Do Sharks Osmoregulate? 958
 - BOX 42.1 Passive and Active Transport: A Quick Review 960
 - How Do Salmon Osmoregulate? 960
- 42.3 Water and Electrolyte Balance in Terrestrial Insects 962
 - How Do Insects Minimize Water Loss from the Body Surface? 962
 - Types of Nitrogenous Wastes: Impact on Water Balance 963
 - Maintaining Homeostasis: The Excretory System 963
- 42.4 Water and Electrolyte Balance in Terrestrial Vertebrates 965
 - The Structure of the Kidney 966
 - Filtration: The Renal Corpuscle 966
 - Reabsorption: The Proximal Tubule 968
 - Creating an Osmotic Gradient: The Loop of Henle 969
 - Regulating Water and Electrolyte Balance: The Distal Tubule and the Collecting Duct 972
- ESSAY Life in the Desert 973
- CHAPTER REVIEW 975

Chapter 43

Animal Nutrition 977

- 43.1 Nutritional Requirements 978
 - Meeting Basic Needs 978
 - Nutrition and Athletic Performance 979
- 43.2 Obtaining Food: The Structure and Function of Mouthparts 981
- 43.3 How Are Nutrients Digested and Absorbed? 984
 - The Mouth and Esophagus 986

- The Stomach 986
- The Small Intestine 989
- The Large Intestine 993

- 43.4 Nutritional Homeostasis—Glucose as a Case Study 993
 - Insulin's Role in Homeostasis 994
 - The Type II Diabetes Mellitus Epidemic 995
- ESSAY Cholesterol, Diet, and Disease 996
- CHAPTER REVIEW 997

Chapter 44

Gas Exchange and Circulation 999

- 44.1 Air and Water as Respiratory Media 999
 - How Do Oxygen and Carbon Dioxide Behave in Air? 1000
 - How Do Oxygen and Carbon Dioxide Behave in Water? 1001
- 44.2 Organs of Gas Exchange 1002
 - Design Parameters: The Law of Diffusion 1002
 - How Do Gills Work? 1003
 - How Do Tracheae Work? 1004
 - How Do Lungs Work? 1005
 - BOX 44.1 Respiratory Distress Syndrome in Premature Infants 1008
 - Homeostatic Control of Ventilation 1009
- 44.3 Blood 1009
 - Oxygen Delivery: The Bohr Effect 1010
 - CO₂ Transport and the Buffering of Blood pH 1011
 - Homeostatic Control of Oxygen-Carrying Capacity 1012
 - BOX 44.2 Blood Doping and Erythropoietin Abuse by Athletes 1013
- 44.4 The Circulatory System 1013
 - Open Circulatory Systems 1014
 - Closed Circulatory Systems 1014
 - The Heart 1016
 - BOX 44.3 Hypertension and the Measurement of Blood Pressure 1020
 - Regulation of Blood Pressure and Blood Flow 1021
- ESSAY Smoking and Lung Function 1023
- CHAPTER REVIEW 1024

Chapter 45

Electrical Signals in Animals 1026

- 45.1 Principles of Electrical Signaling 1027
 - The Anatomy of a Neuron 1027
 - An Introduction to Membrane Potentials 1028
 - The Resting Potential 1029
 - Using the Nernst and Goldman Equations 1030
 - Using Microelectrodes to Measure Membrane Potentials 1031
 - What Is an Action Potential? 1032
- 45.2 Dissecting the Action Potential 1033
 - Distinct Ion Currents Are Responsible for Depolarization and Repolarization 1033

	Voltage-Gated Channels	1033
	The Role of the Sodium-Potassium Pump	1036
	How Is the Action Potential Propagated?	1036
45.3	The Synapse	1038
	What Do Neurotransmitters Do?	1040
	Postsynaptic Potentials and Summation	1041
	What Happens When Ligand-Gated Channels Are Defective?	1042
45.4	The Vertebrate Nervous System	1044
	A Closer Look at the Peripheral Nervous System	1044
	Functional Anatomy of the CNS	1044
	How Does Memory Work?	1046
ESSAY	Can Brain Tissue Transplants Help People with Parkinson's Disease?	1049
CHAPTER REVIEW		1050

Chapter 46

Animal Sensory Systems and Movement 1052

46.1	How Do Sensory Organs Convey Information to the Brain?	1053
	BOX 46.1 Senses That Humans Don't Have	1054
	Sensory Transduction	1054
	Transmitting Information to the Brain	1055
46.2	Hearing	1055
	How Do Sensory Cells Respond to Sound Waves and Other Forms of Pressure?	1055
	The Mammalian Ear	1057
	Sensory Worlds: What Do Other Animals Hear?	1058
46.3	Vision	1059
	The Vertebrate Eye	1060
	BOX 46.2 Vertebrate versus Cephalopod Eyes	1061
	Sensory Worlds: Do Other Animals See Color?	1063
46.4	Taste and Smell	1064
	Taste: Detecting Molecules in the Mouth	1064
	Olfaction: Detecting Molecules in the Air	1065
46.5	Movement	1066
	Skeletons	1067
	How Do Muscles Contract?	1068
ESSAY	Sprinters and Marathoners—Born or Made?	1073
CHAPTER REVIEW		1074

Chapter 47

Chemical Signals in Animals 1076

47.1	Cataloging Hormone Structure and Function	1077
	How Do Researchers Identify a Hormone?	1078
	Chemical Characteristics of Hormones	1079
	The Human Endocrine System—An Overview	1080
47.2	What Do Hormones Do?	1081
	BOX 47.1 A Closer Look at Thyroxine and the Thyroid Gland	1082
	How Do Hormones Coordinate Responses to Environmental Change?	1082
	How Do Hormones Direct Developmental Processes?	1084

	How Are Hormones Involved in Homeostasis?	1084
	BOX 47.2 Is Human Obesity Caused by Leptin Deficiency?	1087
47.3	How Is the Production of Hormones Regulated?	1088
	The Hypothalamus and Pituitary Gland	1088
	Control of Epinephrine by Sympathetic Nerves	1091
47.4	How Do Hormones Act on Target Cells?	1091
	Steroid Hormones and Intracellular Receptors	1091
	Hormones That Bind to Cell-Surface Receptors	1093
ESSAY	Do Humans Produce Pheromones?	1096
CHAPTER REVIEW		1097

Chapter 48

Animal Reproduction 1099

48.1	Asexual and Sexual Reproduction	1100
	Mechanisms of Asexual Reproduction	1100
	Switching Reproductive Modes: A Case History	1100
	Mechanisms of Sexual Reproduction:	
	Gametogenesis	1101
48.2	Fertilization and Egg Development	1102
	External Fertilization	1102
	Internal Fertilization and Sperm Competition	1103
	BOX 48.1 Unusual Aspects of Fertilization	1103
	Oviparity and Viviparity	1105
48.3	Reproductive Structures and Their Functions	1106
	The Male Reproductive System	1106
	The Female Reproductive System	1108
48.4	The Role of Sex Hormones in Mammalian Reproduction	1109
	Puberty	1109
	BOX 48.2 Abuse of Synthetic Steroids	1110
	Female Sex Hormones and the Menstrual Cycle	1111
48.5	Human Pregnancy and Birth	1113
	Major Events during Pregnancy	1114
	How Does the Mother Nourish the Fetus?	1115
	Birth	1116
ESSAY	Contraception	1117
CHAPTER REVIEW		1118

Chapter 49

The Immune System in Animals 1120

49.1	Innate Immunity	1121
	Barriers to Entry	1121
	The Innate Immune Response	1122
49.2	The Acquired Immune Response: Recognition	1124
	An Introduction to Lymphocytes and the Immune System	1124
	The Discovery of B Cells and T Cells	1125
	Antigen Recognition and Clonal Selection	1125
	BOX 49.1 How Are Monoclonal Antibodies Produced?	1128

- How Does the Immune System Distinguish Self from Nonsell? 1132
- 49.3 The Acquired Immune Response: Activation 1132
 - Antigen Presentation by MHC Proteins: Activating T Cells 1133
 - B-Cell Activation and Antibody Secretion 1134
 - Antigen Presentation by Infected Cells: A Signal for Action by $CD8^+$ T Cells 1135
- 49.4 The Acquired Immune Response: Culmination 1135
 - Killing Bacteria 1135
 - Destroying Viruses 1135
 - Why Does the Immune System Reject Foreign Tissues and Organs? 1136
 - Responding to Future Infections: Immunological Memory 1137
 - BOX 49.2 The ELISA Test 1138
- ESSAY Allergies 1139
- CHAPTER REVIEW 1140

UNIT 9

Ecology 1143

Chapter 50

An Introduction to Ecology 1144

- 50.1 Areas of Ecological Study 1145
 - Organismal Ecology 1145
 - Population Ecology 1145
 - Community Ecology 1146
 - Ecosystem Ecology 1146
 - How Do Ecology and Conservation Efforts Interact? 1146
- 50.2 The Nature of the Environment 1146
 - Climate 1147
- 50.3 Types of Terrestrial and Aquatic Ecosystems 1149
 - Terrestrial Biomes 1150
 - Aquatic Environments 1155
 - BOX 50.1 Thermoclines and Lake Turnover 1157
 - BOX 50.2 Deep-Sea Life 1160
- 50.4 Biogeography 1160
 - The Role of History 1161
 - Biotic Factors 1161
 - Abiotic Factors 1162

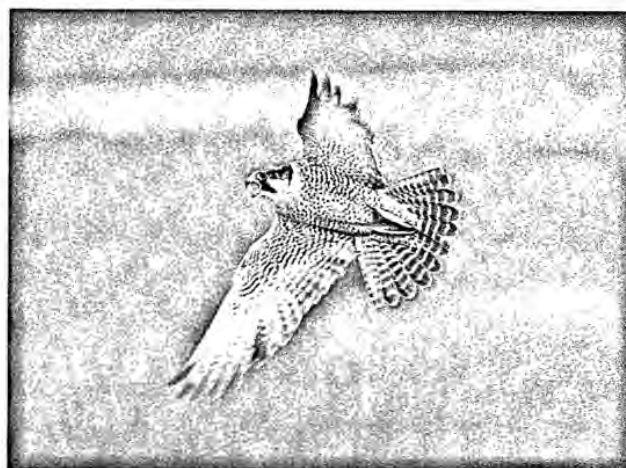
ESSAY Battling Invasive Species 1164

CHAPTER REVIEW 1164

Chapter 51

Behavior 1166

- 51.1 Types of Behavior—An Overview 1167
 - Fixed Action Patterns 1168
 - BOX 51.1 A Closer Look at Behavior Genetics 1169
 - Conditional Strategies 1171



- 51.2 Learning 1172
 - Simple Types of Learning: Classical Conditioning and Imprinting 1172
 - More Complex Types of Learning: Birdsong 1173
 - Can Animals Think? 1174
 - What Is the Adaptive Significance of Learning? 1175
- 51.3 How Animals Act: Hormonal Control 1177
- 51.4 Communication 1178
 - Modes of Communication 1179
 - A Case History: The Honeybee Dance 1180
 - Honesty and Deceit in Communication 1181
- 51.5 Orientation, Navigation, and Migration 1183
 - Migration: Why Do Animals Move with a Change of Seasons? 1184
 - Navigation: How Do Animals Find Their Way? 1184
- 51.6 The Evolution of Self-Sacrificing Behavior 1185
 - Do Organisms Act for the Good of the Species? 1186
 - Kin Selection 1186
 - BOX 51.2 Calculating the Coefficient of Relatedness 1187
 - Reciprocal Altruism 1188
- ESSAY Children at Risk 1189
- CHAPTER REVIEW 1190

Chapter 52

Population Ecology 1192

- 52.1 Demography 1193
 - Life Tables 1193
 - The Role of Life History 1195
 - BOX 52.1 Using Life Tables to Calculate Population Growth Rates 1196
- 52.2 Population Growth 1196
 - Exponential Growth 1196
 - Logistic Growth 1197
 - BOX 52.2 Developing and Applying Population Growth Equations: A Closer Look 1197
 - What Limits Growth Rates and Population Sizes? 1200

xxx Contents

- 52.3 Population Dynamics 1201
 - Population Cycles: The Case of the Red Grouse 1201
 - Age Structure 1203
 - Analyzing Change in the Growth Rate of Human Populations 1205
- 52.4 How Can Population Ecology Help Endangered Species? 1206
 - Preserving Metapopulations 1206
 - Using Life-Table Data to Make Population Projections 1207
 - BOX 52.3 Mark-Recapture Studies 1208
 - Population Viability Analysis 1210
- ESSAY What Limits Human Population Growth? 1211
- CHAPTER REVIEW 1211

Chapter 53

Community Ecology 1214

- 53.1 Species Interactions 1214
 - Competition 1215
 - Consumption 1220
 - Mutualism 1227
- 53.2 Community Structure 1229
 - Clements and Gleason: Two Views of Community Dynamics 1229
 - Disturbance and Change in Ecological Communities 1231
 - What Disturbances Occur, and How Frequent and Severe Are They? 1232
 - Succession: The Development of Communities after Disturbance 1232
- 53.3 Species Diversity in Ecological Communities 1236
 - Global Patterns in Species Diversity 1236
 - BOX 53.1 Measuring Species Diversity 1237
 - The Role of Diversity in Ecological Communities 1238
- ESSAY Let-It-Burn Policies 1239
- CHAPTER REVIEW 1241



Chapter 54

Ecosystems 1243

- 54.1 Energy Flow and Trophic Structure 1244
 - Global Patterns in Productivity 1244
 - How Does Energy Flow through an Ecosystem? 1247
 - Analyzing Energy Flow: A Case History 1251
- 54.2 Biogeochemical Cycles 1252
 - Biogeochemical Cycles in Ecosystems 1253
 - Global Biogeochemical Cycles 1254
- 54.3 Human Impacts on Ecosystems 1259
 - Global Warming 1259
- ESSAY Phosphate Pollution, Acid Rain, and the Ozone Hole: Hope for Ecosystem Recovery? 1261
- CHAPTER REVIEW 1263

Chapter 55

Biodiversity and Conservation 1265

- 55.1 What Is Biodiversity? 1266
 - Biodiversity Can Be Measured and Analyzed at Several Levels 1266
 - Why Is Biodiversity Important? 1267
- 55.2 How Do Biologists Study Biodiversity? 1268
 - Quantifying Genetic Diversity 1268
 - Estimating the Total Number of Species Living Today 1269
 - Steps in Understanding Biodiversity 1269
 - Biodiversity and Ecosystem Function 1270
- 55.3 Threats to Biodiversity 1272
 - Humans Have Affected Biodiversity throughout History 1272
 - Current Threats to Biodiversity 1273
 - How Can Biologists Predict Future Extinction Rates? 1275
- 55.4 Conserving Biodiversity: Biology, Sociology, Economics, and Politics 1277
 - The Role of Governmental and Private Agencies 1278
 - Conservation Strategies 1279
 - Looking to the Future 1279
- ESSAY Models for the Future: The Malpai Borderlands and Guanacaste 1280
- CHAPTER REVIEW 1281

Appendix A-1

Glossary G-1

Image Credits C-1

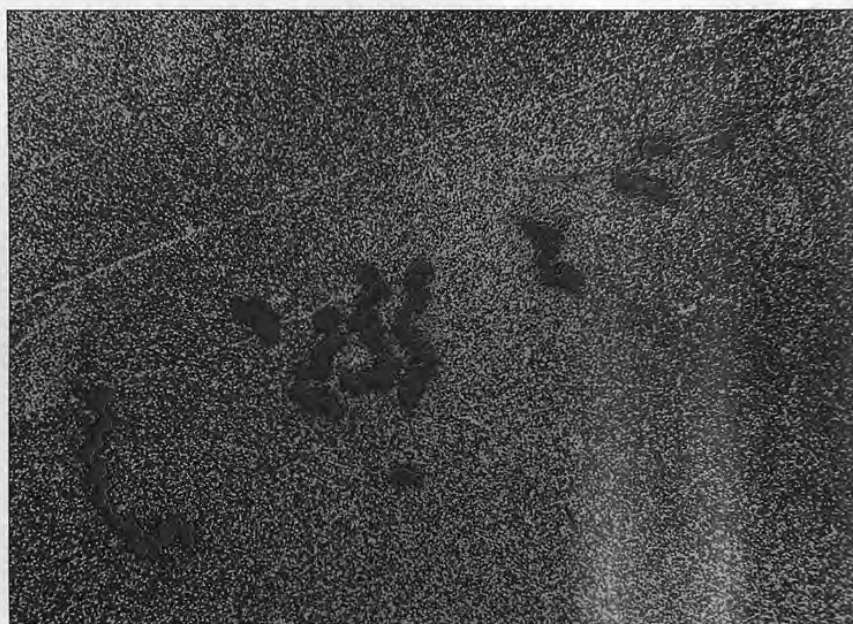
Index I-1

16

Transcription and Translation

KEY CONCEPTS

- After the enzyme RNA polymerase binds to a specific site in DNA with the help of other proteins, it catalyzes the production of an RNA molecule. The base sequence of the RNA produced is complementary to the base sequence of the DNA template strand.
- In eukaryotes, genes consist of coding regions called exons interspersed with noncoding regions called introns. After a eukaryotic mRNA is produced, introns must be spliced out. The ends of the molecule receive a cap and tail.
- Inside ribosomes, mRNAs are translated to proteins via intermediary molecules called transfer RNAs. Transfer RNAs carry an amino acid and have a three-base-pair anticodon, which binds to an mRNA codon. The amino acid carried by the transfer RNA is then added to the growing protein via formation of a peptide bond.
- Many mutations produce changes in the phenotype.

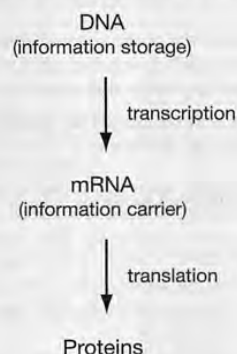


A micrograph showing transcription and translation occurring simultaneously in a bacterial cell. The yellow lines are strands of DNA, and the red structures are ribosomes.

Proteins are the stuff of life. They give cells their shape and control the chemical reactions that go on inside them. No one knows exactly how many different proteins can be made in each of your body's cells, but 100,000 is a reasonable guess. While some of those 100,000 different proteins may not be manufactured at all in certain types of cells, other proteins may be present in quantities ranging from millions of copies to fewer than a dozen. If even one of those types of protein is defective, the cell may sicken or even die.

A cell builds the proteins it needs from instructions encoded in its genome. Chapter 15 explored the basic strategy that cells use to retrieve these instructions and convert them into action in the form of proteins. The first step in the process is the transcription of a gene and the production of a messenger RNA, or mRNA. The RNA message is a short-lived copy of the archived

instructions in DNA. The sequence of ribonucleotides in the mRNA is then translated into a sequence of amino acids. This series of events defines the flow of information in the cell:



The discovery of the relationship between DNA and proteins was one of the great scientific advances of the twentieth century. But once the relationship between transcription and translation had been firmly established, biologists puzzled over how cells actually accomplish the feat of converting genetic information into protein. How do transcription and translation work? For example, how does RNA polymerase know where to start transcribing a gene and where to end? Once an RNA message is produced, how is the linear sequence of ribonucleotides translated into the linear sequence of amino acids in a protein?

This chapter explores the experiments that answered these questions. In addition to providing a deeper understanding of how cells work, the results summarized here have had a significant impact on biomedicine and drug development. Many of the antibiotics prescribed today work by disrupting translation in bacteria that cause disease. In addition, research on transcription and translation set the stage for an explosion of work, introduced in Chapters 17 and 18, on how the expression of specific genes is regulated. Most genes are transcribed and translated only at certain times or in certain types of cells.

Let's begin by examining research on the mechanisms of transcription in bacteria and eukaryotes, then delve into the mechanics of translation. The chapter concludes by examining how changes in DNA can lead to changes in protein structure and function.

16.1 Transcription in Bacteria

The first step in converting genetic information into proteins is to synthesize a messenger RNA version of the instructions archived in DNA. The enzyme RNA polymerase, introduced in Chapters 14 and 15, is responsible for synthesizing mRNA. Like the DNA polymerases introduced in Chapter 14, RNA polymerase performs a template-directed synthesis in the $5' \rightarrow 3'$ direction. (Recall that $5'$ and $3'$ refer to carbons on the sugar subunits of ribonucleotides.) But unlike DNA polymerases, RNA polymerase does not require a primer to begin transcription.

Transcription occurs when RNA polymerase matches the base in a ribonucleotide triphosphate with the complementary base in a gene—a section of DNA. Once a matching ribonucleotide is in place, RNA polymerase catalyzes the formation of a phosphodiester bond between the $3'$ end of the growing mRNA chain and the new ribonucleotide. As this matching-and-catalysis process continues, an RNA that is complementary to the gene is synthesized (Figure 16.1). Note that only one of the two DNA strands is used as a template and transcribed, or “read,” by RNA polymerase. The strand that is read by the enzyme is called the **template strand**; the other strand is called the **non-template strand**. Because RNA has uracil (U) rather than thymine (T), an adenine (A) in the DNA template strand specifies a U in the complementary RNA strand.

Biologists reached these conclusions by studying RNA synthesis in cell-free, or *in vitro*, systems. Once the basic chemical

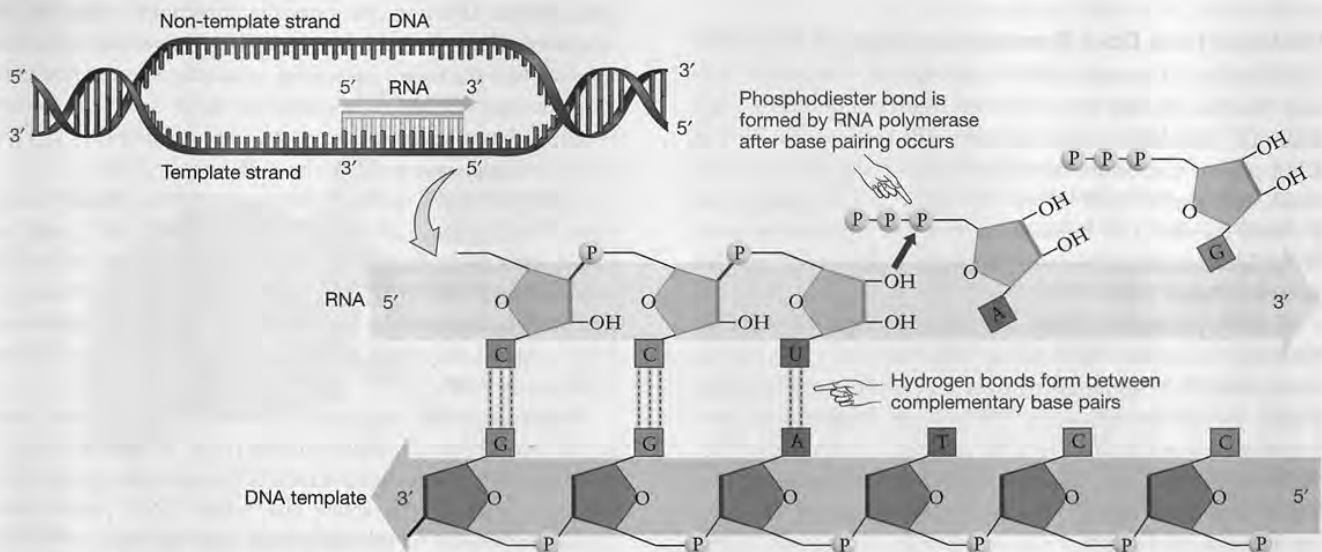


FIGURE 16.1 Transcription Is the Synthesis of RNA from a DNA Template

The reaction catalyzed by RNA polymerase results in the formation of a phosphodiester bond between ribonucleotides. RNA polymerase produces an RNA strand whose sequence is complementary to the bases in the DNA template. **QUESTION** In which direction is RNA synthesized, $5' \rightarrow 3'$ or $3' \rightarrow 5'$? In which direction is the DNA template “read”?

340 Unit 3 Gene Structure and Expression

reaction was understood, an entirely new set of questions arose:

- What does RNA polymerase look like?
- Are any other proteins or factors involved in transcription, or does RNA polymerase act alone?
- How does the enzyme know where to start transcription on the DNA template?

RNA Polymerase Structure and Function

To understand the structure of RNA polymerase, researchers have employed the technique called X-ray crystallography introduced in Chapter 4. Recall that this procedure allows biologists to obtain information about the three-dimensional structure of large, complex molecules.

The most recent models of bacterial RNA polymerase indicate that the enzyme is large and globular and has several prominent channels that run through the interior. The enzyme's active site, where phosphodiester bonds form, is located where several of these channels intersect. The active site contains a strategically placed magnesium atom (Mg^{2+}), which plays a critical role in stabilizing the transition state as the polymerization reaction takes place.

How does this structure correlate with the enzyme's function? It is logical to predict that DNA fits into one of the enzyme's channels and that the two strands in the double helix separate inside the enzyme to expose a single-stranded template at the active site. If so, how do the enzyme and DNA come together?

Initiation: How Does Transcription Begin in Bacteria?

Soon after the discovery of RNA polymerase, researchers realized that the enzyme cannot initiate transcription on its own. Instead, a detachable protein subunit called **sigma** must bind to RNA polymerase before transcription can begin. RNA polymerase and sigma form what biologists call a **holoenzyme** ("whole enzyme"). A holoenzyme consists of a **core enzyme**, which contains the active site for catalysis, and other required proteins (Figure 16.2).

If RNA polymerase is the core enzyme of this holoenzyme, what does sigma do? When researchers mixed RNA polymerase, sigma, and DNA together, they found that the holoenzyme bound tightly to specific sections of DNA. These binding sites were named **promoters**, because they are sections of DNA where transcription begins. The discovery of promoters suggested that sigma's function is regulatory in nature. Sigma appeared to be responsible for guiding RNA polymerase to specific locations where transcription should begin. What is the nature of these specific locations? What do promoters look like, and what do they do?

David Pribnow offered an initial answer to these questions in the mid-1970s. When Pribnow analyzed the base sequence of promoters from a variety of bacteria and from viruses that infect bacteria, he found that the promoters are all 40–50 base

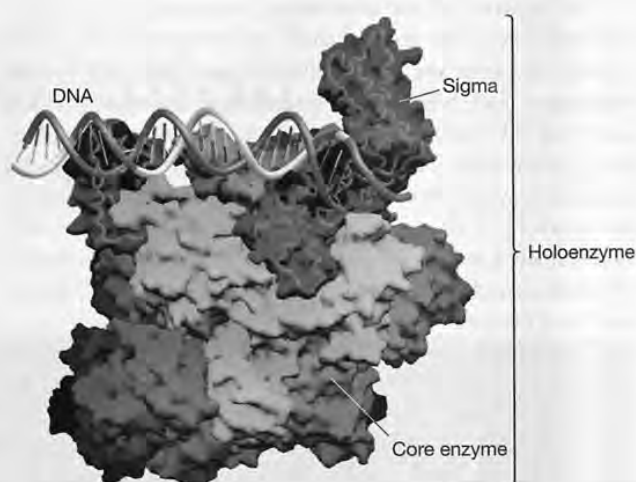


FIGURE 16.2 The Three-Dimensional Structure of RNA Polymerase Bacterial RNA polymerase is a globular enzyme. Sigma is a regulatory component, and the core enzyme contains the active site.

pairs long and that a particular section in each promoter looked similar. This similar segment of DNA had a series of bases identical or similar to TATAAT. This six-base-pair sequence is now known as the **–10 box**, because it occurs about 10 bases from the point where RNA polymerase starts transcription. (DNA that is located in the direction that RNA polymerase moves during transcription is said to be **downstream**; DNA in the opposite direction is said to be **upstream**. Thus, the –10 box is 10 bases upstream from the transcription start site.) Soon after, researchers recognized that the sequence TTGACA occurred in these same promoters about 35 bases upstream from the start of the mRNA. This second key sequence is called the **–35 box**.

Follow-up work showed that transcription begins when sigma binds to the –35 and –10 boxes (Figure 16.3, step 1). This is a key point: Sigma, and not RNA polymerase, makes the initial contact with DNA that starts transcription in bacteria. This finding supports the hypothesis that sigma is a regulatory protein. Sigma tells RNA polymerase where and when to start synthesizing RNA.

Once sigma binds to a promoter, the DNA helix opens and creates two strands of single-stranded DNA. As shown in step 2 of Figure 16.3, the template strand is threaded through a channel that leads to the active site inside RNA polymerase. Monomers known as **ribonucleoside triphosphates**, or NTPs, enter a channel at the bottom of the enzyme and diffuse to the active site. When an incoming NTP pairs with a complementary base on the template strand of DNA, RNA polymerization begins. The reaction catalyzed by RNA polymerase is exergonic and spontaneous because NTPs have so much potential energy, owing to their three phosphate groups. As step 3 of Figure 16.3



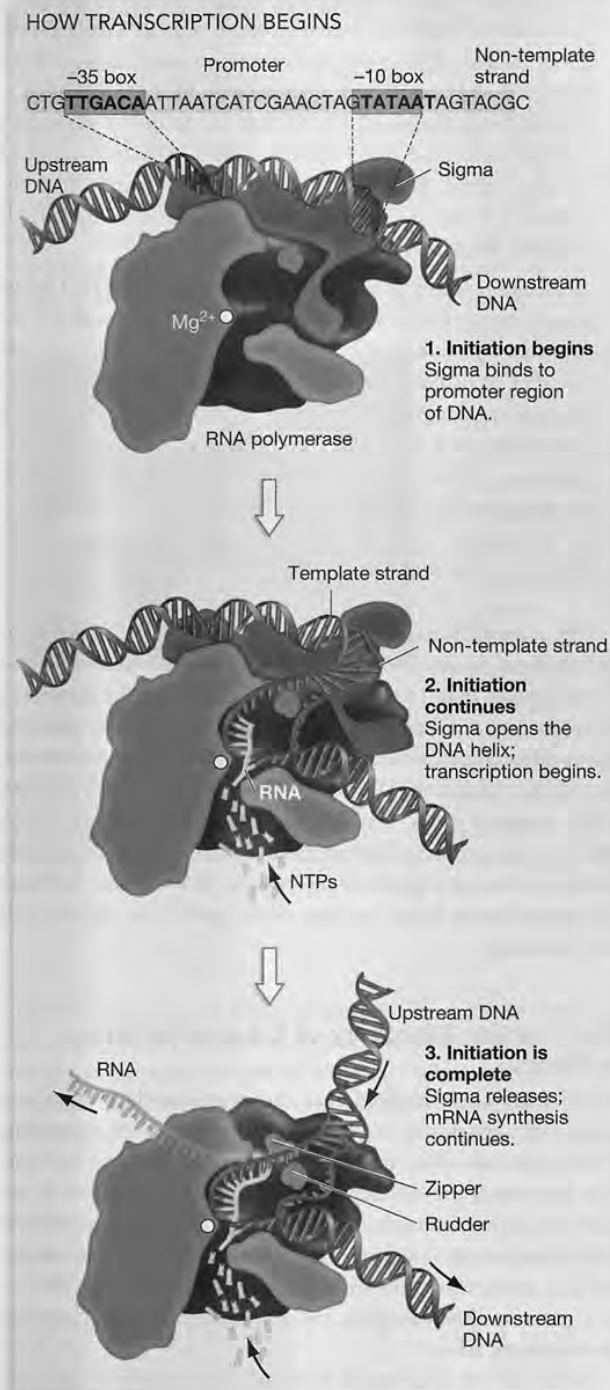


FIGURE 16.3 In Bacteria, Sigma Plays a Key Role in Initiating Transcription

When sigma attaches to RNA polymerase, the holoenzyme is able to bind to the -35 box and the -10 box of a promoter. Once it has bound to a promoter, sigma opens the DNA helix and threads the template strand through the enzyme's active site. When transcription is under way, sigma disengages from RNA polymerase. **EXERCISE** Circle and label RNA polymerase's active site.

shows, sigma is released once RNA synthesis is under way. The **initiation phase** of transcription is complete.

Recent work has shown that most bacteria have several types of sigma proteins, each with a distinctive structure and function. *Escherichia coli* has seven different sigma proteins, for example, while *Streptomyces coelicolor* has more than 60. Each of these proteins binds to promoters with slightly different DNA base sequences. Although all promoters have a -10 box and a -35 box, the sequences within each box vary somewhat. These variations in DNA sequences affect the ability of different sigma proteins to bind. As a result, each type of sigma protein allows RNA polymerase to bind to a different type of gene. The identity of the sigma protein in the RNA polymerase holoenzyme determines which types of genes will be transcribed. Thus by controlling which sigma proteins are active, bacterial cells control which genes are expressed.

Elongation and Termination

During the **elongation phase** of transcription, RNA polymerase moves along the DNA template in the $3' \rightarrow 5'$ direction of the template strand, synthesizing RNA in the $5' \rightarrow 3'$ direction. In the interior of the enzyme, a group of projecting amino acids called the enzyme's *zipper* helps open the double helix at the upstream end, and a nearby group of amino acids called the *rudder* helps steer the template and non-template strands through channels inside the enzyme. Meanwhile, the enzyme's active site catalyzes the addition of nucleotides to the $3'$ end of the growing RNA molecule at the rate of about 50 nucleotides per second.

Note that during the elongation phase of transcription, three prominent channels or grooves in the enzyme are filled. Double-stranded DNA goes into and out of one groove; ribonucleoside triphosphates enter another; and the growing RNA strand exits to the rear. In this way, the enzyme's structure correlates closely with its function.

Transcription ends with a **termination phase**. In most cases, transcription stops when RNA polymerase reaches a stretch of DNA sequence that functions as a **transcription termination signal**. The bases that make up the termination signal code for a stretch of RNA with an unusual property: As soon as it is synthesized, the RNA sequence folds back on itself and forms a short double helix that is held together by complementary base pairing. The secondary structure that results is called a *hairpin* (see Chapter 4). The formation of the hairpin structure is thought to disrupt the interaction between RNA polymerase and the RNA transcript, resulting in the physical separation of the enzyme and its product.

To summarize, transcription begins when sigma binds to the promoter at the start of a gene. Once binding occurs, RNA polymerase begins to synthesize mRNA. Transcription ends when a termination signal at the end of the gene leads to the formation of a hairpin in the mRNA, disrupting the transcription complex.

16.2 Transcription in Eukaryotes

The data reviewed in Section 16.1 established how transcription occurs in bacteria. How do these results compare with research on the mechanism of transcription in eukaryotes? What similarities and differences have come to light?

One of the first conclusions about transcription in eukaryotes is that it is similar to bacterial transcription in an important respect: RNA polymerase does not bind directly to promoter sequences by itself. Instead, proteins that came to be called **basal transcription factors** initiate eukaryotic transcription by matching the enzyme with the appropriate promoter region in DNA. The function of basal transcription factors observed in eukaryotes is analogous to the function of the sigma proteins in bacteria.

The discovery of basal transcription factors emerged from studies with cell-free, or *in vitro*, systems. The basic approach was to purify RNA polymerase from human cells, add template DNA, and analyze the RNAs that were produced. But in early experiments, researchers found that RNA polymerase began copying at random locations on the template DNA instead of specifically at promoter regions. In addition, when RNA polymerase was alone, both strands of the DNA were transcribed instead of just one. This observation inspired the hypothesis that eukaryotic cells must contain sigma-like proteins that are required for normal transcription. To test this hypothesis, biologists added proteins isolated from human cells to the cell-free reaction system. By adding one or a few proteins at a time and recording which ones enabled RNA polymerase to bind to promoters and correctly transcribe the template strand, researchers were gradually able to characterize the suite of proteins that came to be called basal transcription factors.

Follow-up work confirmed several important distinctions about how transcription works in bacteria and eukaryotes:

- In bacteria a single sigma protein binds to a promoter and initiates transcription, but in eukaryotes many basal transcription factors are required to initiate transcription.
- Eukaryotes have three distinct types of RNA polymerase (abbreviated RNA pol) instead of just one. As Table 16.1 shows, RNA pol I, pol II, and pol III each transcribe a dis-

crete class of RNA. Only RNA pol II transcribes the genes that code for proteins. Stated another way, RNA pol II produces mRNA. RNA pol I, in contrast, makes the large RNA molecules that are found in ribosomes. RNA pol III manufactures both the small RNAs that are found in ribosomes and the molecules called transfer RNAs, which are required for translation. Box 16.1 explains how research on toxic mushrooms led to the discovery that eukaryotes have three different RNA polymerases.

- Although eukaryotic genomes contain promoters that signal where transcription should begin, just as bacteria do, the promoters in eukaryotic DNA are much more diverse and complex than bacterial promoters. Many of the eukaryotic promoters recognized by RNA polymerase II include a unique sequence called the **TATA box**, located 30 base pairs upstream of the transcription start site. Some of the promoters recognized by pol II do not contain a TATA box, however. In addition, RNA pol I and pol III interact with entirely different promoters.

The overall message of this research is that the molecular mechanisms involved in transcription are much more complex in eukaryotes than in bacteria. An even more striking contrast between bacterial and eukaryotic gene expression emerged when researchers discovered that eukaryotic genes do not consist of one continuous DNA sequence that codes for a product, as do bacterial genes. Instead, the regions of eukaryotic genes that code for proteins come in pieces that are separated by hundreds or many thousands of intervening DNA bases. Although these intervening bases are part of the gene, they do not code for a product.

The Startling Discovery of Eukaryotic Genes in Pieces

When researchers realized that the protein-coding regions of eukaryotic genes are interrupted by stretches of noncoding DNA, it became clear that the nature of the gene and information processing are different in bacteria and eukaryotes. In eukaryotes, the conversion of information in DNA sequences to mRNA sequences does not occur directly. To make a functional mRNA, eukaryotic cells must dispose of certain sequences inside genes and then combine the separated coding sections into an integrated whole.

What sort of data would provoke such a startling claim? To answer this question, consider work that Phillip Sharp and colleagues carried out in the late 1970s to determine how DNA templates are transcribed. They began one of their experiments by heating DNA molecules enough to break the hydrogen bonds between complementary bases. This treatment separated the two strands. The single-stranded DNA was then incubated with the mRNA encoded by the sequence. The team's idea was to promote base pairing between the mRNA and the single-

TABLE 16.1 Eukaryotic RNA Polymerases

Name of Enzyme	Type of Gene Transcribed
RNA polymerase I (RNA pol I)	Genes that code for most of the large RNA molecules (rRNAs) found in ribosomes (see Section 16.3)
RNA polymerase II (RNA pol II)	Protein-coding genes (produce mRNAs)
RNA polymerase III (RNA pol III)	Genes that code for transfer RNAs (see Section 16.4), and genes that code for one of the small RNA molecules (rRNAs) found in ribosomes
RNA pol II and RNA pol III	RNA molecules found in snRNPs (see Section 16.2)

BOX 16.1 Toxins and Transcription

Although wild mushrooms can be delicious, dining on certain species may be lethal. For example, in late summer and fall the reproductive structures of the death cap mushroom (*Amanita phalloides*) appear in forests, fields, and backyards in North America and Europe (Figure 16.4). The death cap is among



FIGURE 16.4 The Death Cap Mushroom
These mushrooms contain the toxin α -amanitin, which at low concentrations poisons the eukaryotic enzyme RNA polymerase II. **QUESTION** State a hypothesis to explain why the presence of α -amanitin increases the fitness of death cap mushrooms.

the most poisonous of all mushrooms. Worldwide, it is thought to account for 90 percent of mushroom poisoning deaths. Eating a single death cap can kill a healthy adult.

The active agent in death caps is a toxin called α -amanitin. By happenstance, researchers who were investigating α -amanitin's mode of action made a crucial discovery about transcription in eukaryotes. The researchers observed that mRNA synthesis stopped when eukaryotic cells were treated with low concentrations of α -amanitin. The production of other types of RNAs—specifically, transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs)—continued, however. At very high concentrations of the toxin, tRNA production also stopped but transcription of rRNA genes continued. These observations eventually led to the discovery that eukaryotes have three RNA polymerases, each with a distinct role in the transcription of mRNA, tRNA, and rRNA. Low concentrations of α -amanitin inhibit

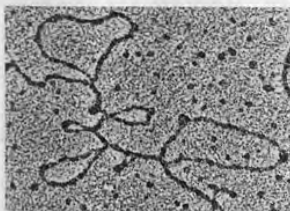
RNA polymerase II, which synthesizes mRNAs. As a result, low concentrations of the toxin halt protein synthesis. High concentrations of this poison block RNA polymerase III, which synthesizes tRNAs. RNA polymerase I is not affected by α -amanitin at any concentration.

Toxins such as α -amanitin are useful in research on transcription and translation for the same reason that knock-out mutations, which completely disable genes, are useful in genetics. Toxins allow researchers to understand what happens when a molecule or a particular stage in a process does *not* work. This is similar, in concept, to learning how a car works by removing a part or a system and carefully recording the machine's response; or to learning how the brain works by studying the behavior of people whose brains have been damaged in specific, quantifiable ways by accidents. Toxins and knock-out mutations provide a molecular dissecting kit. They let biologists explore the intricacies of cell function by inactivating key components.

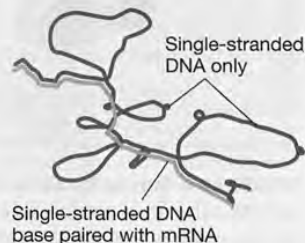
stranded DNA. Under these conditions, the research team expected the mRNA to form base pairs with the DNA sequences that act as the template for its synthesis. When the researchers examined the DNA-RNA hybrid molecules with the electron microscope, however, they observed the structure shown in

Figure 16.5a. Instead of matching up exactly, parts of the DNA formed loops. As **Figure 16.5b** shows, Sharp and his co-workers interpreted these loops as stretches of nucleotides that are present in the DNA template strand but are *not* in the corresponding mRNA.

(a) Micrograph of DNA-RNA hybrid



(b) Interpretation of micrograph



(c) Genes and RNA transcripts differ in length.



FIGURE 16.5 The Discovery of Noncoding Regions of DNA

(a) Electron micrograph of single-stranded DNA that is bonded via complementary base pairing to the mRNA it encodes. **(b)** Researchers' interpretation of the photograph in part (a). The loops represent regions of DNA that do not have an equivalent sequence in the mRNA. These intervening regions are "extra" DNA compared with the sequences in the mRNA. **(c)** Because eukaryotic genes contain both introns and exons, these genes are much larger than the corresponding mature mRNA.

When it became clear that the genes-in-pieces hypothesis was correct, Walter Gilbert suggested that regions of eukaryotic genes that are part of the final mRNA be referred to as **exons** (because they are *expressed*) and the untranslated stretches as **introns** (because they are *intervening*). Introns are sections of genes that are not represented in the final mRNA product. As a result, eukaryotic genes are much larger than their corresponding mature RNA transcripts (Figure 16.5c, page 343).

The transcription of eukaryotic genes by RNA polymerase generates a **primary RNA transcript** that contains both the exon and intron regions (Figure 16.6a). As transcription proceeds, the introns are removed from the growing mRNA strand by a process known as **splicing**. In this phase of information processing, pieces of the primary transcript are removed and the remaining segments are joined together. Splicing occurs while transcription is still under way and results in an mRNA that contains an uninterrupted genetic message.

Once the initial cut in the RNA is made, the free 5' end of the intron becomes attached to the adenine nucleotide and forms a "lariat" of RNA (step 2, Figure 16.6b). Then the free 3' end of the first exon reacts with the 5' end of the second exon. This reaction

DNA 3' Promoter Exon 1 Intron 1 Exon 2 Intron 2 Exon 3 5'

Primary RNA transcript 5' 3'

Spliced transcript 5' 3'

Primary RNA

5' Exon 1 Intron Exon 2 3'

snRNPs

A

Spliceosome

5' 3'

1. Several snRNPs and proteins assemble to form a spliceosome. The 2' hydroxyl group on an adenine nucleotide (A) attacks the 5' end of the intron, breaking RNA.

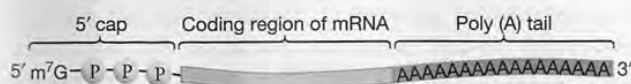
2. The 5' end of the intron becomes attached to the A nucleotide, forming a loop of RNA. The free 3' end of one exon attacks the 5' end of the other.

Excised intron

3. The 3' and 5' ends of adjacent exons bond covalently, releasing the intron (which will then degrade).

5' Exon 1 Exon 2 3'

Motif Exhibit 1059, 542 of 653

**FIGURE 16.7 In Eukaryotes, mRNAs Are Given a Cap and a Tail**

Eukaryotic mRNAs have a cap consisting of a molecule called 7-methylguanylate (symbolized as m^7G) bonded to three phosphate groups; the tail is made up of a long series of adenine residues.

breaks the 3' end of the intron and covalently joins the two exons into a contiguous coding sequence (step 3, Figure 16.6b). Splicing is now complete. In most cases, the excised intron is degraded to ribonucleotide monophosphates.

Current data suggest that both the cutting and rejoining reactions that occur during splicing are catalyzed by RNA molecules in the spliceosome. Recall from Chapter 4 that RNA molecules with catalytic ability are called ribozymes. In eukaryotes, ribozymes play a key role in the production of proteins.

Other Aspects of Transcript Processing: Caps and Tails

Intron splicing is not the only way that primary mRNA transcripts are processed in eukaryotes. As soon as the 5' end of a eukaryotic mRNA emerges from RNA polymerase, enzymes add a structure called the **5' cap** (Figure 16.7). The cap consists of the molecule 7-methylguanylate and three phosphate groups. When transcription is complete, an enzyme cleaves the 3' end of most mRNAs, and other enzymes add a long tract of 100–250 adenine nucleotides. This sequence is known as the **poly (A) tail**.

Not long after the caps and tails on eukaryotic mRNAs were described, evidence began to accumulate that they protect mRNAs from degradation by ribonucleases and that they enhance the efficiency of translation. For example, experimental mRNAs that have a cap and a tail last longer when they are introduced into cells than do experimental mRNAs that lack a cap, a tail, or both a cap and a tail. The mRNAs with caps and

tails also produce more proteins than do mRNAs without caps and tails.

Recent work has shown why the cap and tail are so critical to mRNAs. The 5' cap serves as a recognition signal for the translation machinery, and the poly (A) tail extends the life span of an mRNA by protecting the message from degradation by ribonucleases in the cytoplasm.

✓CHECK YOUR UNDERSTANDING

Transcription occurs when RNA polymerase catalyzes the $5' \rightarrow 3'$ synthesis of an RNA molecule. RNA synthesis is driven by the potential energy stored in nucleotide triphosphates and is based on matching complementary base pairs to the sequence in a template DNA strand. **Table 16.2** summarizes some of the similarities and differences observed in transcription in bacteria versus eukaryotes. You should be able to diagram (1) how transcription initiation, elongation, and termination occur in bacteria and eukaryotes and (2) how mRNAs are processed in eukaryotes.

16.3 An Introduction to Translation

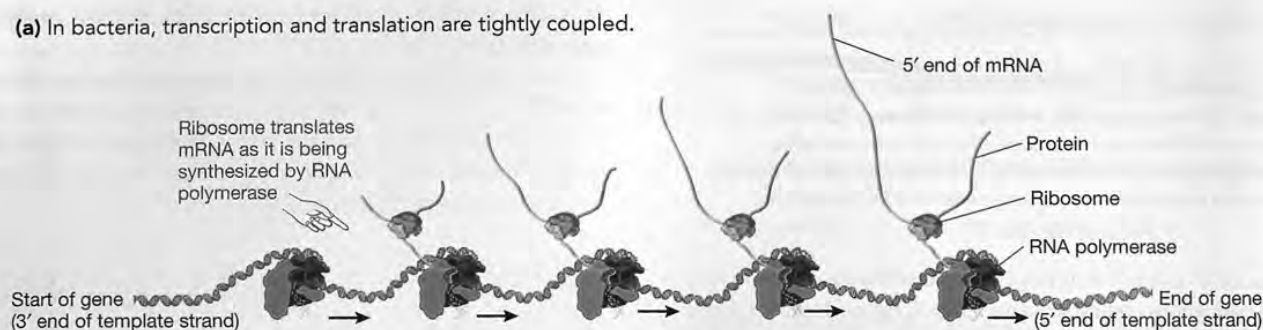
In translation, the sequence of bases in a messenger RNA molecule is converted into a sequence of amino acids in a polypeptide. The genetic code specifies the relationship between the bases of a triplet codon in mRNA and the amino acid it codes for (see Chapter 15). But how are the amino acids assembled into a polypeptide according to the information in messenger RNA?

Studying how translation occurred in cell-free systems proved to be an extremely productive approach to answering this question. Once in vitro translation systems had been developed from human cells, *E. coli*, and a variety of other organisms, it became clear that the basic mechanisms of translation are fundamentally the same throughout the tree of life. The sequence of events that occurs during protein synthesis is similar in bacteria and eukaryotes.

TABLE 16.2 Comparing Transcription in Bacteria and Eukaryotes

Aspect	Bacteria	Eukaryotes
RNA polymerase	One	Three; each produces a different class of RNA
Promoter structure	Typically contains a –35 box and a –10 box	Complex and variable; often includes a TATA box –30 from start of gene
Protein(s) involved in contacting promoter	Sigma; different versions of sigma bind to different promoters	Many basal transcription factors
RNA processing steps	None; translation occurs while transcription is still under way	Extensive; several processing steps occur in nucleus before RNA is exported to cytoplasm for translation: 1. Enzyme-catalyzed addition of 5' cap 2. Splicing (intron removal) by spliceosome 3. Enzyme-catalyzed addition of 3' poly (A) tail

(a) In bacteria, transcription and translation are tightly coupled.



(b) In eukaryotes, transcription and translation are separated in space and time.

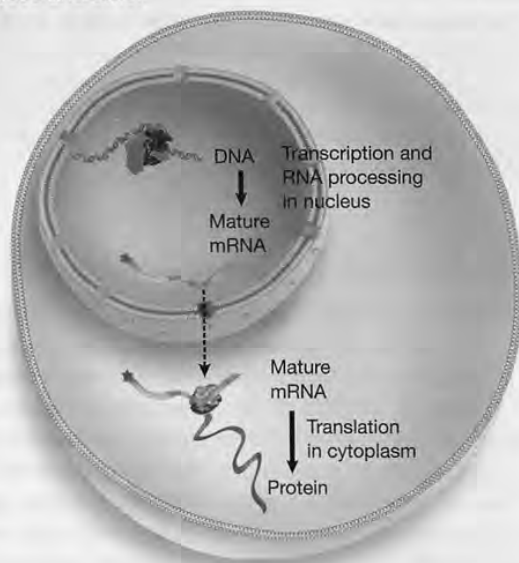


FIGURE 16.8 Transcription and Translation Occur Simultaneously in Bacteria but Not in Eukaryotes

(a) In bacteria, ribosomes attach to mRNA transcripts and begin translation while RNA polymerase is still transcribing the DNA template strand. (b) In eukaryotes, mRNA transcripts are processed inside the nucleus before they are exported to the cytoplasm. There, they are translated by ribosomes. **EXERCISE** In bacteria, many ribosomes attach to an mRNA transcript at the same time—so many copies of the protein product are made simultaneously. Add two ribosomes to the longest mRNA transcript in part (a), along with the protein each is producing.

of radioactive atoms that are incorporated into the molecule as it is being synthesized, followed by a chase of unlabeled atoms. Britten's group wanted to label amino acids with radioactive atoms and then track the molecules as translation occurred.

The biologists began by feeding a pulse of radioactive sulfate ($^{35}\text{SO}_4^{2-}$) to growing cultures of *E. coli*. They expected the cells to incorporate the radioactive sulfur into the amino acids methionine and cysteine—which contain sulfur (see Chapter 3)—and then into newly synthesized proteins. Fifteen seconds after adding the radioactive sulfate, the researchers “chased” the label by adding a large excess of nonradioactive sulfate to the culture medium. The chase vastly reduced the incorporation of radioactive sulfur into proteins. If the ribosome hypothesis was correct, then the radioactive signal should be associated with ribosomes for a short period of time. Later, all of the radioactivity should be found in proteins independently of the ribosomes.

This is exactly what the researchers found. Soon after the pulse of labeled sulfate ended, the radioactive atoms were found in free amino acids or in ribosomes. Later, all of the radioactive atoms were found on completed proteins. Based on these data, biologists concluded that proteins are synthesized at ribosomes and then released.

About a decade after the ribosome hypothesis was confirmed, electron micrographs like this chapter's opening photograph showed bacterial ribosomes in action. The images confirmed that, in bacteria, ribosomes attach to mRNAs and begin synthesizing proteins even before transcription is complete (Figure 16.8a). Transcription and translation can occur concurrently in bacteria because there is no nuclear envelope to

Ribosomes Are the Site of Protein Synthesis

The first question that biologists answered about translation concerned where it occurs. The answer was inspired by a simple observation: There is a strong, positive correlation between the number of small cellular structures known as **ribosomes** in a given type of cell and the rate at which that cell synthesizes proteins. For example, immature human red blood cells divide rapidly, synthesize millions of copies of the protein hemoglobin, and contain large numbers of ribosomes. In contrast, the same cells at maturity have low rates of protein synthesis and very few ribosomes. Based on this correlation, investigators proposed that ribosomes are the site of protein synthesis in the cell.

To test this hypothesis, Roy Britten and collaborators did a pulse-chase experiment similar in design to experiments introduced in Chapter 7. Recall that the goal of a pulse-chase experiment is to label a population of molecules as they are being produced. The location of the tagged molecules is then followed through time. The tagging is done by supplying a pulse

separate the two processes. In bacteria, transcription and translation are physically connected. But in eukaryotes, mRNAs are spliced in the nucleus and then exported to the cytoplasm. Once RNA messages are outside the nucleus, ribosomes attach to them and begin translation (Figure 16.8b). In eukaryotes, transcription and translation are separated in time and space.

How Does an mRNA Triplet Specify an Amino Acid?

During translation, genetic information is converted from nucleic acids to proteins. In effect, the hereditary instructions are translated from one distinctive chemical language into another. The discovery of the genetic code revealed that each triplet codon in mRNA specifies a particular amino acid in a protein. How does this conversion happen?

One early hypothesis was that mRNA codons and amino acids interact directly. The proposal was that the bases in a particular codon were complementary in shape or charge to the side group of a particular amino acid (Figure 16.9a). But Francis Crick pointed out that the chemistry involved didn't make

sense. For example, how could the nucleic acid bases interact with a hydrophobic amino acid side group, which does not form hydrogen bonds?

Crick proposed an alternative hypothesis. As Figure 16.9b shows, he suggested that some sort of adapter molecule holds amino acids in place while interacting directly and specifically with a codon in mRNA via hydrogen bonding. In essence, Crick predicted the existence of a chemical go-between that produced a physical connection between the two types of molecules. Crick, as it turns out, was right.

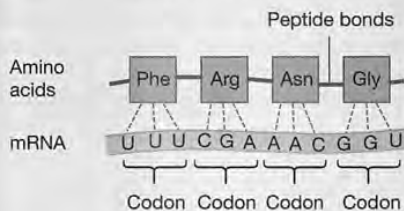
16.4 The Role of Transfer RNA

Crick's adapter molecule was discovered by accident. Biologists were trying to work out a cell-free protein synthesis system derived from mammalian liver cells and had discovered that ribosomes, mRNA, amino acids, ATP, and a molecule called guanosine triphosphate, or GTP, had to be present for translation to occur. (GTP is similar to ATP but contains guanosine instead of adenosine.) These results were logical—ribosomes provide the catalytic machinery, mRNAs contribute the message to be translated, amino acids are the building blocks of proteins, and ATP and GTP supply potential energy to drive the endergonic polymerization reactions responsible for forming proteins. But in addition, a cellular fraction that contained a previously unknown type of RNA turned out to be indispensable. If this type of RNA is missing, protein synthesis does not occur. What is this mysterious RNA, and why is it essential to translation?

The novel class of RNAs eventually became known as **transfer RNA (tRNA)**. The role of tRNA in translation was a mystery until some researchers happened to add a radioactive amino acid—leucine—to an in vitro protein synthesis system. The treatment was actually done as a control for an unrelated experiment. To the researchers' amazement, some of the radioactive leucine attached to tRNA molecules. Follow-up experiments showed that the attachment of an amino acid to a tRNA requires an input of energy in the form of ATP. A tRNA molecule that becomes covalently linked to an amino acid, such as any of the molecules observed in the lucky experiment, is called an **aminoacyl tRNA**. More recent research has shown that enzymes called **aminoacyl tRNA synthetases** are responsible for catalyzing the addition of amino acids to tRNAs. For each of the 20 major amino acids, there is a different aminoacyl synthetase and one or more tRNAs. Figure 16.10, page 348, summarizes how amino acids are transferred from aminoacyl tRNA synthetases to tRNAs.

What happens to the amino acid that is carried by an aminoacyl tRNA? Biologists who tracked the fate, over time, of radioactive leucine molecules that were attached to tRNAs discovered that the amino acids are transferred from aminoacyl tRNAs to proteins. The data that support this conclusion are

(a) **Hypothesis 1:** Amino acids interact directly with mRNA codons.



(b) **Hypothesis 2:** Adapter molecules hold amino acids and interact with mRNA codons.

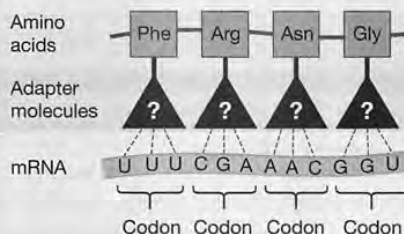


FIGURE 16.9 Two Hypotheses for How mRNA Codons Interact with Amino Acids

(a) An early hypothesis suggested that amino acids and mRNA codons interact directly. (b) The adapter-molecule hypothesis predicts that mRNA codons do not interact directly with the amino acids being added to a growing polypeptide. Instead, an intermediary adapter molecule carries an amino acid and aligns it with the appropriate mRNA codon.

HOW AMINO ACIDS ARE LOADED ONTO tRNAs

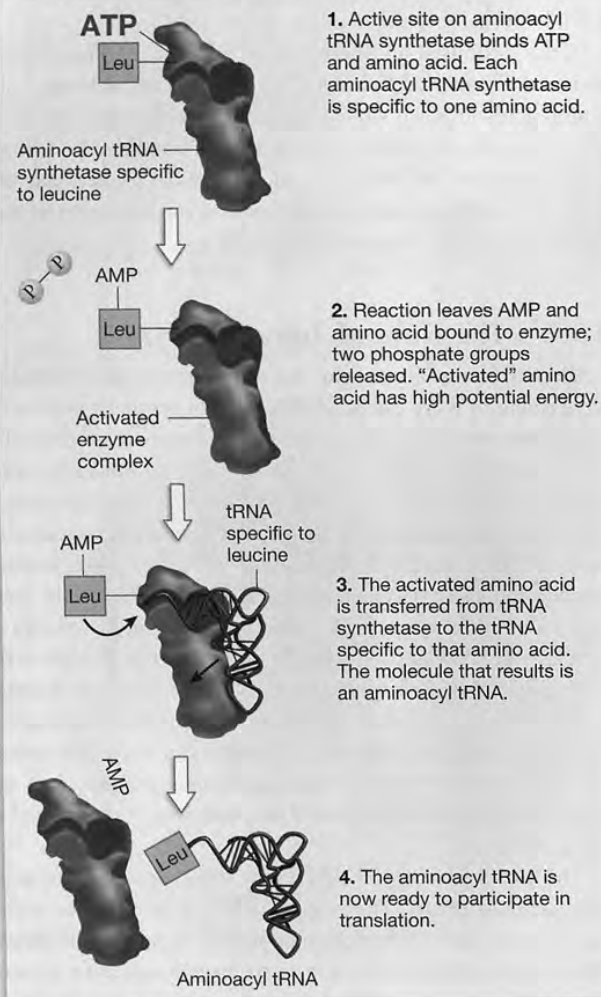


FIGURE 16.10 Aminoacyl tRNA Synthetases Load Amino Acids onto the Appropriate tRNA

An aminoacyl tRNA synthetase—in this case, for leucine—becomes activated in a reaction that requires ATP. Activated aminoacyl tRNA synthetases transfer an amino acid to a tRNA, producing an aminoacyl tRNA.

shown in Figure 16.11. The graph's message is that radioactive amino acids are lost from tRNAs and incorporated into polypeptides synthesized in ribosomes. These results inspired the use of *transfer* in tRNA's name. The experiment also confirmed that aminoacyl tRNAs act as the interpreter in the translation process. tRNAs are Crick's adapter molecules.

Connecting Structure with Function

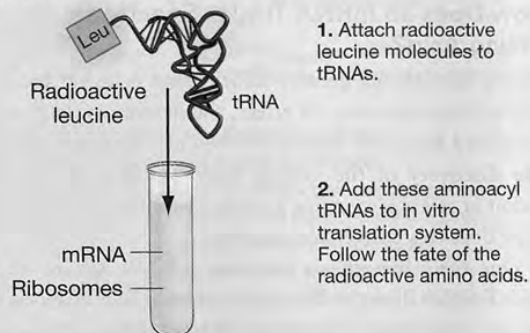
Transfer RNAs serve as the chemical go-betweens that allow amino acids to interact with an mRNA template. But precisely how does this connection occur?

Question: Are amino acids that are attached to tRNA molecules transferred to proteins?

Hypothesis: Aminoacyl tRNAs transfer amino acids to growing polypeptides.

Null hypothesis: The amino acids held on tRNAs are not involved in protein synthesis.

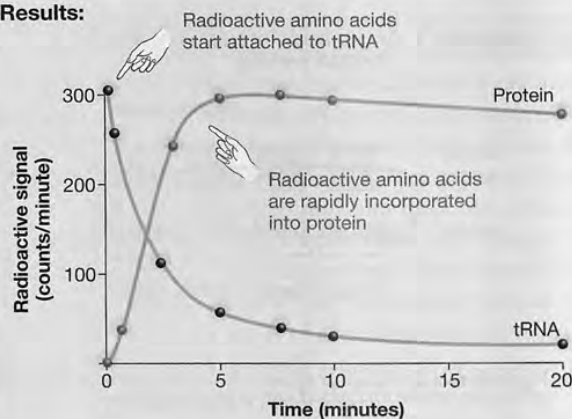
Experimental setup:



Prediction: Radioactive amino acids will be found in proteins.

Prediction of null hypothesis: Radioactive amino acids will not be found in proteins.

Results:

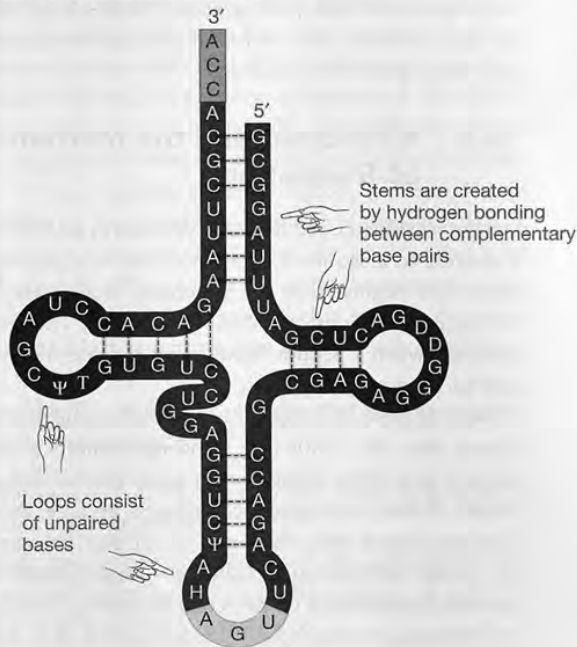


Conclusion: Aminoacyl tRNAs transfer amino acids to growing polypeptides.

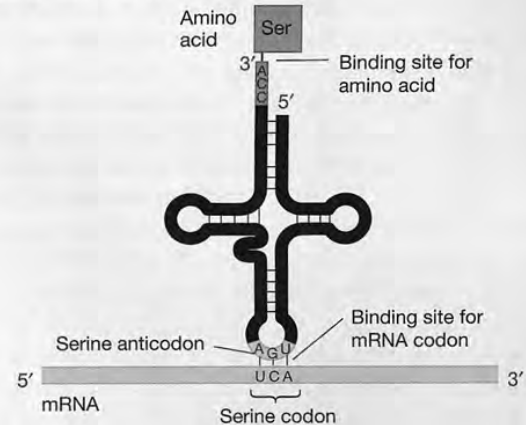
FIGURE 16.11 Experimental Evidence That Amino Acids Are Transferred from tRNAs to Proteins

This question was answered by research on tRNA's molecular structure. The initial studies established the sequence of nucleotides in various tRNAs, or what is termed their primary structure. Transfer RNAs are relatively short, ranging from 75 to 85 nucleotides in length. When biologists studied the primary sequence closely, however, they noticed that certain parts of the molecules can form secondary structures. Specifically, some of the bases in the tRNA molecule can form hydrogen bonds with complementary bases in a different region of the

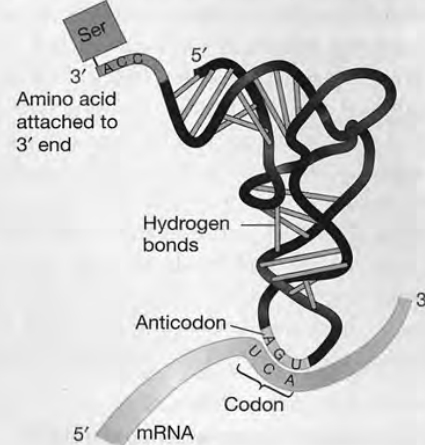
(a) Secondary structure of tRNA



(b) Early model of tRNA function



(c) Revised model incorporating tertiary structure of tRNA

**FIGURE 16.12 The Structure of Transfer RNA**

(a) The secondary structure of tRNA resembles a cloverleaf. (b) If amino acids are attached to the 3' end of a tRNA with an anticodon appropriate for that amino acid, then the tRNA anticodon will form a complementary base pair with the mRNA codon. (c) Recent data indicate that tRNAs have an L-shaped tertiary structure.

same molecule. As a result, the entire tRNA molecule could assume the cloverleaf shape illustrated in Figure 16.12a. The stems in the cloverleaf are produced by complementary base pairing between different portions of the molecule, while the loops consist of single-stranded RNA.

Two aspects of this secondary structure proved especially interesting. The CCA sequence at the 3' end of the molecule offered a binding site for amino acids, while the triplet on a loop on the opposite end of the cloverleaf could serve as an anticodon. An **anticodon** is a set of three nucleotides that forms base pairs with the mRNA codon. Figure 16.12b depicts an early model of how the anticodon of a tRNA connects an mRNA codon with the appropriate amino acid.

This model had to be modified, however, when X-ray crystallography studies revealed the tertiary structure of tRNAs. Recall from Chapter 3 that the tertiary structure of a molecule is defined by the three-dimensional arrangement of its atoms and is usually a product of folding. According to the X-ray crystallography data, the cloverleaf structure folds over to produce an L-shaped molecule (Figure 16.12c). All of

the tRNAs in a cell have the same L-shaped structure, but each has a distinct anticodon and attached amino acid. The tertiary structure of tRNAs is important because it results in a precise separation between the anticodon and the attached amino acid. As we'll see, this separation is key to the positioning of the amino acid and the anticodon in the ribosome.

How Many tRNAs Are There?

When research succeeded in characterizing all of the different types of tRNAs available in cells, a paradox arose. According to the genetic code introduced in Chapter 15, the 20 amino acids found in proteins are specified by 61 different mRNA codons. Instead of containing 61 different tRNAs with 61 different anticodons, though, most cells contain only about 40. How can all 61 mRNA codons in the genome be translated with only two-thirds of the tRNAs required?

To resolve this paradox, Francis Crick proposed what is known as the **wobble hypothesis**. To understand his logic, recall from Chapter 15 that many amino acids are specified by more than one codon. Further, recollect that codons for the same

amino acid tend to have the same nucleotides at the first and second positions but a different nucleotide at the third position. For example, both of the codons CAA and CAG code for the amino acid glutamine. (Codons are always written in the 5' → 3' direction.) Surprisingly, experimental data have shown that a tRNA with an anticodon of GUU can base pair with both CAA and CAG in mRNA. (Anticodons are always written in the 3' → 5' direction.) The GUU anticodon matches the first two bases (C and A) in both codons, but the U in the third position forms a nonstandard base pair with a G in the CAG codon.

Crick proposed that inside the ribosome, tRNAs with nonstandard base pairing at the third position can still bind successfully to an anticodon. If so, it would allow a limited flexibility, or “wobbling,” in the base pairing. According to the wobble hypothesis, a nonstandard base pair in the third position—such as G-U—is acceptable as long as it does not change the amino acid for which the codon codes.

Follow-up research has shown that certain tRNAs are especially prone to pairing with more than one mRNA codon. These tRNAs have a nitrogen-containing base called *inosine* in the third position of their anticodon. (Inosine is derived from adenine by a chemical modification that occurs after the tRNA is produced.) During protein synthesis, an inosine in the anticodon can base pair with a A, U, or C in an mRNA codon. For example, the tRNA with the anticodon GAI carries the amino acid leucine and binds to four of the six codons that specify leucine: CUA, CUC, CUU, and UUA. (The UUA is possible because the G in the anticodon sometimes pairs with U in the mRNA. This interaction between

anticodons and codons is one of the only known exceptions to the standard base-pairing rules.) In this way, “wobble” in the third position of a codon allows just 40 or so tRNAs to bind to all 61 mRNA codons.

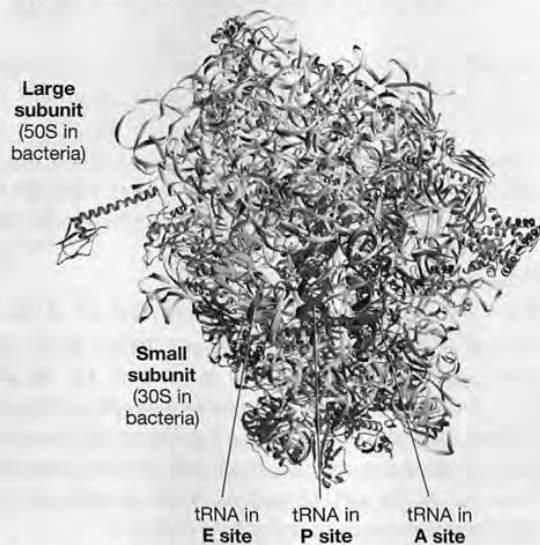
16.5 Ribosomes and the Mechanism of Translation

During translation, the sequence of bases in an RNA message is converted to a sequence of amino acids in a polypeptide. The conversion begins when the anticodon of a tRNA that carries an amino acid binds to a codon in mRNA. The conversion is complete when a peptide bond forms between that amino acid and the growing polypeptide chain.

Both of these events occur inside a ribosome. Biologists have known since the 1930s that ribosomes contain a considerable amount of protein along with a great deal of **ribosomal RNA (rRNA)**. When high-speed centrifuges became available, researchers found that ribosomes could also be separated into two major substructures, called the *large subunit* and *small subunit*. Each subunit consists of a complex of RNA molecules and proteins. For example, the large subunit in *E. coli* contains two different rRNA molecules and 34 proteins. More recent work has shown that the small subunit holds the mRNA in place during translation and that peptide bond formation takes place in the large subunit.

The three-dimensional model of the ribosome in **Figure 16.13a** shows how all of the molecules and reactions required for translation fit together. Note that during protein synthesis,

(a) Ribbon model of ribosome during translation



(b) Diagram of ribosome during translation

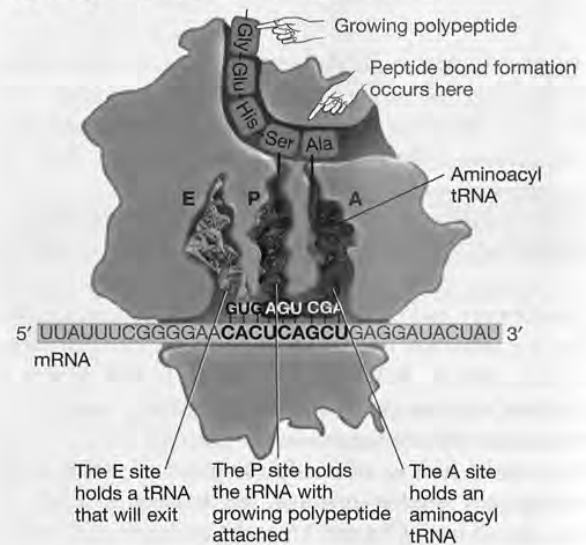


FIGURE 16.13 The Structure of the Ribosome

(a) During translation, three tRNAs line up side by side inside the ribosome. (b) Different events occur at each of the three sites inside the ribosome where tRNAs are found. **EXERCISE** Circle and label the active site in the ribosome.

three distinct tRNAs are lined up inside the ribosome. All three are bound to their corresponding mRNA codon at the base of the structure. The tRNA that is on the right (colored red) in the figure carries an amino acid. This tRNA's position in the ribosome is called the **A site**, for acceptor or aminoacyl. The tRNA that is in the middle (green) holds the growing polypeptide chain and occupies the **P site**, for peptidyl, inside the ribosome. (Think of "P" for "peptide-bond formation.") The left-hand (yellow) tRNA no longer has an amino acid attached and is about to leave the ribosome. It occupies the ribosome's **E site**, for exit. **Figure 16.13b** summarizes what happens to each tRNA inside the ribosome. Because all tRNAs have the same secondary and tertiary structure, they all fit equally well in the A, P, and E sites.

The ribosome is a molecular machine that synthesizes proteins in a three-step sequence:

1. An aminoacyl tRNA diffuses into the A site; its anticodon binds to a codon in mRNA.
2. A peptide bond forms between the amino acid held by the aminoacyl tRNA in the A site and the existing polypeptide, which is held by a tRNA in the P site.
3. The ribosome moves ahead, and all three tRNAs move one position down the line. The tRNA in the E site exits; the tRNA in the P site moves to the E site; and the tRNA in the A site switches to the P site.

The protein that is being synthesized grows by one amino acid each time this three-step sequence repeats. The process occurs up to 20 times per second in bacterial ribosomes and about 2 times per second in eukaryotic ribosomes.

This introduction to how tRNAs, mRNAs, and ribosomes interact during protein synthesis leaves a number of key questions unanswered, however. How do mRNAs and ribosomes get together to start the process? Once protein synthesis is under way, how is peptide bond formation catalyzed inside the ribosome? And how does protein synthesis conclude when the ribosome reaches the end of the message? Let's consider each question in turn.

Initiation

To translate an mRNA, a ribosome must begin at a specific point in the message, translate the mRNA up to the message's termination codon, and then stop. Biologists call these three phases of protein synthesis **initiation**, **elongation**, and **termination**, respectively. One key to understanding initiation is to recall, from Chapter 15, that the codon AUG is found near the 5' end of all mRNAs and codes for the amino acid methionine. The presence of this start codon is an aspect of initiation that is common to both bacteria and eukaryotes.

Figure 16.14 shows how translation gets under way in bacteria. The process begins when a section of rRNA in the small subunit of a ribosome binds to a complementary sequence on an mRNA. The mRNA region is called the **ribosome binding site** or **Shine-Dalgarno sequence**, after the biologists who discovered it. The site is about six nucleotides upstream from the AUG start codon. It consists of all or part of the bases 5' AGGAGGU 3'. The complementary sequence in the rRNA of the small subunit reads 3' UCCUCCA 5'. The interaction between the small subunit and the message is mediated by proteins called **initiation factors**. In eukaryotes, initiation factors bind to the 5' cap on mRNAs and guide it to the ribosome.

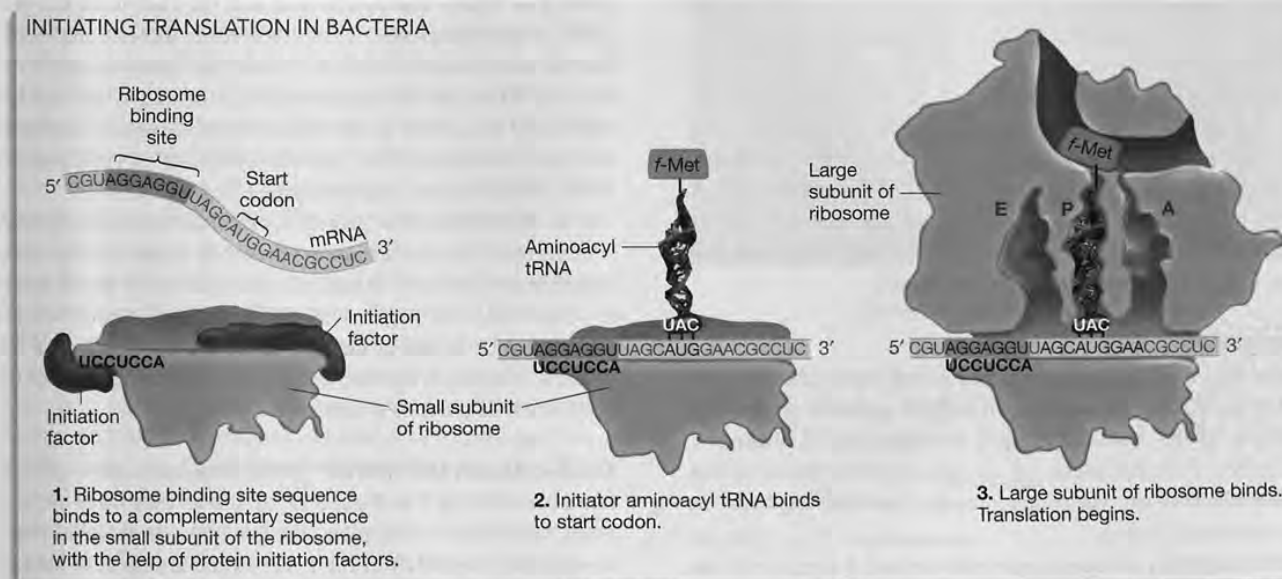


FIGURE 16.14 The Initiation Phase of Translation in Bacteria

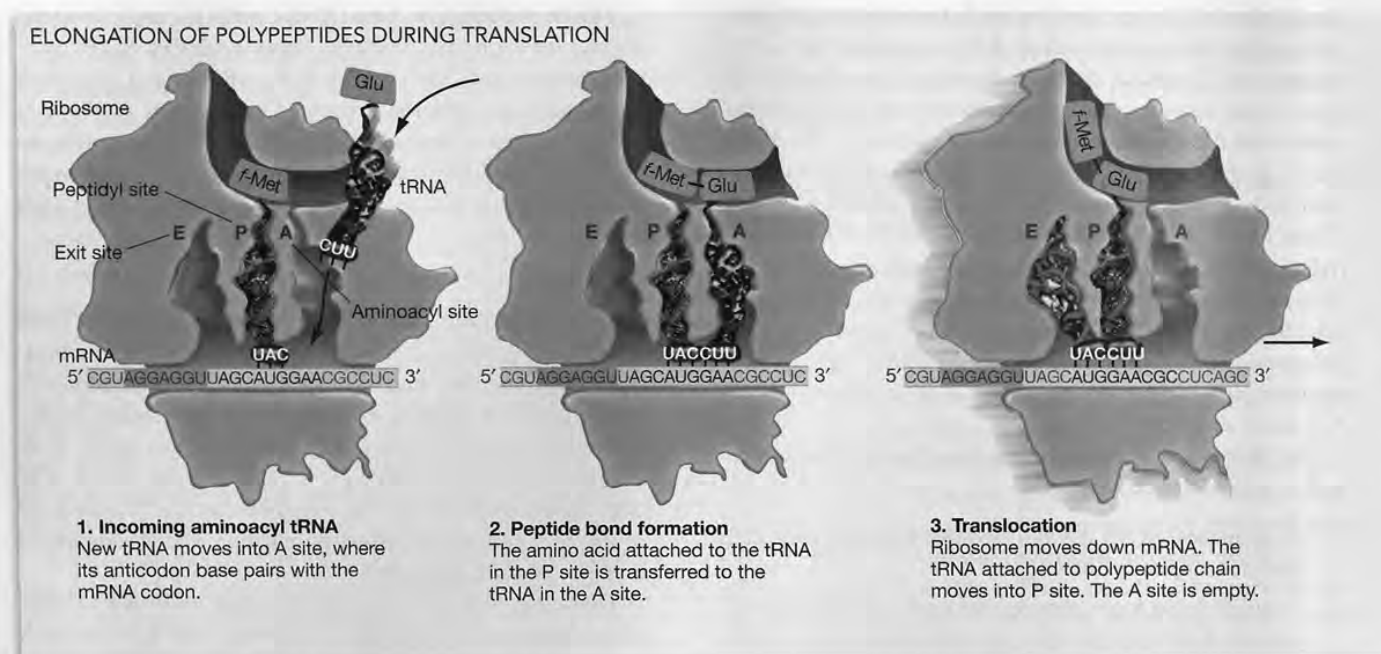


FIGURE 16.15 The Elongation Phase of Translation

Once the Shine-Dalgarno sequence has attached to the small ribosomal subunit, an aminoacyl tRNA bearing a modified form of methionine called *N*-formylmethionine (abbreviated *f*-met) binds to the AUG start codon. Initiation is complete when the large subunit joins the complex. When the ribosome is completely assembled, the tRNA bearing *f*-met occupies the P site.

To summarize, initiation is a three-step process in bacteria: (1) The mRNA binds to the small subunit of the ribosome, (2) the initiator aminoacyl tRNA binds to the start codon, and (3) the large subunit of the ribosome completes the complex.

It's important to note that in bacteria a single mRNA may contain several start codons, termination codons, and ribosome binding sites and thus code for several distinct proteins. Such mRNAs are said to be **polycistronic**. In polycistronic messages, an AUG start sequence near the 5' end of the mRNA is followed by a termination signal farther along, which in turn is followed by an AUG start sequence for a second protein and then a second termination signal, and so on.

Elongation

At the start of elongation, the E and A sites in the ribosome are empty of tRNAs. As a result, an mRNA codon is exposed at the base of the A site. As step 1 in Figure 16.15 illustrates, elongation proceeds when an aminoacyl tRNA binds to the codon in the A site via complementary base pairing between anticodon and codon.

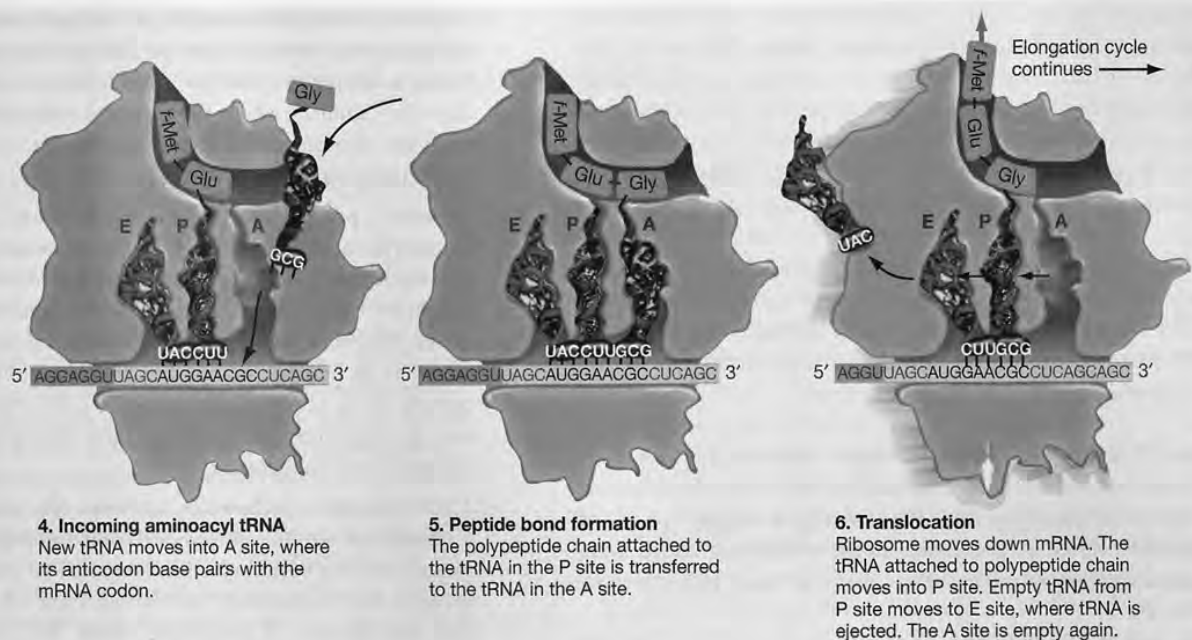
When a tRNA occupies both the P site and A site, the amino acids they hold are placed in the ribosome's active site. This is where peptide bond formation—the essence of protein

synthesis—occurs. Peptide bond formation is considered one of the most important reactions that takes place in cells, because manufacturing proteins is among the most fundamental of all cell processes. The question is, how does it happen?

Is the Ribosome an Enzyme or a Ribozyme? Because ribosomes contain both proteins and RNA, researchers had argued for decades over whether the active site consisted of protein or RNA. The debate was not resolved until the year 2000, when researchers completed three-dimensional models that were detailed enough to view the structure of the active site. These models confirmed that the active site consists entirely of ribosomal RNA. Based on these results, biologists are now convinced that protein synthesis is catalyzed by RNA. The ribosome is a ribozyme—not an enzyme.

The observation that protein synthesis is catalyzed by RNA is important, because it supports the RNA world hypothesis introduced in Chapter 4. Recall that proponents of this hypothesis claim that life began with RNA molecules and that the presence of DNA and proteins in cells evolved later. If the RNA world hypothesis is correct, then it is logical to observe that the production of proteins is catalyzed by RNA.

Moving Down the mRNA What happens after a peptide bond forms? Step 2 in Figure 16.15 shows that when peptide bond formation is complete, the polypeptide chain is transferred from the tRNA in the P site to the amino acid held by the tRNA in the A site. Step 3 shows the process called **translocation**, which occurs when the ribosome moves down



the mRNA in the 5' → 3' direction. Translocation does several things: It moves the empty tRNA into the E site; it moves the tRNA containing the growing polypeptide into the P site; and it opens the A site and exposes a new mRNA codon. If the E site is occupied when translocation occurs, the tRNA there is ejected into the cytoplasm.

The three steps in elongation—(1) arrival of aminoacyl tRNA, (2) peptide bond formation, and (3) translocation—repeat down the length of the mRNA. By adding specific molecules to *in vitro* translation systems, researchers have confirmed that each elongation cycle depends on an input of energy from several GTP molecules as well as assistance from proteins called **elongation factors**. Further, recent three-dimensional models of ribosomes in various stages of the translation sequence show that the machine as a whole is highly dynamic during the process. The ribosome constantly changes shape as tRNAs come and go and catalysis and translocation occur. Its structure changes in conjunction with its multipart function: coordinating interactions between tRNAs and an mRNA, and catalyzing peptide bond formation.

Electron microscopy has also confirmed that once elongation is under way, additional ribosomes bind to the start site in the same message and initiate translation. As this process continues, strings of ribosomes called **polyribosomes** assemble along an mRNA. Each ribosome synthesizes a copy of the protein (Figure 16.16). Even though it may take a minute or more for a single ribosome to produce a large polypeptide, the presence of polyribosomes can quicken the overall pace of protein production. Polyribosomes are observed routinely in both bacteria and eukaryotes.

Termination

How does protein synthesis end? To answer this question, recall from Chapter 15 that the genetic code includes three stop codons: UAA, UAG, and UGA. In most cells, no aminoacyl tRNA has an anticodon that binds to these sequences. When

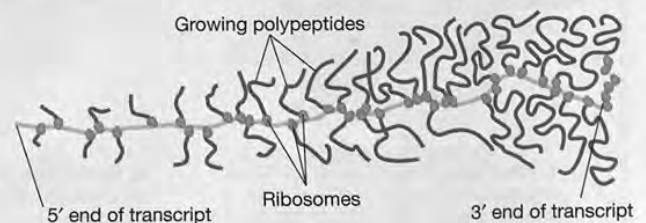
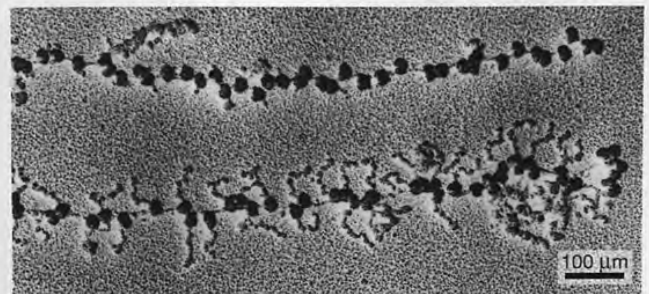


FIGURE 16.16 Polyribosomes

In both bacteria and eukaryotes, many ribosomes synthesize proteins from the same mRNA at the same time. The resulting structures are called polyribosomes.

translocation opens the A site and exposes one of the stop codons, a protein called a **release factor** fills the A site (Figure 16.17). Release factors are proteins that fit tightly into the A site because their size, shape, and electrical charge are extremely similar to the same properties in a tRNA. Release factors do not carry an amino acid, however. When a release factor occupies the A site, the active site catalyzes the hydrolysis of the bond linking the tRNA in the P site with the polypeptide chain. This reaction frees the polypeptide, and it is released from the ribosome. The ribosome then separates from the mRNA, and the two ribosomal subunits dissociate. The subunits are ready to attach to the start codon of another message and start translation anew.

Post-Translational Modifications Proteins are not fully formed and functional when termination occurs. From earlier chapters, it should be clear that most proteins go through an extensive series of processing steps, collectively called **post-translational modification**, before they are ready to go to work in a cell:

- Recall from Chapter 3 that a protein's function depends on its shape and that a protein's shape depends on how it folds. Folding actually begins during elongation, long before termination occurs and ribosomes disassemble. Although it occurs spontaneously, in the sense that no energy input is

required, folding is frequently speeded up by proteins called **molecular chaperones**. Recent data have shown that, in some bacteria, chaperone proteins actually bind to the ribosome near the "tunnel" where the growing polypeptide emerges from the ribosome. This result suggests that folding occurs as the polypeptide is emerging from the ribosome.

- Chapter 7 pointed out that many eukaryotic proteins are extensively modified after they are synthesized. For example, small chemical groups may be added to proteins in the organelles called the rough endoplasmic reticulum and the Golgi apparatus. Some proteins receive a sorting signal that serves as an address and ensures that the molecule will be carried to the correct location in the cell. Certain proteins are also augmented with sugar or lipid groups that are critical for normal functioning.
- Many proteins are altered by enzymes that add or remove a phosphate group. Phosphorylation (addition of phosphate) and dephosphorylation (removal of phosphate) of proteins were introduced in Chapter 3 and Chapter 8. Recall that because a phosphate group has two negative charges, adding or removing a phosphate group causes major changes in the shape and chemical reactivity of proteins. These changes have a dramatic effect on the protein's activity—often switching it from an inactive state to an active state or vice versa.

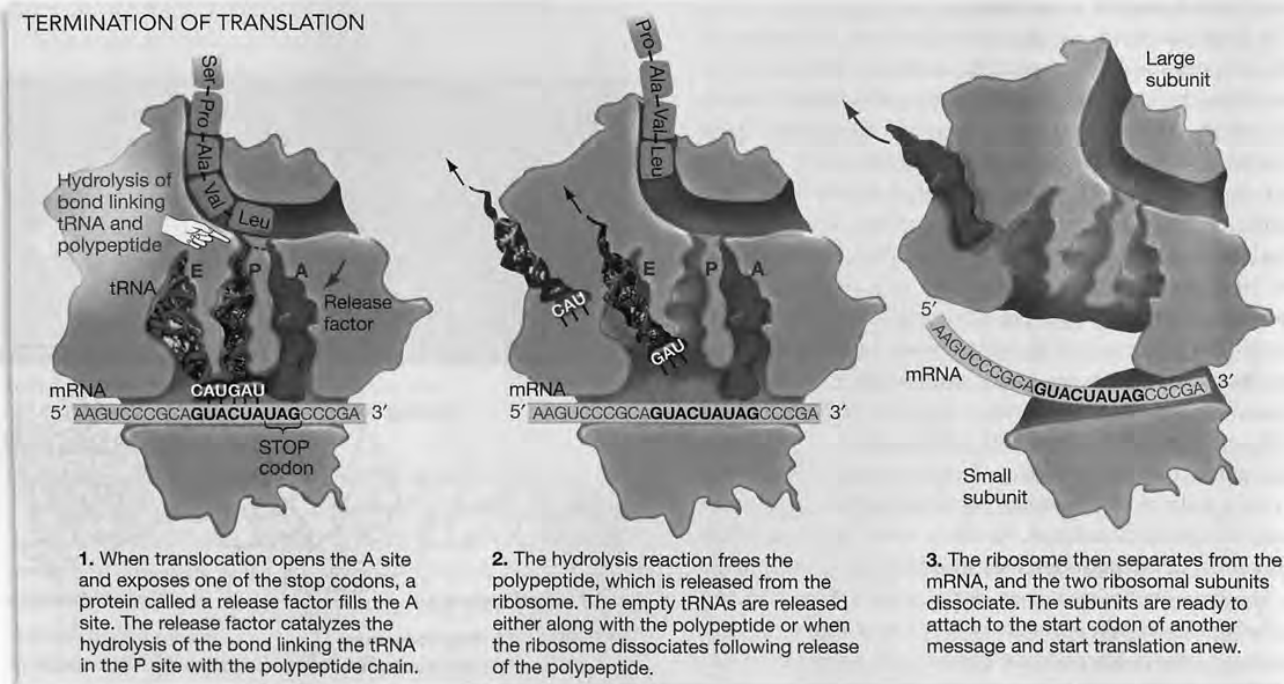


FIGURE 16.17 The Termination Phase of Translation

The overall message is that the manufacture of a completely functional protein depends not only on translation in the ribosome but also on a wide array of other molecules and events that take place throughout the cell.

✓CHECK YOUR UNDERSTANDING

In bacteria, translation begins when the ribosome binding site on an mRNA binds to an rRNA sequence in the small subunit of the ribosome. The initiator aminoacyl tRNA binds to the start codon in the mRNA, and the large subunit of the ribosome attaches to the small subunit to complete the ribosome.

In both bacteria and eukaryotes, polypeptides elongate when the appropriate aminoacyl tRNA enters the A site. A peptide bond forms between the amino acid held by that tRNA in the A site and the polypeptide held by the tRNA in the P site. The ribosome then moves down the mRNA one codon, and the process occurs again. Translation ends when the ribosome reaches a stop codon. You should be able to (1) diagram a ribosome during translation initiation, elongation, and termination; (2) label each major structure in the diagrams; and (3) describe the function of each major structure.

16.6 The Molecular Basis of Mutation

This chapter has focused on how the information in DNA is transcribed to RNA and then translated into a functioning protein.

Although biologists continue to work on issues such as how basal transcription factors in eukaryotes interact with RNA polymerase and how ribosomes change shape during translation, the molecular mechanisms responsible for information flow in the cell, from DNA to RNA to proteins, are now considered fairly well understood.

Before we go on to consider how transcription and translation are regulated in bacteria and eukaryotes in the next two chapters, it will be helpful to tie together some major points covered in the last three chapters by analyzing the molecular

basis of mutation. A **mutation** is any change in an organism's DNA. It is a change in a cell's information archive—a change in its genotype. What are the consequences of mutation?

The molecular machinery introduced in this chapter faithfully transcribes and translates mutations in DNA sequences. The result is the production of novel types of proteins. In this way, changes in DNA lead to changes in molecules that affect the organism's phenotype.

Figure 16.18 shows how a common type of mutation occurs. If DNA polymerase mistakenly inserts the wrong base as it synthesizes a new strand of DNA, and if proofreading by DNA polymerase and the mismatch repair system fail to correct the mismatched base, a change in the sequence of bases in DNA results. A single base change such as this is called a **point mutation**.

What happens when point mutations are transcribed and translated? To answer this question, consider the first point mutation ever described. This mutation occurs in the human gene for the protein hemoglobin. Recall that hemoglobin is abundant in red blood cells and that it carries oxygen to tissues.

Figure 16.19a, page 356, shows a small section of DNA sequence from the normal gene for hemoglobin, along with the same sequence from the mutant allele found in individuals who suffer from sickle-cell disease. The mutant form has a thymine in place of an adenine at the second position in the sixth codon specified by this gene. A glance back at the genetic code in Chapter 15 confirms why this point mutation is significant: During protein synthesis, the mutant codon specifies valine instead of glutamic acid in the amino acid chain of hemoglobin. Point mutations that cause changes in the amino acid sequence of proteins are called **missense mutations** or **replacement mutations**. In the case of hemoglobin, the single change in primary structure causes the protein to crystallize when oxygen levels in the blood are low. When hemoglobin crystallizes, it causes red blood cells to become sickle shaped (Figure 16.19b). The misshapen cells get stuck in blood vessels, so nearby cells are starved for oxygen. The result is intense pain and anemia.

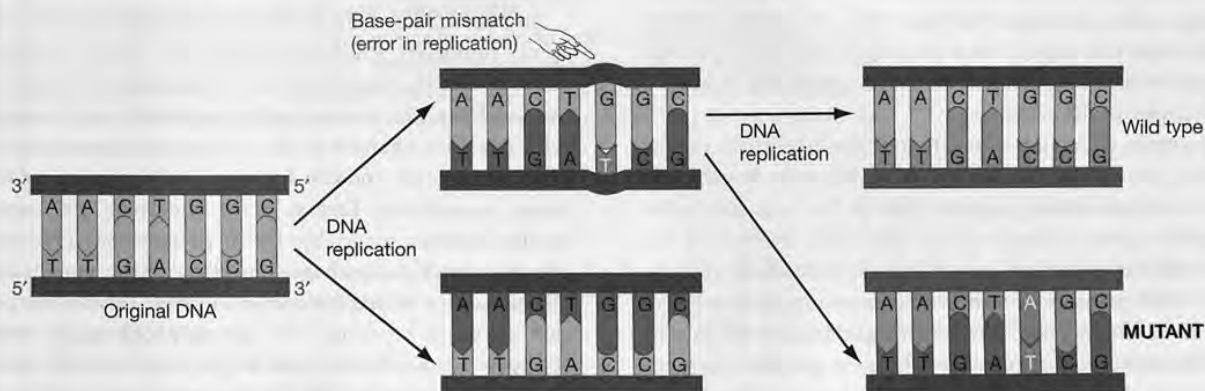
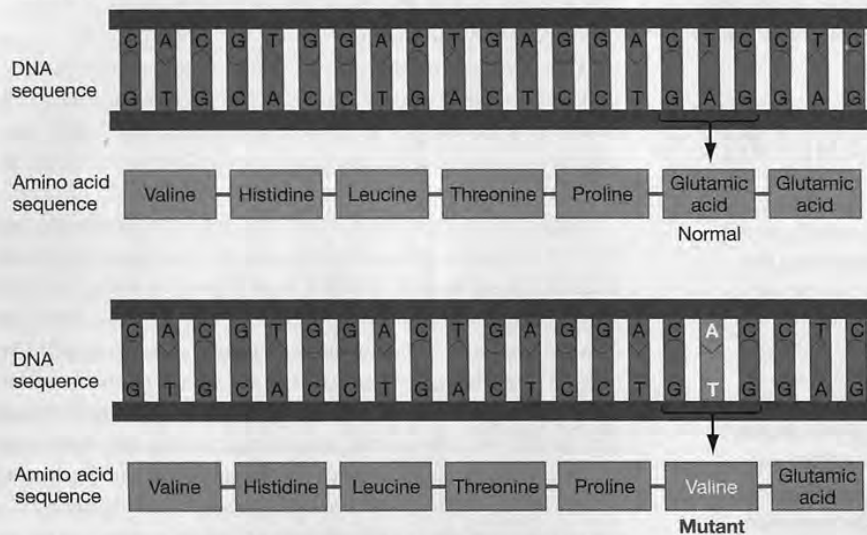


FIGURE 16.18 Unrepaired Mistakes in DNA Synthesis Lead to Point Mutations

QUESTION Why is it logical that the type of mutation illustrated here is termed a point mutation?

(a) DNA point mutation can lead to a different amino acid sequence.



(b) Phenotype

**FIGURE 16.19 Sickle-Cell Disease Results from a Point Mutation in the Gene for Hemoglobin**

(a) These normal and mutant genes that code for hemoglobin differ by a single nucleotide. The sequences shown are just a small portion of the gene. (b) The mutation in part (a) changes the primary sequence of the gene product. The change in amino acid sequence causes hemoglobin molecules to crystallize when oxygen levels in the blood are low. As a result, red blood cells sickle and get stuck in small blood vessels. **EXERCISE** The mRNA codon GUG specifies the valine at the start (left) of this protein. Label the template strand and the non-template strand of DNA, and indicate their 5' and 3' ends.

Missense mutations such as that found in the sickling allele can be **deleterious**, meaning they reduce an individual's fitness—the ability to survive and reproduce. But missense mutations may also be beneficial. For example, the sickling allele confers a fitness advantage in regions of the world where the life-threatening disease malaria is common. The advantageous effect is possible because the parasite that causes malaria infects red blood cells. If an individual has one normal hemoglobin allele and one sickling allele, red blood cells that become infected with the parasite tend to sickle while others function normally. The deformed cells are quickly destroyed by the body, eliminating the parasitic cells inside the blood cells. As a result, individuals who have one copy of the mutant gene have lower parasite loads and tend to be much healthier than people who have two normal copies of the gene.

In contrast to having a negative or positive effect on the organism, many point mutations have virtually no effect at all. To understand why, suppose that at the next site in the hemoglobin gene of Figure 16.19 (the third position in the same sixth codon), a thymine had been substituted for the cytosine. This point mutation would have no consequence. Why? Both GAA and GAG specify glutamic acid. In this case, the amino acid sequence of the gene product does not change even though the DNA sequence is altered due to mu-

tation. This type of alteration in the base sequence is called a **silent mutation**. Silent mutations are said to be **neutral** in their effect on an individual's fitness. It is possible for point mutations to be silent and neutral with respect to fitness because of the redundancy in the genetic code.

The take-home message from this analysis of hemoglobin, sickle-cell disease, and malaria is simple: It is possible for point mutations to be deleterious, beneficial, or neutral. Figure 16.20a summarizes the types of point mutations that have been documented and reviews their consequences for the amino acid sequences of proteins.

In addition to documenting various types of point mutations, biologists study mutations that involve larger-scale changes in the composition of chromosomes. Chapter 9 introduced the phenomena called polyploidy and aneuploidy, which involve changes in the number of chromosome complements present and the addition or deletion of a chromosome, respectively. Just as point mutations are caused by random errors made by DNA polymerase, chromosome changes result from chance mistakes in the partitioning of chromosomes during meiosis or mitosis. In addition, errors that occur in crossing over during prophase of meiosis I (Chapter 12) can lead to gene duplications or deletions, as illustrated in Figure 16.20b. Finally, when accidental breaks in

(a) Types of point mutations

Definition		Example	Consequence
		Original DNA sequence — TAT TGG CTA GTA CAT Original polypeptide — Tyr — Trp — Leu — Val — His	
Silent	Change in nucleotide that does not change amino acid specified by codon	TAC TGG CTA GTA CAT Tyr — Trp — Leu — Val — His	Change in genotype but no change in phenotype
Missense (Replacement)	Change in nucleotide that changes amino acid specified by codon	TAT TGT CTA GTA CAT Tyr — Cys — Leu — Val — His	Change in primary structure of protein
Nonsense	Change in nucleotide that results in early stop codon	TAT TGA CTA GTA CAT Tyr — STOP	Premature termination—polypeptide is truncated
Frameshift	Addition or deletion of a nucleotide	TAT TCG GCT AGT ACA T Tyr — Ser — Ala — Ser — Thr	Reading frame is shifted—massive missense

(b) Other types of mutations

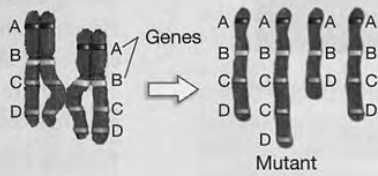
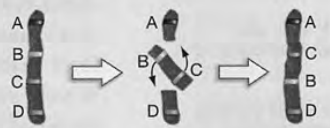
Definition	Example	Consequence
Gene duplication		Produces an extra copy or deletion of one or more genes. Families of related genes arise by gene duplication.
Chromosome inversion		Changes gene order along chromosome. Other types of chromosome breaks can lead to deletion or addition of chromosome segments.

FIGURE 16.20 A Mutation Is a Change in an Organism's DNA

(a) Point mutations in DNA produce an array of effects on the amino acid sequences of proteins, ranging from no change to extensive change. (b) Many mutations involve large-scale changes in the composition of chromosomes. In addition to gene duplication and chromosome inversions, Chapter 9 introduced polyploidy (multiple chromosome complements), trisomy (additional chromosomes), and aneuploidy (loss or gain of a chromosome).

chromosomes occur, the chromosome segments may become flipped and rejoin—a phenomenon known as a **chromosome inversion** (Figure 16.20b).

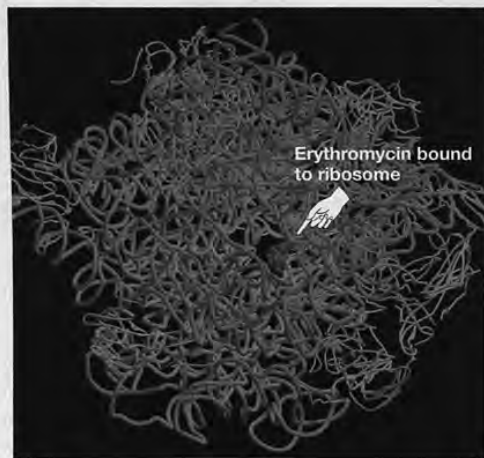
When changes in an organism's genotype are transcribed and translated, they may result in changes in its phenotype. In

this way, mutations produce new alleles and new traits. At the level of populations, mutations furnish the heritable variation that Mendel and Morgan analyzed and that makes evolution possible. At the level of individuals, mutations can cause disease or death or lead to increases in fitness.

ESSAY Antibiotics That Poison the Ribosome

Antibiotics are molecules that kill bacteria or slow their growth. Since their initial discovery by biomedical researchers in the 1930s, their use has saved tens of millions of human lives that were threatened by bacterial infections. It is no exaggeration to claim that the era of modern medicine began with the widespread availability of inexpensive antibiotics.

(a) Erythromycin binds to ribosomes and plugs the exit tunnel for the polypeptide.



(b) Mode of action of some antibiotics that poison translation

Antibiotic	Mode of Action
Chloramphenicol	Binds at the active site in the large subunit of ribosome
Clindamycin	Binds to A and P sites and prevents correct positioning of tRNAs
Erythromycin, carbomycin A, clarithromycin, roxithromycin, spiramycin, tylosin, azithromycin	Plugs the tunnel where growing polypeptide exits ribosome
Puromycin	Enters A site and bonds to growing polypeptide, but blocks addition of more amino acids
Spectinomycin	Inhibits translocation of ribosome
Streptomycin, paromomycin	Binds to the small subunit of ribosome and prevents accurate reading of the mRNA
Tetracycline	Prevents aminoacyl tRNAs from entering A site in ribosome

FIGURE 16.21 Many Antibiotics Work by Disrupting Translation in Bacteria

[(a) "Structural basis for the interaction of antibiotics with peptidyl transferase centre in eubacteria: by Schlunzen et al." Figure 5 in *Nature* (2001) 413: 814–821 by Schlunzen et al.]

Many of the antibiotics that physicians prescribe today are produced naturally by soil-dwelling fungi. These organisms compete with bacterial cells for space and resources. For a fungus, the ability to synthesize antibiotics is an adaptation that

Recent research on antibiotics has focused on two seemingly distinct topics: the evolution of bacteria that resist them and the molecular mechanism of drug action.

helps individuals destroy competitors and increase their access to nutrients. Antibiotics are weapons in chemical warfare that has gone on among competing soil-dwelling organisms for millions of years. Pharmaceutical companies still search for new antibiotics by studying naturally occurring fungi in the soil.

Recent research on antibiotics has focused on two seemingly distinct topics: the evolution of bacteria that resist them and the molecular mechanism of drug action. **Antibiotic resistance** is the ability to grow and reproduce efficiently in an environment containing antibiotics. But as often happens in biological science, efforts to answer questions that seem unrelated produce closely connected answers.

How are evolutionary and molecular studies linked? Early research showed that several medically important antibiotics work by inhibiting translation in bacterial cells. In the presence of certain drugs, bacterial ribosomes stop working, protein synthesis ceases, and the targeted cell dies. Fortunately, bacterial and eukaryotic ribosomes are different enough at the molecular level that ribosome-poisoning antibiotics tend not to disrupt protein synthesis in humans. This is why antibiotics that target bacterial ribosomes have few side effects when administered to humans as a drug.

To see how evolutionary and molecular studies of ribosomes intersect, consider the latest three-dimensional models of bacterial ribosomes from bacteria treated with the antibiotic erythromycin. As **Figure 16.21a** shows, the drug binds to the tunnel where growing polypeptide chains leave the ribosome. When erythromycin binds to this site, it inhibits translation and kills the cell. A number of bacterial strains have evolved resistance to this antibiotic, however. When researchers analyzed the structure of ribosomes from drug-resistant populations of bacteria, they found that the ribosomes contained novel amino acids and rRNA sequences at the site where erythromycin binds. The altered amino acids and bases make erythromycin much less likely to bind to the ribosome. As a result, the drug no longer inhibits translation effectively.

Similar studies have now been done on a wide array of antibiotics. The results have documented the binding site and mode of action of many commonly prescribed drugs (Figure 16.21b). In many cases, investigators have also been able to confirm that drug-resistant strains of bacteria have mutations precisely at these binding sites.

To make sense of these findings, biomedical researchers acknowledge that mutations are constantly creating variations in the makeup of ribosomes within bacterial populations. The researchers hypothesize that extensive use of antibiotics has led to intensive natural selection for cells with mutations that make them resistant to one or more

drugs. In areas where antibiotic use is widespread, the mutant bacteria have higher fitness and come to dominate their population.

Antibiotic resistance is now so pervasive among disease-causing bacteria that some physicians claim that we have entered a post-antibiotic era. Researchers and physicians have already spent several decades in a race to develop new antibiotics that are effective against drug-resistant strains of bacteria. Research on ribosome structure and the mechanisms of translation has taken on an urgent tone. The more we know about the molecular mechanisms of protein synthesis, the better we can run the race against drug-resistant bacteria.

CHAPTER REVIEW

Summary of Key Concepts

The instructions for making and operating a cell are archived in DNA, transcribed into messenger RNA, and then translated into protein.

- After the enzyme RNA polymerase binds to a specific site in DNA with the help of other proteins, it catalyzes the production of an RNA molecule. The base sequence of the RNA produced is complementary to the base sequence of the DNA template strand.

Early experiments established that RNA polymerase begins transcription by binding to promoter sequences in DNA. In bacteria, this binding occurs in conjunction with a regulatory protein called sigma. Sigma recognizes particular sequences within promoters that are located 10 bases and 35 bases upstream from the start of the actual genetic message. These binding sites ensure that the correct DNA strand is transcribed in the correct direction. Eukaryotic promoters are more complex and variable than bacterial promoters, and interact with many more transcription factors.

Web Tutorial 16.1 Transcription

- In eukaryotes, genes consist of coding regions called exons interspersed with noncoding regions called introns. After a eukaryotic mRNA is produced, introns must be spliced out. The ends of the molecule receive a cap and tail.

After transcription of a eukaryotic mRNA is complete, several key events occur. Stretches of noncoding RNA called introns are spliced out by complex molecular machines called spliceosomes, a “cap” signal is added to the 5′ end, and a poly (A) tail is added to the 3′ end. Experiments with modified mRNAs established that both the 5′ cap and the poly (A) tail serve as recognition signals for the translation machinery and protect the message from degradation by ribonucleases.

- Inside ribosomes, mRNAs are translated to proteins via intermediary molecules called transfer RNAs. Transfer RNAs

carry an amino acid and have a three-base-pair anticodon, which binds to an mRNA codon. The amino acid carried by the transfer RNA is then added to the growing protein via formation of a peptide bond.

Experiments with radioactively labeled amino acids confirmed that ribosomes are the site of protein synthesis, and that transfer RNAs (tRNAs) serve as the chemical bridge between the RNA message and the polypeptide product. tRNAs have a predicted secondary structure that resembles a cloverleaf and an L-shaped tertiary structure. One leg of the L contains the anticodon, which forms a base pair with the mRNA codon, while the other leg holds the amino acid appropriate for that codon. Because imprecise pairing—or “wobbling”—is allowed in the third positions of anticodons, only about 40 different tRNAs are required to translate the 61 codons that code for amino acids.

Through intensive study of ribosome structure and function, researchers established that the process of elongating a polypeptide involves three steps: (1) an incoming aminoacyl tRNA occupies a position in the ribosome called the A site; (2) the growing polypeptide chain is transferred from a peptidyl tRNA in the ribosome’s P site to the amino acid bound to the tRNA in the A site, and a peptide bond is formed; and (3) the ribosome is translocated to the next codon on the mRNA, accompanied by ejection of the empty tRNA from the E site. Recent data on the three-dimensional structure of the ribosome has confirmed that peptide bond formation is catalyzed by a ribozyme, not an enzyme.

While translation is in progress, proteins fold into their three-dimensional conformation (tertiary structure), sometimes with the aid of chaperone proteins. Some proteins are targeted to specific locations in the cell by the presence of signal sequences, while others remain inactive until modified by phosphorylation or the removal of certain amino acids.

Web Tutorial 16.2 Translation

■ Many mutations produce changes in the phenotype.

When mutations occur in DNA, the altered sequences are faithfully transcribed and translated into gene products that are part of the organism's phenotype. Depending on the location and type of alteration in DNA and its impact on the re-

sulting RNA or protein product, a mutation can be beneficial, deleterious, or neutral with respect to fitness. Mutations produce novel proteins and RNAs. They are the source of the heritable variation that makes evolution possible.

Questions

Content Review

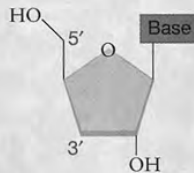
- How did the A site of the ribosome get its name?
 - It is where amino acids are affixed to tRNAs, producing aminoacyl tRNAs.
 - It is where the amino group on the growing polypeptide chain is available for peptide bond formation.
 - It is the site occupied by incoming aminoacyl tRNAs.
 - It is surrounded by α -helices of ribosomal proteins.
- How did the P site of the ribosome get its name?
 - It is where the promoter resides.
 - It is the site where peptide bond formation takes place.
 - It is where peptidyl tRNAs reside.
 - It is the site where the growing polypeptide chain is phosphorylated.
- What is a molecular chaperone?
 - a protein that recognizes the promoter and guides the binding of RNA polymerase
 - a protein that activates or deactivates another protein by adding or removing a phosphate group
 - a protein that is a component of the large ribosomal subunit and that assists with peptide bond formation
 - a protein that helps newly translated proteins fold into their proper three-dimensional configuration
- The three types of RNA polymerase found in eukaryotic cells transcribe different types of genes. What does RNA polymerase II produce?
 - rRNAs
 - tRNAs
 - mRNAs
 - spliceosomes
- What is an anticodon?
 - the part of an RNA message that signals the termination of translation
 - the part of an RNA message that signals the start of translation
 - the part of a tRNA that binds to a complementary codon in mRNA
 - the part of a tRNA that accepts an amino acid, via a reaction catalyzed by tRNA synthetase
- Which of the following questions about transcription was *not* answered by the discovery of the -35 and -10 sequences in bacterial promoters?
 - Does RNA polymerase act alone to initiate transcription, or are other proteins involved?
 - How do sigma proteins bind to RNA polymerase?
 - How does RNA polymerase "know" where to initiate transcription?
 - How does RNA polymerase "know" which strand of DNA acts as the template?

Conceptual Review

- Explain the relationship among eukaryotic promoter sequences, basal transcription factors, and RNA polymerase. Explain the relationship among bacterial promoter sequences, sigma, and RNA polymerase.
- According to the wobble rules, the correct amino acid can be added to a growing polypeptide chain even if the third base in an mRNA codon does not correctly match the corresponding base in a tRNA anticodon. How do the wobble rules relate to the redundancy of the genetic code?
- Why does splicing occur in eukaryotic mRNAs? Where does it occur, and how are snRNPs involved?
- Sketch the structure of a tRNA molecule. Label the anticodon and the CCA sequence where amino acids bind. Explain how a tRNA interacts with tRNA synthetase. Sketch the structure of an aminoacyl tRNA. Explain how an aminoacyl tRNA interacts with an mRNA in the ribosome.
- Explain the sequence of events that occurs during translation as a protein elongates by one amino acid. At each step, specify what is happening in the ribosome's A site, P site, and E site.
- What evidence supports the hypothesis that peptide bond formation is catalyzed by a ribozyme?

Group Discussion Problems

1. The 5' cap and poly (A) tail that are attached to eukaryotic mRNAs appear to help the message last longer by protecting it from degradation by ribonucleases. But why is an enzyme such as a ribonuclease in the cell in the first place? What function would an enzyme that destroys messages serve?
2. Look back at Figure 16.1, which shows the formation of a phosphodiester bond during RNA polymerization. Then study the structure of the nucleotide shown below, called cordycepin. If cordycepin triphosphate, which has three phosphate groups bonded to the 5' hydroxyl group in the figure, is added to a cell-free transcription reaction, the nucleotide is added onto the growing RNA chain. This observation confirms that synthesis occurs by the addition of monomers in the form of triphosphates to the 3' end of the growing chain. Briefly explain why. Be sure to describe the expected result if synthesis occurred at the 5' end of the polymer.
3. Carl Woese has determined the nucleotide sequence of rRNA molecules from a diverse array of organisms and compared them with each other. According to his data, certain portions of the rRNAs in the large subunit are very similar in all organisms. To make sense of this result, Woese suggests that the conserved sequences have an important functional role. His logic is that these conserved sequences are so important to cell function that any changes in the sequence cause death. Woese also claims that the existence of the conserved sequences supports other data indicating that peptide bond formation is catalyzed by the large rRNA, not by ribosomal proteins. Explain the thinking behind his claim.
4. Recent work has revealed why α -amanitin inhibits transcription by RNA pol II. Structural models show that α -amanitin binds to a site inside the enzyme, but not to the active site itself. Based on the model of bacterial RNA polymerase in Figure 16.2, predict where α -amanitin binds and why it inhibits transcription.



Answers to Multiple-Choice Questions 1. c; 2. b; 3. d; 4. c;
5. c; 6. d

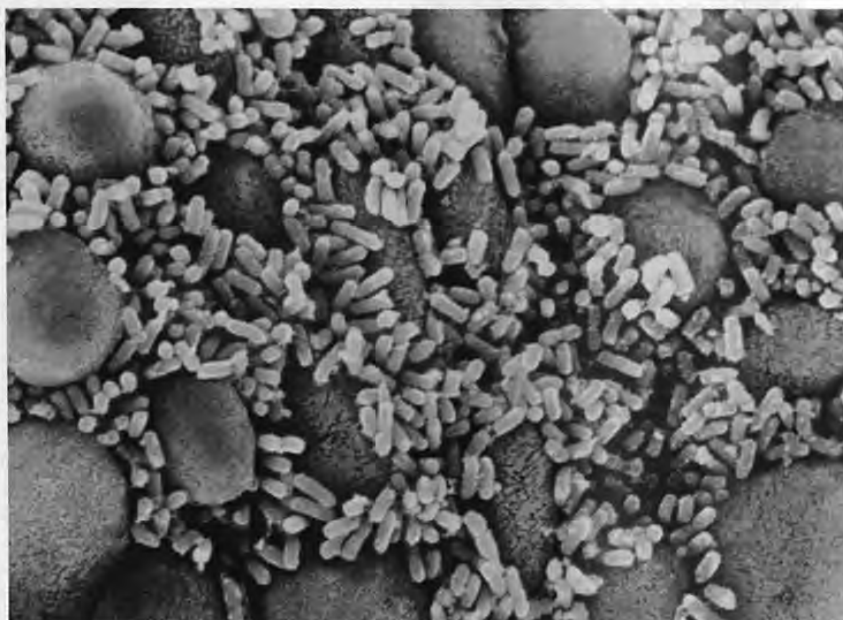
www.prenhall.com/freeman is your resource for the following:
Web Tutorials; Online Quizzes and other Online Study Guide materials; Answers to Conceptual Review Questions; Solutions to Group Discussion Problems; Answers to Figure Caption Questions and Exercises; and Additional Readings and Research.

17

Control of Gene Expression in Bacteria

KEY CONCEPTS

- Changes in gene expression allow bacterial cells to respond to environmental changes.
- Gene expression can be controlled at three levels: transcription, translation, or post-translation (protein activation).
- Transcriptional control can be positive or negative. In the case of the genes for lactose metabolism in *E. coli*, positive control occurs when a regulatory protein binds to DNA and increases the transcription rate. Negative control occurs because another regulatory protein must be released before transcription can occur.
- Many regulatory proteins bind to specific sites in DNA. Because different regulatory proteins have different amino acid sequences, they bind to different DNA sequences.



The red mounds in this micrograph are human intestinal cells; the yellow structures are bacteria. In the intestine, the nutrients available to bacteria constantly change. This chapter explores how changes in gene expression help bacteria respond to environmental changes.

Imagine waiting anxiously to hear the opening lines of a wonderfully melodic symphony, played by a renowned orchestra. The crowd applauds as the celebrated conductor comes onstage, then hushes as he takes the podium. He cocks the baton; the musicians raise their instruments. As the baton comes down, every instrument begins blaring a different tune at full volume. A tuba plays “Dixie”; a violinist renders “In-A-Gadda-Da-Vida”; and a cellist begins Mexico’s national anthem. A snare drum lays down beats for “Baby Got Back” while the bass drum simulates the cannons in the “1812 Overture.” Instead of music, there is pandemonium. The conductor staggers off stage, clutching his heart.

A cacophony like this would result if a bacterial cell “played” all its genes at full volume all the time. The *Escherichia coli* cells living in your gut right now have over 4300 genes. If all of those genes were expressed at the fastest possible

rate at all times, the *E. coli* cells would stagger off the stage, too. But this does not happen. Cells are extremely selective about which genes are expressed, in what amounts, and when.

This chapter explores how bacterial cells control the activity of their genes. **Gene expression** is said to occur when a protein or other gene product is synthesized and is active in the cell.

Understanding how cells regulate gene expression is a fundamental issue in biological science, for two reasons. First, bacteria are the most abundant organisms on Earth, and they occupy virtually every habitat known. In many cases, bacterial cells are able to grow and reproduce successfully because they respond rapidly to dramatic changes in temperature, pH, light, competitors, and nutrients. Changes in gene expression give bacteria the ability to cope with these types of environmental change. Second, questions about bacterial gene expression have enormous practical significance. For example, biologists are trying

to understand which genes are expressed when disease-causing bacteria colonize a person and start an infection in the hope that drugs can be developed that alter the expression of key genes. Bacterial cells are also used commercially to manufacture important products such as human insulin and growth hormone. Efficient use of bacteria depends on a solid understanding of how transcription and translation are regulated.

Understanding gene expression is fundamental to our understanding of life. This chapter's goal is to introduce key concepts in the regulation of gene expression. Let's begin by reviewing some of the environmental challenges that bacterial cells face and then explore how these organisms meet them.

17.1 Gene Regulation and Information Flow

Bacteria are found in virtually every habitat on Earth, from boiling hot springs to alpine snowfields and from the open ocean to crevices miles underground. The millions of bacterial species existing today have evolved a bewildering variety of ways to solve a central problem of living: obtaining the materials and energy required for growth and reproduction. Although some bacteria specialize by using just one type of food, the vast majority of species are able to switch among several different sources of carbon and energy, according to which nutrients are available in the environment at the time. The fundamental question addressed in this chapter is *how* this switching occurs. Each type of nutrient requires a different membrane transport protein to bring the nutrient molecule into the cell and a different suite of enzymes to process it. How does a bacterial cell turn some genes on and others off so that it can take advantage of alternative food sources? More generally, how do bacteria regulate gene expression so that a cell makes only the products that it needs?

As a case study, we explore the strategies of *Escherichia coli*. These cells can use a wide array of carbohydrates to supply the carbon and energy they need. The carbohydrates are in your food. As your diet changes from day to day, the availability of different sugars in your intestines varies. Precise control of gene expression gives *E. coli* the ability to respond to these changes in its environment and use the different sugars.

To appreciate why precise control over gene expression is so important to these cells, it is critical to realize that bacteria can be packed an inch thick along your intestinal walls. The organisms represent many different species that are competing for space and nutrients. For a cell to survive and reproduce in this environment, it must use resources efficiently—particularly resources that provide energy and carbon compounds. For example, recall from Chapter 4 that carbohydrates are polymers made up of simple sugars and from Chapter 9 that carbohydrates provide cells with energy and chemical building blocks. *Escherichia coli* is capable of metabolizing a wide variety of carbohydrates. But it would be energetically wasteful for *E. coli* to produce all of the enzymes required to process all of the var-

ious carbohydrates all of the time. Instead, it is logical to predict that the enzymes being produced match the sugars that are available at a given time. Efficient use of resources, via tight control over gene expression, is critical if cells are going to be able to compete successfully for space and nutrients.

On the basis of this reasoning, transcription and translation of individual genes in bacteria should be triggered by specific signals from the environment, such as the presence of specific sugars. Did you drink milk at your last meal, or eat french fries and a candy bar? Each type of food contains different sugars. Each sugar should induce a different response by the *E. coli* cells in your intestine. Just as a conductor needs to regulate the orchestra's musicians, cells need to regulate which proteins they produce at any given time.

Mechanisms of Regulation—an Overview

Gene expression can be controlled at any step between the synthesis of RNA and the activation of the final gene product. Three steps occur as information flows from DNA to proteins, represented by arrows in the following expression:

$$\text{DNA} \rightarrow \text{mRNA} \rightarrow \text{protein} \rightarrow \text{activated protein}$$

The arrow from DNA to RNA represents transcription—the making of messenger RNA (mRNA). The arrow from RNA to protein represents translation, in which ribosomes read the information in mRNA and use that information to synthesize a protein. The arrow from protein to activated protein represents post-translational modifications—including folding, addition of carbohydrate or lipid groups, or perhaps phosphorylation.

How can a bacterial cell conserve ATP and amino acids by avoiding the production of proteins that are not needed at a particular time? A look at the flow of information from DNA to protein suggests three possible mechanisms:

1. The cell could avoid making the mRNAs for particular enzymes. If there is no mRNA, then ribosomes cannot make the gene product. For example, various regulatory proteins affect the ability of RNA polymerase to bind to a promoter and initiate transcription. Genes that are controlled in this way are said to undergo **transcriptional control**:

$$\text{DNA} \not\rightarrow \text{mRNA} \rightarrow \text{protein} \rightarrow \text{activated protein}$$

2. If the mRNA for an enzyme has been transcribed, the cell might have a way to prevent the mRNA from being translated into protein. Mechanisms that alter the length of time an mRNA survives before it is degraded by ribonucleases, that affect translation initiation, or that affect elongation factors and other proteins during the translation process are forms of **translational control**:

$$\text{DNA} \rightarrow \text{mRNA} \not\rightarrow \text{protein} \rightarrow \text{activated protein}$$

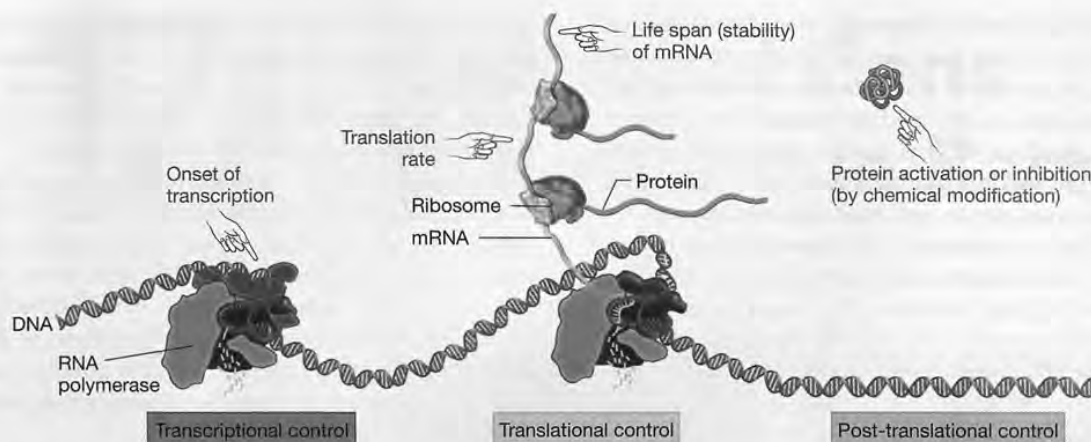


FIGURE 17.1 Gene Expression Can Be Regulated at Three Levels

Although this chapter focuses on how regulatory molecules affect the ability of RNA polymerase to initiate transcription, translational and post-translational controls also occur in bacteria.

- Chapter 16 pointed out that some proteins are manufactured in an inactive form and have to be activated by chemical modification, such as the addition of a phosphate group. This type of regulation is **post-translational control**:

DNA → mRNA → protein → $\times$ → activated protein

Which of these three forms of control occur in bacteria? The short answer to this question is “all of the above.” As **Figure 17.1** shows, many factors affect how much active protein is produced from a particular gene. Transcriptional control is particularly important due to its efficiency—it saves the most energy for the cell, because it stops the process at the earliest possible point. Translational control is advantageous because it allows a cell to change quickly which proteins are produced. Post-translational control is significant as well, and it provides the most rapid response of all three mechanisms. Among these mechanisms of gene regulation, there is a clear trade-off between the speed of response and the conservation of ATP, amino acids, and other resources. Transcriptional control is slow but efficient in resource use. Post-translational control is fast but energetically expensive.

Although this chapter focuses almost exclusively on mechanisms of transcriptional control in bacteria, it is important to keep in mind that both translational and post-translational controls occur in these organisms. It's also important to recognize that some genes—such as those that code for the enzymes required for glycolysis—are transcribed all the time, or **constitutively**. But the expression of other genes is regulated, meaning that they may be induced or repressed. Finally, it is critical to realize that gene expression is not an all-or-none proposition. Genes are not just “on” or “off”—instead, the level of expression is highly variable. What factors determine how much a particular gene is expressed at a particular time?

Metabolizing Lactose—A Model System

As Chapters 13 through 16 have shown, many of the fundamental advances in genetics have been achieved through the analysis of a model system. Studying the inheritance of seed shape in garden peas revealed the fundamental patterns of gene transmission. Exploring transcription in viruses and *E. coli* led to the discovery of RNA polymerase, transcription factors (such as the sigma proteins), and promoters. In studies of gene regulation, the key model system has been the metabolism of the sugar lactose in *E. coli*.

Jacques Monod, François Jacob, and many colleagues introduced lactose metabolism in *E. coli* as a model system during the 1950s and 1960s. Although they worked with a single species of bacterium, their results had a profound effect on thinking about gene regulation in all organisms. Some details turned out to be specific to the *E. coli* genes responsible for lactose metabolism, but many of Monod and Jacob's results are universal.

Escherichia coli can use a wide variety of sugars for ATP production, via glycolysis and the electron transport chain or via fermentation. These sugars also serve as a raw material in the synthesis of amino acids, vitamins, and other complex compounds. Glucose is *E. coli*'s preferred carbon source, however, meaning that it is the source of energy and carbon atoms that the organism uses most efficiently. This observation is logical, because glycolysis begins with glucose and is the main pathway for the production of ATP. Lactose, the sugar found in milk, is also used by *E. coli*, but only when glucose supplies are depleted. Recall from Chapter 5 that lactose is a disaccharide made up of one molecule of glucose and one molecule of galactose.

To use lactose, *E. coli* must first transport the sugar into the cell. Once lactose is inside the cell, *E. coli* can break the sugar down into glucose and galactose by means of the enzyme β -galactosidase. The glucose released by this reaction is used

directly via glycolysis; other enzymes convert the galactose to an intermediate in the glycolytic pathway.

In the early 1950s, biologists discovered that *E. coli* produces high levels of β -galactosidase only when lactose is present in the environment. If lactose is absent, little β -galactosidase is synthesized. Because lactose appears to induce the production of the enzyme that breaks itself down, researchers proposed that lactose itself regulates the gene for β -galactosidase. More formally, the hypothesis was that lactose acts as an inducer. An **inducer** is a molecule that stimulates the expression of a specific gene.

In the late 1950s Jacques Monod investigated how the presence of glucose affects the regulation of the β -galactosidase gene. Would *E. coli* produce high levels of β -galactosidase when both glucose and lactose were present in the surrounding environment? Figure 17.2 summarizes the experimental approach by

which Monod found that the answer was no. This meant that β -galactosidase production is induced by the availability of lactose, but only when glucose is unavailable. The key observations can be summarized as follows:

- Cells grown on glucose only: no β -galactosidase production
- Cells grown on glucose + lactose: no β -galactosidase production
- Cells grown on lactose only: β -galactosidase production

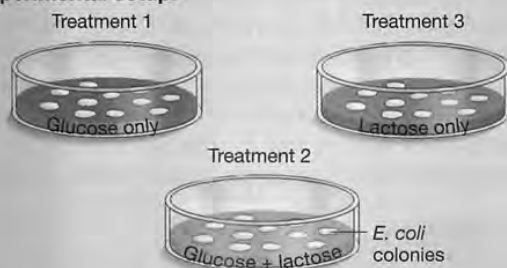
Monod teamed up with François Jacob to investigate exactly how lactose and glucose regulate the genes responsible for lactose metabolism. Research on how these genes are regulated opened a window on how genes in all organisms are controlled. Research on this system is still continuing, some 50 years later.

Question: *E. coli* produces β -galactosidase when lactose is present. Does *E. coli* produce β -galactosidase when both glucose and lactose are present?

Hypothesis: *E. coli* does not produce β -galactosidase when glucose is present, even if lactose is present. (Glucose is the preferred food source.)

Alternative hypothesis: *E. coli* produces β -galactosidase whenever lactose is present, regardless of the presence or absence of glucose.

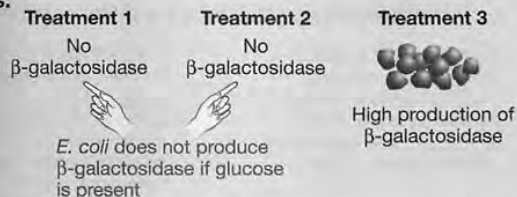
Experimental setup:



Prediction: β -Galactosidase will be produced only in treatment 3.

Prediction of alternative hypothesis: β -Galactosidase will be produced in treatments 2 and 3.

Results:



Conclusion: Glucose prevents the expression of the gene for β -galactosidase. The presence of lactose without glucose stimulates the expression of that gene.

17.2 Identifying the Genes Involved in Lactose Metabolism

To understand how *E. coli* controls production of β -galactosidase and the membrane transport protein that brings lactose into the cell, Monod and Jacob first had to find the genes that code for these proteins. To do this, they employed the same tactic used in the pioneering studies of DNA replication, transcription, and translation reviewed in earlier chapters: They isolated and analyzed mutant individuals. In this case, their goal was to find *E. coli* cells that were *not* capable of metabolizing lactose. Cells that can't use lactose must lack either β -galactosidase or the lactose-transporter protein.

Finding mutants with respect to a particular trait was a two-step process. The researchers' first step was to generate a large number of individuals with mutations at random locations in their genomes. To produce mutant cells, Monod and others exposed *E. coli* populations to X-rays, UV light, or chemicals that damage DNA and increase mutation rates. The second step is to screen the mutants to find individuals with defects in the process or biochemical pathway in question—in this case, defects in lactose metabolism. A technique that allows researchers to identify individuals with a particular type of mutation is called a **screen**.

Monod and colleagues were looking for cells that cannot grow in an environment that contains only lactose as an energy source. Normal cells grow well in this environment. How could the researchers select cells on the basis of *lack* of growth?

Screening Mutants—Replica Plating and Indicator Plates

Replica plating and growth on indicator plates were key techniques in the search for mutants with defects in lactose metabolism. **Replica plating** begins with spreading mutagenized

FIGURE 17.2 Glucose Affects the Regulation of the β -Galactosidase Gene

bacteria on a plate that is filled with gelatinous agar containing glucose, as illustrated in **Figure 17.3**. This plate is known as the **master plate**. The bacteria are then allowed to grow, so that each cell produces a single colony. (A bacterial colony consists of a large number of identical cells that are descended from a single cell.) Next, a block covered with a piece of sterilized velvet is pressed onto the master plate. Because of the contact, cells from each colony on the master plate are transferred to the velvet. Then the velvet is pressed onto a plate containing a medium that differs from that of the master plate by a single component. For instance, in Figure 17.3, the second medium has only lactose as a source of carbon and energy. Cells from the velvet stick to the plate's surface, producing an exact copy of the colonies on the master plate. This copy is called the **replica plate**. After these cells grow, an investigator can compare the colonies that thrive on the replica plate's medium with those on the master plate. In this case, colonies that grow on the master plate but that are missing on the replica plate represent mutants deficient in lactose metabolism. By picking these particular colonies from the master plate, researchers build a collection of lactose mutants.

Monod also used an alternative strategy based on **indicator plates**, where mutants with metabolic deficiencies are observed directly. Consider what happens when mutagenized bacteria are grown on agar plates containing lactose and are then sprayed with a solution containing the colorless compound ONPG (o-nitrophenyl- β -D-galactoside). ONPG consists of two subunits—a galactose molecule and a molecule called o-nitrophenol—that are joined by the same β -1,4 glycosidic linkage that joins the galactose and glucose subunits of lactose. Consequently, ONPG acts as an alternative substrate for

β -galactosidase. When the enzyme cleaves ONPG, galactose and o-nitrophenol are released. This is significant, because o-nitrophenol is intensely yellow. When colonies exposed to ONPG turn yellow, it means that they have β -galactosidase activity. Colonies that stay white are unable to cleave ONPG, meaning they have a defect in the enzyme or its production.

Different Classes of Lactose Metabolism Mutants

The initial mutant screen yielded the three types of mutants summarized in **Table 17.1**. In one class, the mutant cells were unable to cleave ONPG even if lactose was present inside the cells to induce production of the β -galactosidase protein. The investigators concluded that these mutants must lack a functioning version of the β -galactosidase protein—meaning the gene that encodes β -galactosidase is defective. This gene was designated *lacZ*, and the mutant allele *lacZ*⁻.

In the second class of mutants, the cells failed to accumulate lactose inside the cell. In contrast, the concentration of lactose in wild-type cells is about 100 times that of lactose in the surrounding environment. To interpret this result, Jacob and Monod hypothesized that the mutant cells had defective copies of the membrane protein responsible for transporting lactose into the cell. This protein was identified and named **galactoside permease**; the gene that encodes it was designated *lacY*.

The third and most surprising class of mutants did not regulate expression of β -galactosidase and galactoside permease normally. For example, when these mutant cells were grown on lactose alone and then sprayed with ONPG, they turned yellow just as normal cells do. But if they were grown on a medium that contained glucose but no lactose, they still turned

REPLICA PLATING

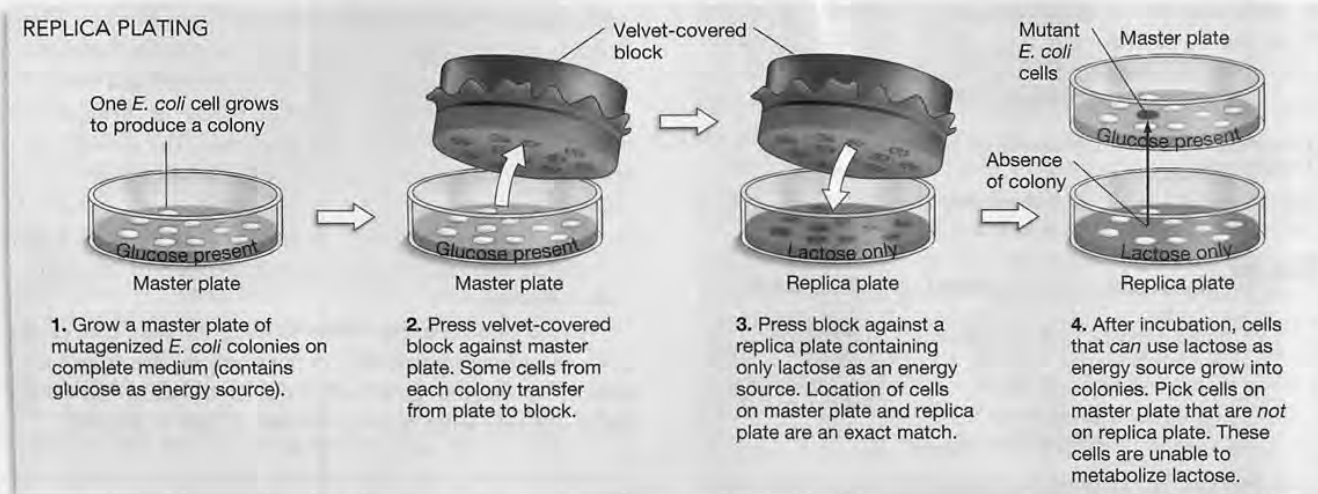


FIGURE 17.3 Replica Plating Is a Technique for Identifying Mutant Cells

Replica plating, used here to isolate mutant *E. coli* cells with a deficiency in lactose metabolism. There are two keys to this technique: (1) The location of colonies on the master plate and replica plate must match exactly, and (2) the media in the two plates must differ by just one type of component. **QUESTION** How would you alter this protocol to isolate mutant cells with a deficiency in the enzymes required to synthesize tryptophan?

TABLE 17.1 Three Distinct Types of Mutants in Lactose Metabolism of *E. coli*

Mutant Phenotype	Interpretation	Inferred Genotype
Cells cannot cleave ONPG even if lactose is present as an inducer.	No β -galactosidase; gene for β -galactosidase is defective. Call this gene <i>lacZ</i> .	<i>lacZ</i> ⁻
Cells cannot accumulate lactose inside the cell.	No membrane protein (galactoside permease) required for import of lactose; gene for galactoside permease is defective. Call this gene <i>lacY</i> .	<i>lacY</i> ⁻
ONPG is cleaved even if lactose is absent (no inducer).	Constitutive expression of <i>lacZ</i> and <i>lacY</i> ; gene for regulatory protein that shuts down <i>lacZ</i> and <i>lacY</i> is defective—it does not need to be induced by lactose. Call this gene <i>lacI</i> .	<i>lacI</i> ⁻

yellow after being sprayed with ONPG. Normal cells, in contrast, do not produce β -galactosidase under these conditions. Normal cells remain white when grown on glucose and then sprayed with ONPG.

Cells that are abnormal because they produce a product at all times are called **constitutive mutants**. Unlike normal cells, their product is always part of the cell's constitution. The gene that mutated to produce constitutive β -galactosidase expression was named *lacI*. The use of "I" was appropriate because these mutants did not need an inducer to express β -galactosidase or galactoside permease. Recall that, in normal cells, the expression of these genes is induced by the presence of lactose. But in cells with a mutant form of *lacI* (*lacI*⁻ mutants), gene expression occurred with or without lactose. This meant that *lacI*⁻ mutants have a defect in gene regulation. In these mutants, the gene remains "on" when it should be turned off.

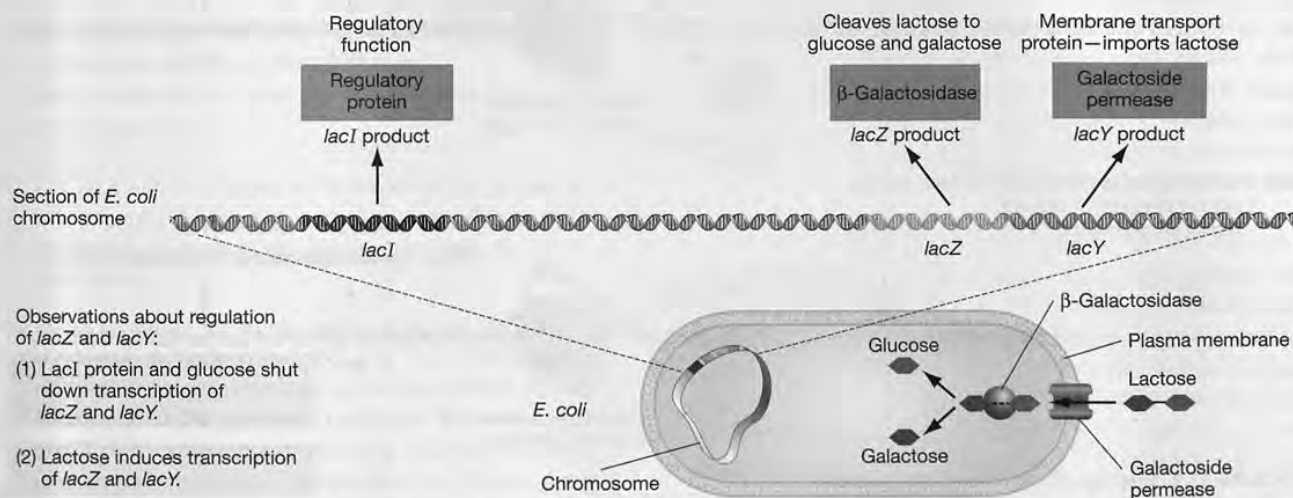
Based on these observations, it is logical to infer that the normal product of the *lacI* gene prevents the transcription of *lacZ* and *lacY* when lactose is absent. Because lactose acts as an inducer for the production of β -galactosidase, it is reasonable to expect that the *lacI* gene or gene product interacts with lactose in some way. (Later work showed that the inducer is a de-

rivative of lactose called *allolactose*. For the sake of historical accuracy and simplicity, however, this discussion refers to lactose itself as the inducer.)

Several Genes Are Involved in Metabolizing Lactose

Jacob and Monod had succeeded in identifying three genes involved in lactose metabolism: *lacZ*, *lacY*, and *lacI*. They had concluded that *lacZ* and *lacY* code for proteins involved in the metabolism and import of lactose, while *lacI* is responsible for some sort of regulatory function. The *lacI* gene or gene product repressed the expression of *lacZ* and *lacY* in the absence of lactose, but when lactose was present, the opposite occurred—transcription of *lacZ* and *lacY* was induced.

When Jacob and Monod followed up on these experiments by mapping the physical location of the three genes on *E. coli*'s circular chromosome, they found that the genes are close together (Figure 17.4). This was a crucial result, because it suggested that both *lacZ* and *lacY* might be controlled by *lacI*. Could one regulatory gene manage more than one protein-encoding gene? If so, how does *lacI* actually work? And why do lactose and glucose have opposite effects on it?

**FIGURE 17.4 The *lac* Genes Are in Close Physical Proximity**

Early experiments on lactose use in *E. coli* identified the *lacI*, *lacZ*, and *lacY* genes, documented the function of their protein products, and mapped their locations on the *E. coli* chromosome.

17.3 The Discovery of the Repressor

In principle, there are two ways that transcription can be regulated: via positive control or negative control. **Positive control** is in effect when something must be added for transcription to occur; **negative control** is in effect when something must be taken away for transcription to occur. In starting up a car, positive control occurs when you step on the gas pedal, meaning that you add fuel; negative control occurs when the parking brake is set. It turned out that the *lacZ* and *lacY* genes in *E. coli* are controlled by both a gas pedal and a parking brake—that is, they are under positive control *and* negative control.

The hypothesis that the *lacZ* and *lacY* genes might be under negative control originated with Leo Szilard in the late 1950s. Szilard suggested to Jacques Monod that the *lacI* gene codes for a product that represses transcription of the *lacZ* and *lacY* genes. Recall from Section 17.2 that, in *lacI*⁻ mutants, the production

of β -galactosidase and galactoside permease is constitutive, even if lactose is not available. Szilard proposed that regulation fails because the *LacI*⁺ protein—which normally prevents enzyme synthesis—is inactive. Stated another way, the *lacI* gene produces an inhibitor that exerts negative control over the *lacZ* and *lacY* genes. This transcription inhibitor was called a **repressor** and was thought to bind directly to DNA near or on the promoter for the *lacZ* and *lacY* genes (Figure 17.5a). To explain how the presence of lactose triggers transcription in normal cells, Szilard and Monod proposed that lactose interacts with the repressor in a way that makes the repressor release from its binding site (Figure 17.5b). The idea was that lactose induces transcription by removing negative control. Lactose releases the parking brake. As Figure 17.5c shows, it was logical to observe that constitutive transcription occurs in *lacI*⁻ mutants because a functional repressor is absent—the parking brake is broken.

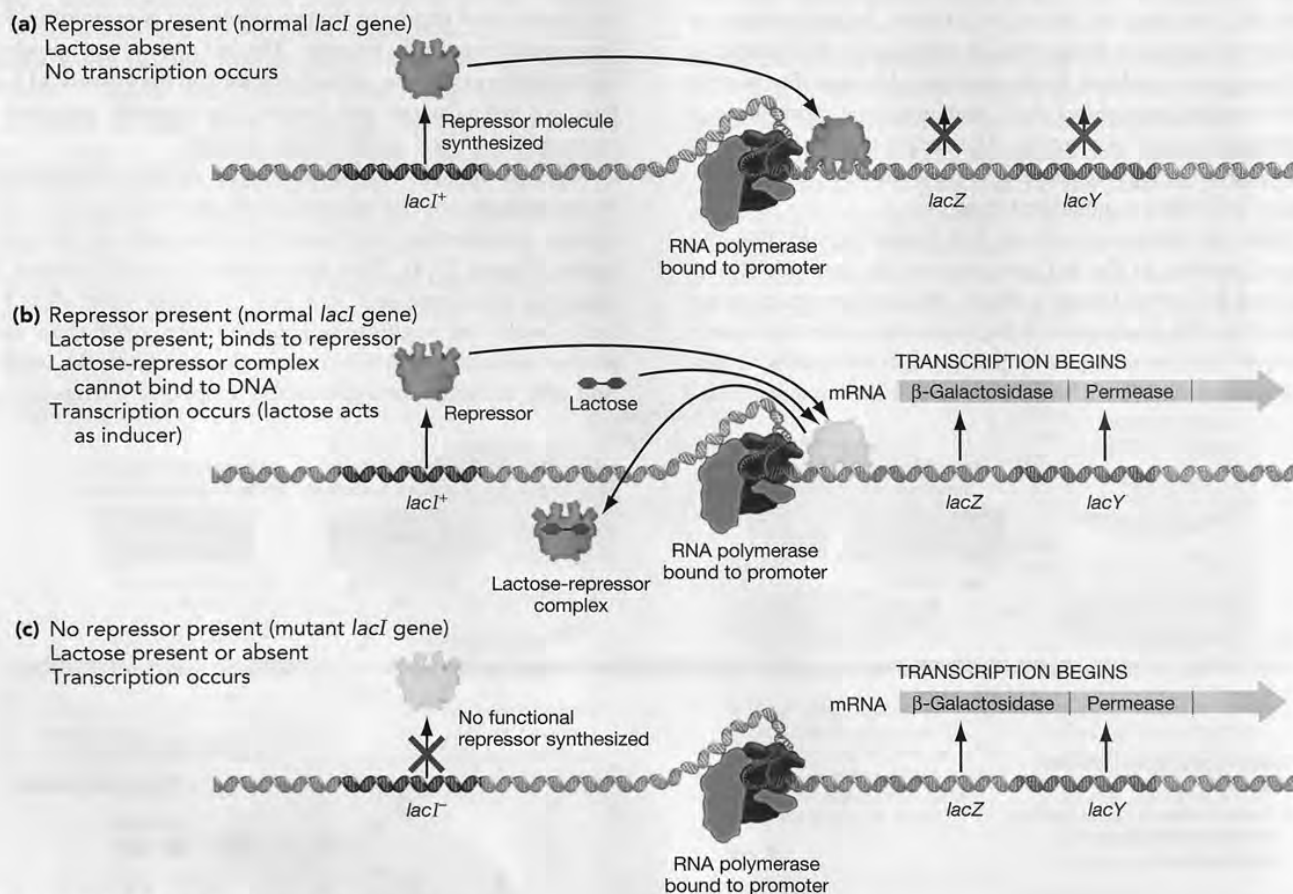


FIGURE 17.5 The Hypothesis of Negative Control

The negative-control hypothesis maintains (a) that transcription of genes involved in lactose use is normally blocked by a repressor molecule that binds to DNA on or near the promoter for *lacZ* and *lacY*, (b) that lactose induces transcription of *lacZ* and *lacY* by interacting with the repressor, and (c) that, when a functional repressor is absent, transcription proceeds.

How could researchers assess these hypotheses? This was no trivial task. But in a set of brilliant experiments, each of the predictions of the negative-control hypothesis was tested and confirmed. The key studies were based on creating *E. coli* cells that were diploid for the genes involved in lactose metabolism. For example, consider an experiment that begins with cells that are haploid and have the genotype $I^- lacZ^+ lacY^+$. These cells lack the hypothesized repressor but have normal copies of the genes for β -galactosidase or galactoside permease. As predicted by the hypothesis of negative control, they produce β -galactosidase constitutively. But what happens when these cells are made diploid by the addition of DNA with the genotype $I^+ lacZ^- lacY^-$? The second set of genes codes for a functioning copy of the repressor, but not of β -galactosidase or galactoside permease. Once the cells are made diploid, β -galactosidase production declines and then stops. This result supports the hypothesis that the repressor codes for a protein that shuts down transcription. As predicted, the repressor is the “parking brake” on transcription. But if an inducer such as lactose is added to these diploid cells, β -galactosidase activity continues. This result supports the hypothesis that lactose removes the repressor—it releases the parking brake. Stated another way, lactose acts as an inducer by removing negative control.

The *lac* Operon

Jacob and Monod summarized the results of their experiments with a comprehensive model of negative control that was published in 1961. In essence, their experimental data had confirmed the hypotheses illustrated in Figure 17.5. One of their key conclusions was that the genes for β -galactosidase and galactoside permease are controlled together. To capture this point, they coined the term **operon** for a set of coordinately regulated bacterial genes that are transcribed together into one mRNA. Logically enough, the group of genes involved in lactose metabolism was termed the ***lac* operon**.

Three hypotheses are central to the Jacob-Monod model of *lac* operon regulation:

1. The *lacZ* and *lacY* genes are adjacent and are transcribed into one mRNA initiated from the single promoter of the *lac* operon. As a result, the expression of these two genes is coordinated.
2. The repressor is a protein encoded by *lacI* that binds to DNA and prevents transcription of *lacZ* and *lacY*. The binding site for the repressor is a section of DNA in the *lac* operon called the **operator**. Jacob and Monod proposed that *lacI* is expressed constitutively and that its product (the repressor) prevents transcription by blocking RNA polymerase from making contact with the promoter.
3. The inducer (lactose) interacts directly with the repressor by binding to it. As a result, the repressor changes shape in

a way that causes it to drop off the DNA strand. Transcription is then able to proceed because RNA polymerase can access the promoter. Recall from Chapter 3 that this form of control over protein function is called **allosteric regulation**. When allosteric regulation occurs, a small molecule binds directly to a protein and causes it to change its shape and activity. When the inducer binds to the repressor, negative control ends.

Testing the *lac* Operon Model

Did Jacob and Monod “get it right”? After their model of negative control was published, Jacob and Monod confirmed the existence of the operator. They did this by finding *E. coli* mutants that expressed *lacZ* constitutively yet had normal forms of the repressor. In these new mutants, the repressor protein was unable to function because the nucleotide sequence of the operator was altered. In 1967 Walter Gilbert and Benno Müller-Hill were able to tag copies of the repressor protein with a radioactive atom. Their experiments showed that the repressor physically binds to the DNA sequences of the operator. This result confirmed that the operator is not a protein or an RNA product, but instead is part of the DNA sequence of the *lac* operon.

The only major modification of the original *lac* operon model is based on the recent recognition that the repressor does not physically block RNA polymerase from contacting the promoter, as Jacob and Monod had proposed. Instead, the repressor prevents the RNA polymerase holoenzyme from opening the DNA helix once that enzyme has bound to the promoter. As a result, the template strand cannot be threaded through the enzyme’s active site so that transcription can begin.

In addition, a third gene of the *lac* operon, *lacA*, was discovered after the original model of negative control was published. The *lacA* gene is tightly linked to *lacY* and *lacZ* and is transcribed as part of the same mRNA strand that carries *lacZ* and *lacY*. The gene codes for the enzyme transacetylase, which transfers acetyl (CH_3COO^-) groups to β -galactoside sugars. The protein’s function is protective in nature: The acetylation step allows sugars that cannot be metabolized to be exported from the cell instead of accumulating to dangerous levels within the cell.

An mRNA like the one produced by the *lac* operon is known as a polycistronic mRNA. **Polycistronic mRNAs** contain more than one protein-encoding segment. In most cases, bacterial genes that are transcribed as a polycistronic mRNA code for products devoted to a common end, such as the *lacZ* and *lacY* products required for lactose metabolism. Similarly, the polycistronic *trp* (pronounced “trip”) operon in *E. coli* consists of five genes, each of which is required for different steps in the synthesis of the amino acid tryptophan. The *trp* operon is also under negative control—but with a twist explained in Box 17.1, page 370. Biologists interpret the polycistronic structure of many bacterial genes as an adaptation

BOX 17.1 Almost Parallel Worlds: Similarities and Contrasts in Control of the *trp* and *lac* Operons

The genes in the *lac* operon code for proteins that are responsible for breaking down a compound in the cell, lactose. Logically enough, the expression of these genes depends on the presence of lactose, the compound that is to be broken down. When the inducer, a lactose derivative, binds to the repressor, the repressor comes off the operator and transcription begins.

Now imagine how a cell would regulate an operon that is responsible for synthesizing a molecule instead of breaking it down. In this case, the cell should stimulate gene expression when the molecule is absent but should reduce transcription when the molecule is present. To see how cells accomplish this task, let's explore how *E. coli* regulates the ***trp* operon**—a set of five cotranscribed genes, each of which is required for the synthesis of the amino acid tryptophan.

If RNA polymerase binds to the promoter for the *trp* operon and begins transcription, all five *trp* operon genes are transcribed and translated together. Once translation is complete, the synthesis of tryptophan commences.

The *trp* operon is under negative control. It has an operator sequence that overlaps the promoter, and a repressor protein that binds to the operator to prevent transcription—as in the *lac* operon.

In the *trp* operon, tryptophan binds to the repressor to regulate transcription, just as allolactose binds to the *lac* operon's repressor to regulate transcription. There is a crucial difference in the two systems, however. If tryptophan binds to the repressor, it means that tryptophan is present in the cell and transcription should not proceed.

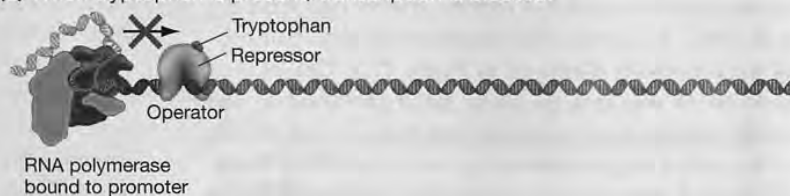
This is exactly what biologists observe: As Figure 17.6a shows, the *trp* repressor binds operator DNA and exerts negative control only when the repressor is complexed with tryptophan. When tryptophan is absent, the repressor does not bind to the operator and the operon is expressed constitutively (Figure 17.6b). This is exactly the opposite of how the presence of lactose affects the *lac* repressor. And it makes perfect sense for the *trp* operon: Bacterial cells will have higher fitness if the operon is transcribed when tryptophan levels are low but shut down when tryptophan is abundant.

In addition to offering a new twist on negative control, the *trp* operon has provided other fascinating insights into gene regulation. Experiments have revealed that negative control in the *trp* operon is supplemented by a regulatory mechanism called **attenuation**.

Attenuation works like a dimmer switch, because it superimposes fine-tuned control on what would otherwise be a simple on-off system. In the case of the *trp* operon, the dimmer is sensitive to tryptophan levels. When tryptophan levels are high, attenuation causes transcription to terminate early in the operon. When tryptophan levels are low, transcription proceeds through the entire operon. Attenuation that is based on amino acid concentration is seen in many other operons devoted to amino acid synthesis.

After considering how the *lac* and *trp* operons are regulated in *E. coli*, it's difficult not to have new respect for “simple” forms of life. Bacteria may be small and unicellular, but they perform fantastically complex, finely controlled chemistry. Even the simplest cells are marvelous machines.

(a) When tryptophan is present, transcription is blocked.



(b) When tryptophan is absent, transcription occurs.

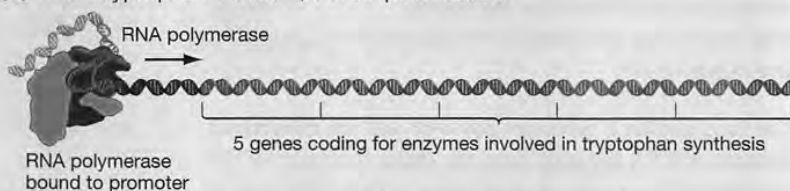


FIGURE 17.6 The *trp* Operon Is under Negative Control

(a) When tryptophan is present, it binds to a repressor protein. The tryptophan-repressor complex binds to the operator and shuts down transcription, which ends the synthesis of tryptophan. (b) The *trp* operon contains five coding genes.

that increases the efficiency of gene expression. By binding to one promoter, regulatory proteins can control the expression of several genes that are required for the same process. Polycistronic mRNAs are common in bacteria but rare or nonexistent in most eukaryotes.

The Impact of the *lac* Operon Model

Regulation of the *lac* operon provided an immensely important model system in genetics, partly because aspects of the model turned out to be important in many other species, genes, and operons. For example, follow-up work has

shown that numerous bacterial genes are under negative control via repressor proteins. Even more important, the *lac* operon model introduced the idea that gene expression is regulated by physical contact between specific regions in regulatory proteins and specific regulatory sites in DNA. Stated another way, the model introduced the idea that gene expression could be modified by proteins that bind directly to regulatory sequences in DNA. Publication of the *lac* operon model was a watershed event in the history of biological science. Its impact on thinking about gene regulation is hard to overestimate.

As elegant as the model presented in Figure 17.5 is, however, something is missing. Where does glucose fit in? This is the final element of the regulatory system for lactose metabolism. As it turns out, it is also the element that is based on positive control of transcription.

✓CHECK YOUR UNDERSTANDING

Negative control occurs when something must be taken away for transcription to occur. The *lac* operon repressor exerts negative control over three protein-coding genes by binding to the operator site in DNA near the promoter. For transcription to occur, an inducer molecule (a derivative of lactose) must bind to the repressor, causing it to release from the operator. You should be able to (1) explain why it is logical to observe that a molecule derived from lactose induces transcription of the *lac* operon; (2) diagram the *lac* operon, showing the relative positions of the operator, the promoter, and the three protein-coding genes; and (3) show what that operon looks like in the presence and absence of lactose.

17.4 Catabolite Repression and Positive Control

When glucose is present in the environment, transcription of the *lac* operon is reduced. This is true even when lactose is available to induce β -galactosidase expression. Given that glucose is the preferred carbon source, this aspect of control over the *lac* operon is logical. Glucose is produced when β -galactosidase cleaves lactose. When glucose is present, it is not necessary for the cell to cleave lactose and produce still more glucose.

Recall from Chapter 9 that biologists use the term *catabolism* to refer to reactions that break larger molecules into simpler subunits. The hydrolysis of lactose into its glucose and galactose subunits is an example of catabolism (Figure 17.7a). In many cases, operons that encode catabolic enzymes are inhibited when the end product of the reaction, the **catabolite**, is abundant. This form of end-product inhibition is **catabolite repression**. In the case of the *lac* operon, glucose is the catabolite. When glucose is abundant in the cell, transcription of the *lac* operon is decreased by catabolite repression (Figure 17.7b).

How does glucose prevent expression of the *lac* operon? An answer to this question began to emerge when researchers discovered a second major control element in the *lac* operon. This regulatory switch consists of a DNA sequence known as the **CAP binding site**, which is located just upstream of the *lac* promoter, and a regulatory protein called **catabolite activator protein (CAP)**, which binds to this DNA sequence.

To understand how CAP works, notice that not all promoters are created equal. Strong promoters allow efficient initiation of transcription by RNA polymerase; weak ones support much less efficient initiation of transcription. The *lac* promoter is weak. But when the CAP regulatory protein is bound to the

(a) Catabolism



(b) Catabolite repression of the *lac* operon

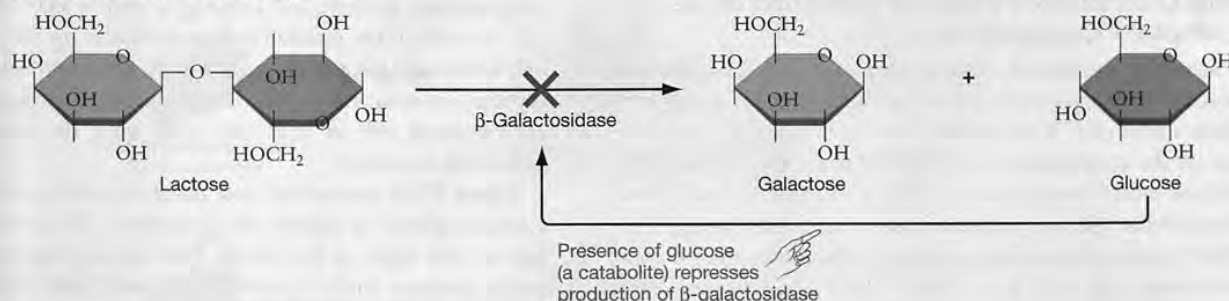


FIGURE 17.7 Catabolite Repression Is a Mechanism of Gene Regulation

(a) A generalized example of catabolism. (b) Catabolite repression occurs when one of the small product molecules represses the production of the enzyme responsible for the reaction. In the case of lactose metabolism, the production of β -galactosidase is suppressed when glucose is present.

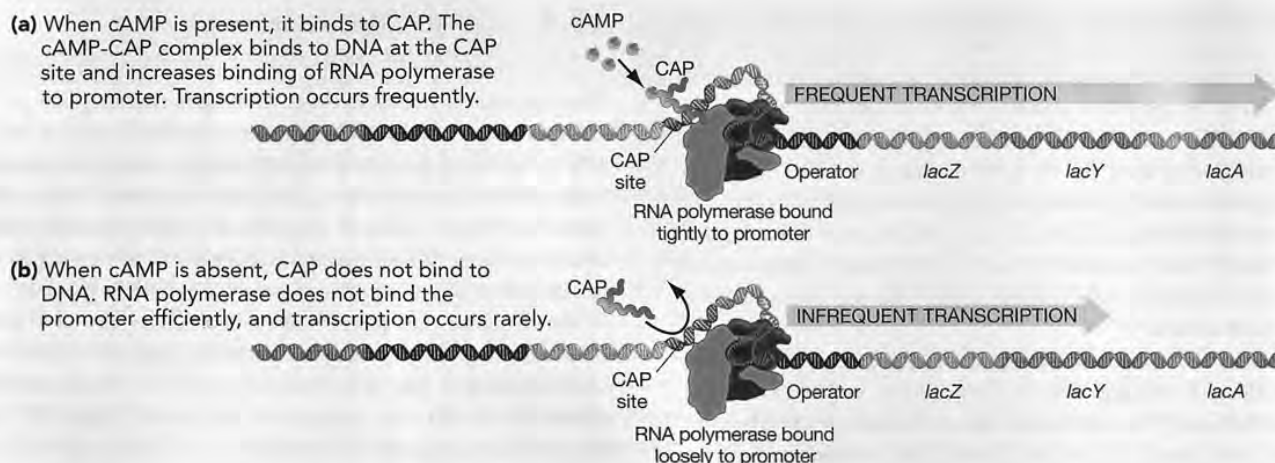


FIGURE 17.8 Positive Control of the *lac* Operon

(a) When glucose levels in an *E. coli* cell are low, cAMP is produced. cAMP then interacts with CAP to increase transcription of the *lac* operon. (b) When glucose is abundant, cAMP is rare in the cell and positive control does not occur.

CAP site just upstream of the *lac* promoter, the protein interacts with RNA polymerase in a way that allows transcription to begin much more frequently. CAP binding greatly strengthens the *lac* promoter. In this way, CAP exerts positive control of the *lac* operon. When CAP is active, transcription increases.

Researchers also discovered that CAP, like the repressor protein, is allosterically regulated. CAP changes shape when the regulatory molecule **cyclic AMP (cAMP)** binds to it. Only when cAMP is attached to it can CAP bind to DNA (Figure 17.8a). The same type of allosteric regulation goes on with the repressor and allolactose during negative control, except that the repressor binds to DNA only in the *absence* of the inducer. During positive control, the CAP-cAMP complex binds to the CAP binding site. As long as a repressor isn't bound to the operator, the complex increases the efficiency of transcription. If no cAMP is bound to CAP, then CAP has a conformation that does not allow binding to the CAP site (Figure 17.8b).

How Does Glucose Influence Formation of the CAP-cAMP Complex?

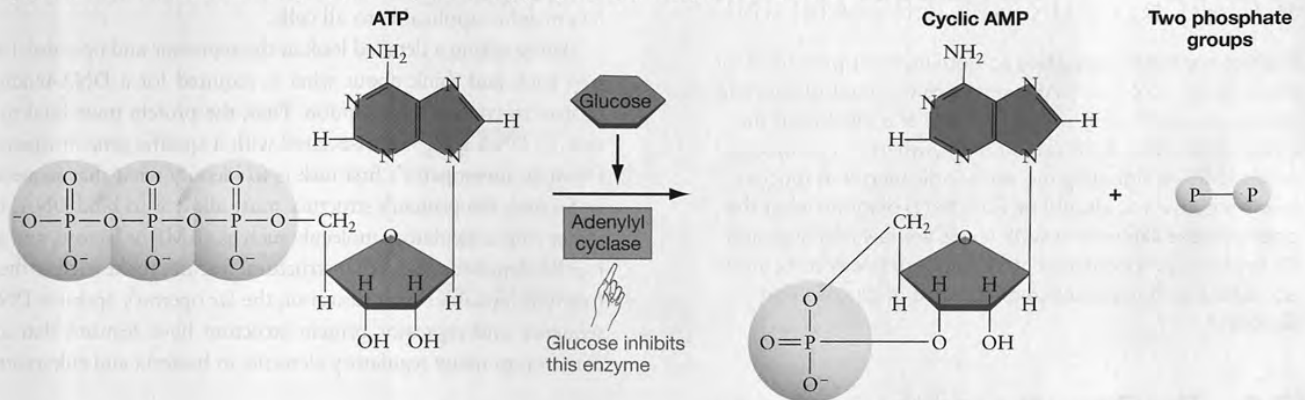
Where does the elusive molecule glucose fit into this scheme? Glucose's role in positive control of the *lac* operon is indirect. More specifically, it is mediated by the influence that glucose has on the concentration of cAMP. Glucose levels outside the cell and cAMP levels inside the cell are inversely related. When extracellular glucose concentrations are high, intracellular cAMP concentrations are low; when extracellular glucose concentrations are low, intracellular cAMP concentrations are high. This seesaw is driven by the enzyme **adenylyl cyclase**, which produces cAMP from ATP (Figure 17.9a). Adenylyl cyclase's activity is inhibited by extracellular glucose. To see the consequences of this fact, imagine a situation in which glucose is abundant outside the cell (Figure 17.9b). In this state, adeny-

lyl cyclase activity is low. Therefore, cAMP levels inside the cell are low. CAP is not in a CAP-cAMP complex, so it does not have the conformation that allows it to bind to the CAP site and stimulate *lac* operon transcription. Conversely, when the extracellular concentration of glucose is low, there is an increase in the intracellular level of cAMP. In this case, the CAP-cAMP complex forms, binds to the CAP site, and allows RNA polymerase to initiate transcription efficiently. As long as the repressor is not bound to the operator, transcription of the *lac* operon is initiated frequently and lactose can be employed as an alternative energy source.

It is important to note that the CAP-cAMP system influences many genes in addition to the *lac* operon. CAP sites are found adjacent to the promoters for several operons that are required for the metabolism of sugars other than glucose. When glucose levels fall and cAMP concentrations rise, the effect on gene regulation is similar to ringing an alarm bell. In response to the alarm, genes that encode enzymes required for the use of lactose, maltose, glycerol, and other food sources may be turned on—but only if the metabolite of a particular operon is available to the cell. Conversely, cAMP levels are low when glucose supplies are adequate. In this case, the genes for these enzymes are expressed only at low levels even when the appropriate metabolite is present.

Figure 17.10 summarizes how positive control and negative control combine to regulate the *lac* operon. The general message of this figure is that interactions among regulatory elements produce finely tuned control over gene expression. Because positive and negative control elements are superimposed, *E. coli* fully activates the genes for lactose metabolism only when lactose is available and glucose is scarce or absent. In this way, control over gene expression increases the ability of these cells to compete, grow, and reproduce efficiently.

(a) Glucose inhibits the activity of the enzyme adenylyl cyclase, which produces cAMP from ATP.



(b) The amount of cAMP and the rate of transcription of *lac* operon are inversely related to the concentration of glucose.

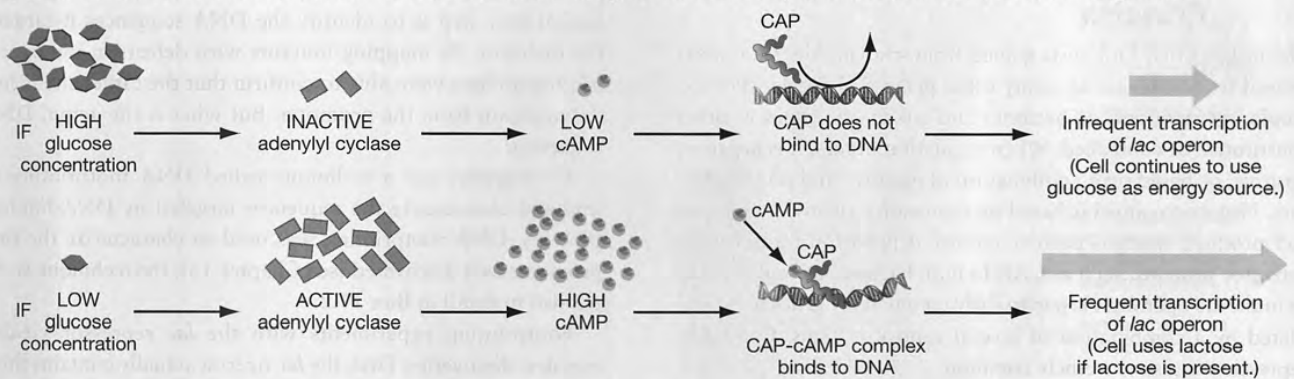


FIGURE 17.9 Cyclic AMP (cAMP) Is Synthesized When Glucose Levels Are Low

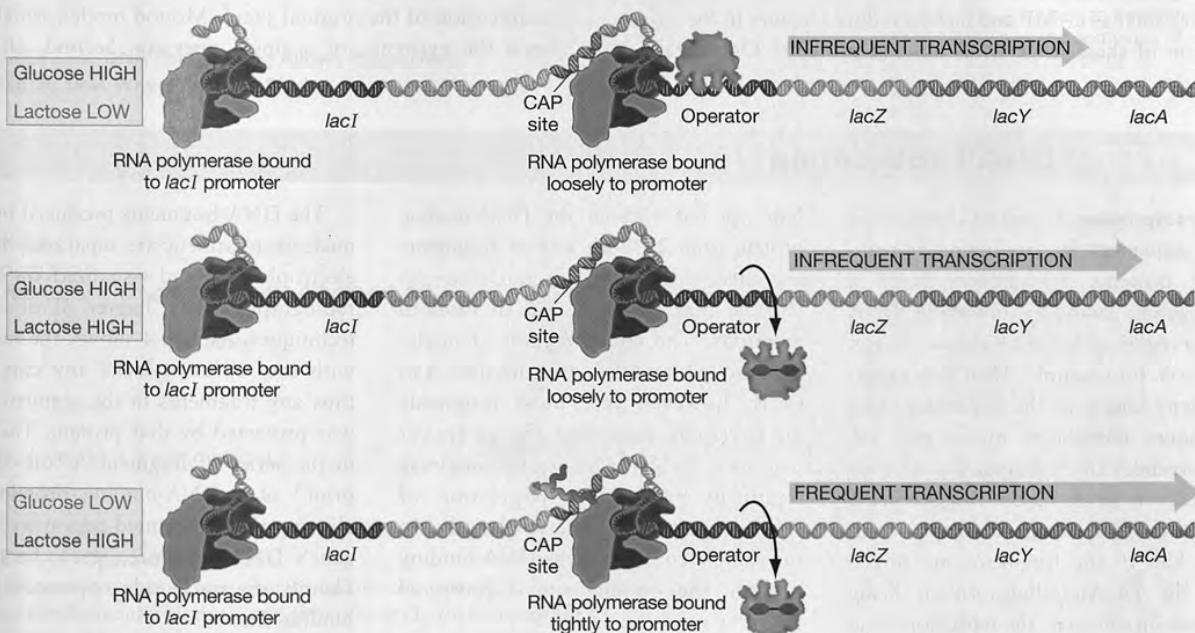


FIGURE 17.10 An Overview of Lactose Use

✓ CHECK YOUR UNDERSTANDING

Positive control occurs when something must be added for transcription to occur. CAP exerts positive control over the *lac* operon by binding to the CAP site and increasing the transcription rate. CAP can bind only when it is complexed with cAMP—a signaling molecule indicating that glucose levels are low. You should be able to (1) diagram what the operon looks like when cAMP levels are low and high and (2) explain how positive control and negative control interact, based on the presence or absence of glucose and lactose.

17.5 The Operator and the Repressor— an Introduction to DNA-Binding Proteins

The insights that biologists gained from studying the *lac* operon proved to be relevant to many other genes and species. For example, transcription of bacterial and eukaryotic genes is either constitutive or controlled. When regulation occurs, it is negative, positive, or based on a combination of negative and positive factors. Negative control is based on repressor proteins, such as the *lacI* product, whereas positive control depends on transcription activator proteins, such as CAP. In both bacteria and eukaryotes, as in the *lac* operon, it is particularly common for genes to be regulated by a combination of several control systems. Catabolite repression is also extremely common.

Currently, biologists who are interested in gene regulation are trying to understand how DNA-binding proteins such as CAP and the repressor can control transcription. Precisely how do these proteins interact with DNA? How do small regulatory molecules such as cAMP and lactose induce changes in the conformation of these large DNA-binding proteins? Once again,

investigators are using the *lac* operon as a model system that offers insights applicable to all cells.

Before taking a detailed look at the repressor and operator, let's step back and think about what is required for a DNA-binding protein to control transcription. First, the protein must bind to a specific DNA sequence associated with a specific gene or operon. Often an investigator's first task is to identify what that sequence is. Second, the protein's structure must allow it to bind DNA, interact with a regulatory molecule such as cAMP or lactose, and affect RNA polymerase. What structural features make each of these functions possible? As it turns out, the *lac* operon's operator DNA sequence and repressor protein structure have features that are common to many regulatory elements in bacteria and eukaryotes.

Finding the Operator

To understand how the *lac* repressor controls transcription, a logical first step is to identify the DNA sequences it targets: the operator. By mapping mutants with defects in the operator, researchers were able to confirm that the target site is just downstream from the promoter. But what is the actual DNA sequence?

Investigators use a technique called DNA footprinting to find and characterize the sequences targeted by DNA-binding proteins. DNA footprinting was used to characterize the first promoter ever discovered (see Chapter 16); the technique is explained in detail in Box 17.2.

Footprinting experiments with the *lac* repressor yielded two new discoveries. First, the *lac* operon actually contains three sites where the repressor protein may bind. As Figure 17.12a shows, these sites are called O_1 , O_2 , and O_3 . When the repressor protein is bound to DNA, the protein simultaneously binds O_1 and one of the two other operators. This finding required a modification of the original Jacob-Monod model, which proposed the existence of a single operator. Second, all three

BOX 17.2 DNA Footprinting

DNA footprinting is used to characterize DNA sequences that are bound by regulatory proteins. Researchers begin a footprinting study by obtaining DNA from a region of interest—the *lac* operon of *E. coli*, for example. Then they generate many copies of the sequence, using techniques introduced in Chapter 19. The identical DNA fragments are then radioactively labeled at one end (see step 1 of Figure 17.11). Once labeling is complete, half of the fragments are mixed with the DNA-binding protein being studied—in this case, the repressor—and

half are left without the DNA-binding protein (step 2). Both sets of fragments are then treated with a nuclease—an enzyme that cleaves DNA in random locations. The concentration of nuclease is so low and the exposure time is so short, however, that most fragments are never cut; those that are cut are cut just once. As step 3 shows, the nuclease treatment produces a population of DNA fragments of different lengths. In the sample containing the DNA-binding protein, the binding site is protected from the nuclease.

The DNA fragments produced by the nuclease treatment are separated by gel electrophoresis and visualized via autoradiography (see Chapter 3) or other techniques. As step 4 shows, the sample with the repressor lacks any cuts and thus any fragments in the segment that was protected by that protein. The gap in the series of fragments is the “footprint” of a DNA-binding protein. By aligning the footprinted region with the gene's DNA sequence, researchers can identify the nucleotide sequence of the binding site.

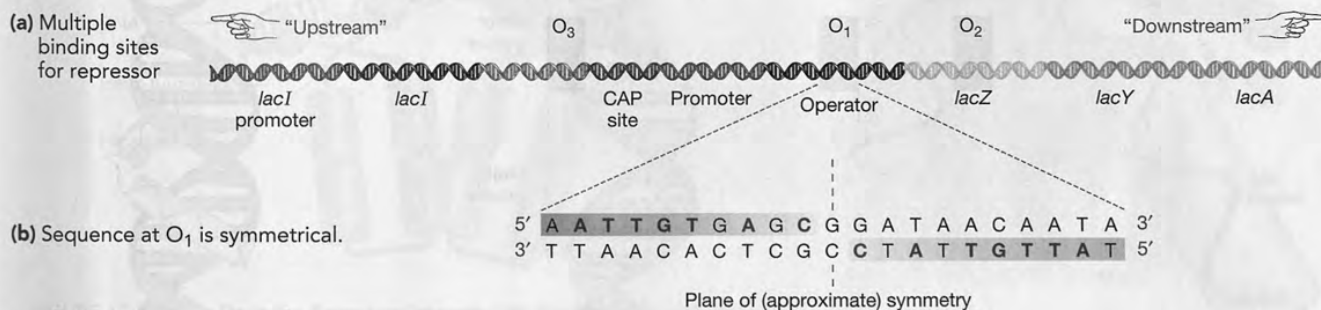
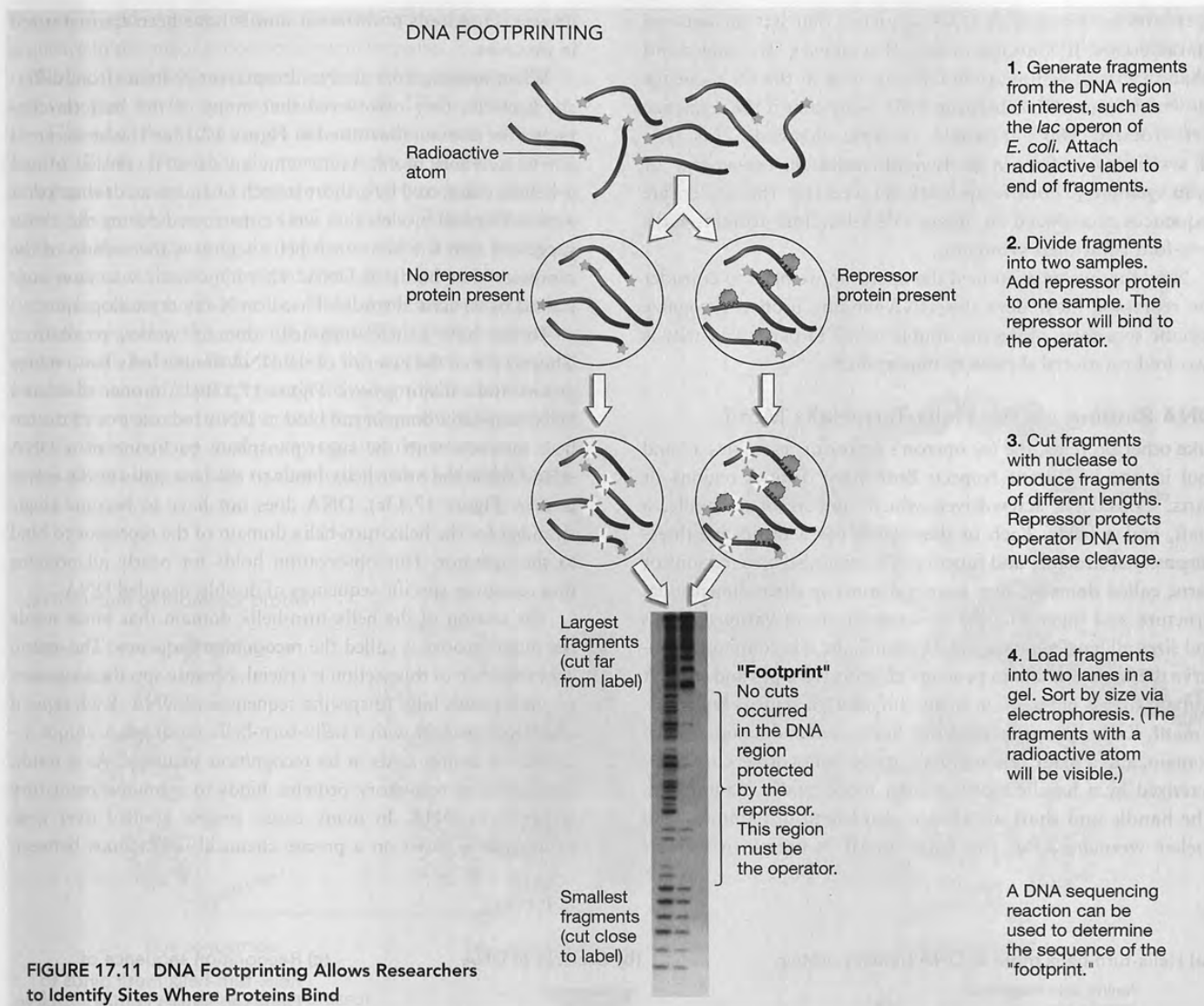


FIGURE 17.12 Operator Sequences

(a) The *lac* operon contains three operator sequences, called O₁, O₂, and O₃, where the repressor can bind. A repressor simultaneously binds O₁ and either O₂ or O₃. (b) Many sequences targeted by DNA-binding proteins are symmetrical. In this example, all but three bases are identical on the two sides of the plane of symmetry.

operators have a similar DNA sequence that has an unusual characteristic: It contains an axis of symmetry. To understand what an axis of symmetry in DNA is, look at the O_1 sequence shown in **Figure 17.12b** (page 375). Note that if the sequence were rotated 180° , it would be little changed. This type of symmetry is known as **two-fold rotational symmetry** or **dyad symmetry**. Follow-up work showed that the nucleotide sequences recognized by many DNA-binding proteins have two-fold rotational symmetry.

Now that we've examined the operator, we need to consider the repressor. How does this DNA-binding protein recognize specific sequences along the double helix? In particular, why is two-fold rotational symmetry important?

DNA Binding via the Helix-Turn-Helix Motif

Like other proteins, the *lac* operon's repressor resembles a hand tool in one important respect: Both have distinct regions or parts. Consider a screwdriver, which consists of a handle, a shaft, and a blade. Each of these parts has a distinctive three-dimensional structure and function. Proteins also have regions or parts, called **domains**, that have a distinctive three-dimensional structure and function. Just as screwdrivers of various shapes and sizes all have a handle, shaft, and blade, it is common to observe the same domain in proteins of various shapes and sizes. A domain that is observed in many different proteins is known as a **motif**. Each discrete screwdriver has a handle domain, a shaft domain, and a blade domain. As a group, screwdrivers are characterized by a handle motif, a shaft motif, and a blade motif. The handle and shaft motifs are also found in hammers and socket wrenches, but the blade motif is unique to screw-

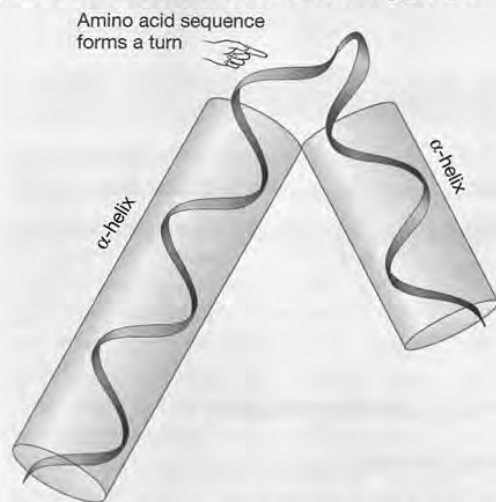
drivers. Hundreds of different motifs have been characterized in proteins.

When investigators analyzed repressor proteins from different bacteria, they discovered that many of the bacteria contained the domain illustrated in **Figure 17.13a**. This structure is a **helix-turn-helix motif**. As its name indicates, it consists of two α -helices connected by a short stretch of amino acids that forms a turn. Physical models that were constructed during the 1980s suggested that the helix-turn-helix region is the section of the repressor that binds to DNA. This hypothesis was later supported by structural studies based on X-ray crystallography.

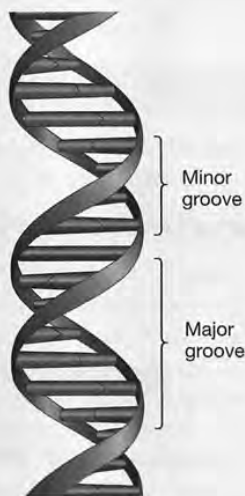
To see how a helix-turn-helix domain works, recall from Chapter 4 that the exterior of the DNA double helix has a minor groove and a major groove (**Figure 17.13b**). A protein that has a helix-turn-helix domain can bind to DNA because one of the helices interacts with the sugar-phosphate backbone of a DNA strand while the other helix binds to the base pairs in the major groove (**Figure 17.13c**). DNA does not have to become single stranded for the helix-turn-helix domain of the repressor to bind to the operator. This observation holds for nearly all proteins that recognize specific sequences of double-stranded DNA.

The section of the helix-turn-helix domain that binds inside the major groove is called the **recognition sequence**. The amino acid sequence of this section is crucial, because specific sequences of amino acids bind to specific sequences of DNA. Each type of regulatory protein with a helix-turn-helix motif has a unique sequence of amino acids in its recognition sequence. As a result, each of these regulatory proteins binds to a unique regulatory sequence in DNA. In many cases, precise control over gene expression is based on a precise chemical interactions between

(a) Helix-turn-helix motif of DNA-binding protein



(b) Grooves of DNA



(c) Recognition sequence of helix-turn-helix motif binds to DNA sequences in major groove.



FIGURE 17.13 Helix-Turn-Helix Motifs in DNA-Binding Proteins

QUESTION The part of a helix-turn-helix motif that interacts directly with DNA is called the recognition sequence. Why?

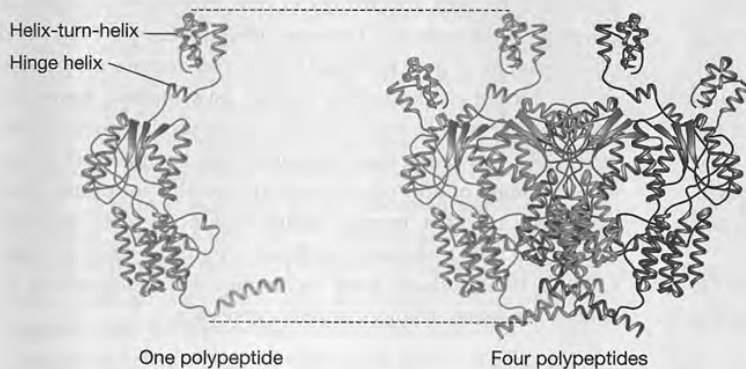
the amino acid sequence in a recognition sequence and the DNA sequence in the major groove of a regulatory site.

Biologists used X-ray crystallography to document the precise way that the *lac* repressor binds to the DNA of the operator. Consider the structure of the repressor alone, shown in Figure 17.14a. Note that the protein consists of four identical polypeptide chains, and that four helix-turn-helix domains project from the main body of the molecule. When researchers analyzed the structure of the repressor while it was bound to the operator sequence, they confirmed that each of these helix-turn-helix motifs binds operator DNA inside the major groove.

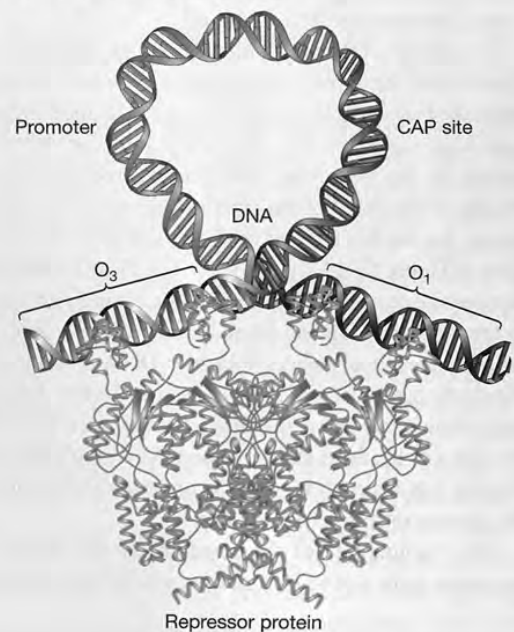
Now recall that the operator essentially is composed of two symmetrical halves. Each half of the operator sequence is bound by one of the repressor's helix-turn-helix motifs. The major grooves in the operator and the recognition helices of the repressor can pair up in this way because the operator is symmetrical and because the repressor consists of identical poly-

peptides that are flipped 180° with respect to each other. In addition, because the repressor contains four helix-turn-helix domains and two of the four bind per operator, a total of two operators can be bound simultaneously by one repressor. When this occurs, a loop of DNA is created between O_1 and whichever other operator (O_2 or O_3) is bound (Figure 17.14b). This type of DNA looping is observed at many other bacterial promoters and at virtually all eukaryotic promoters.

(a) Structure of repressor protein



(b) DNA binding of repressor



(c) Interaction of inducer and repressor

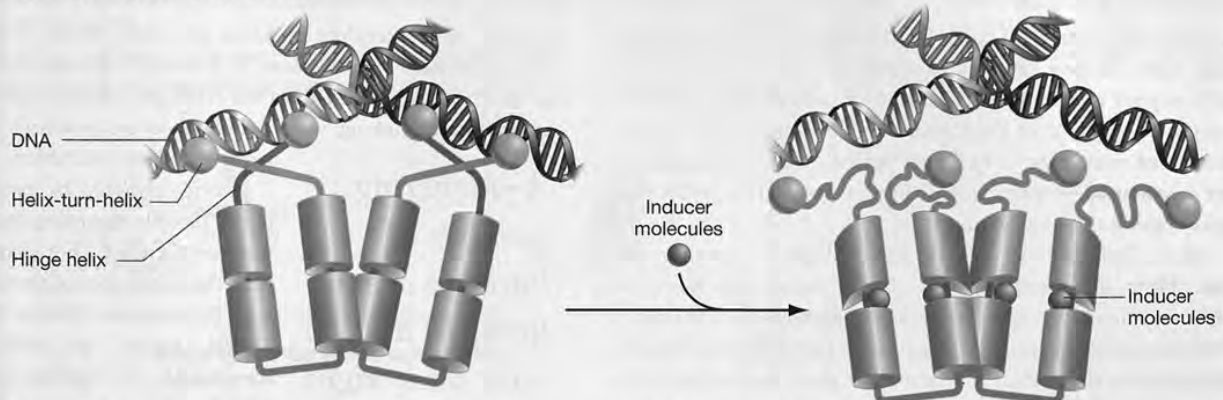


FIGURE 17.14 How Does the Repressor Interact with the Operator and the Inducer?

(a) The repressor is a complex of four polypeptides, each a product of the *lacI* locus. (b) The repressor protein has four helix-turn-helix motifs to bind DNA at O_1 and either O_2 or O_3 . (c) "Hinge helices" in the repressor exist only when the inducer is not present (left). When the inducer binds to the repressor, the hinge helices relax and cause the protein to fall off the operator DNA (right).

How Does the Inducer Change the Repressor's Affinity for DNA?

Negative control of the *lac* operon occurs when the repressor binds to the operator. In this state, the helix-turn-helix motifs of the repressor are attached to DNA. Even though RNA polymerase can bind to the promoter nearby, the enzyme cannot open the double helix and begin transcription. This physical interference is typical of how negative regulators work. Positive regulators, in contrast, work by stabilizing the interaction between RNA polymerase and the promoter.

Negative control continues as long as the repressor is bound to DNA. But when lactose is present, the repressor dissociates from the operator. How does the inducer affect the repressor in a way that disrupts DNA binding?

To answer this question, researchers analyzed the three-dimensional structure of the repressor when it was bound to a molecule called IPTG, which acts as an inducer in the *lac* system. Their data showed that IPTG binds at locations in the middle of the repressor and that it produces an important change in the shape of the protein near the helix-turn-helix domains. To see this change in shape, compare the left and right sides of **Figure 17.14c** (page 377). The diagram on the left summarizes the repressor's shape without the inducer. Pay attention to the structures labeled **hinge helices** near the helix-turn-helix motifs. The hinge helices lock the DNA-binding motifs into place when they contact the operator. As the diagram on the right shows, however, the hinge helices uncoil when IPTG or the inducer binds to the repressor. As a result, the two DNA-binding regions can't bind to the operator at the same time, and the protein releases from the site.

The "unhinging" of the hinge helices that occurs in the *lac* repressor turns out to be a unique mechanism for changing the

activity of a regulatory protein. Still, several general lessons emerge from the study of this repressor. First, the shape and activity of a protein may change when a small molecule binds to it. Allosteric shifts like that observed in the *lac* repressor occur in thousands of different proteins. Second, key elements in transcriptional control are regulated at the **post-translational level**. For example, the repressor protein is always present because it is transcribed and translated constitutively at low levels. When a rapid change in *lac* operon activity is required, it does not occur via changes in the transcription or translation of the repressor. Rather, the activity of the already translated protein is altered. In virtually all cases, the activity of key regulatory proteins is controlled by post-translational modifications.

Future Directions

Now that biologists understand the physical basis of negative control and induction, what does the future hold? One promising line of research is focused on how the activity of many genes and operons is coordinated. For example, suppose that an *E. coli* cell in your gut is faced with starvation. Does it adjust the activity of hundreds or even thousands of its genes in order to survive? If so, how are changes in gene expression coordinated throughout the genome? Because biologists have a fairly solid understanding of how individual genes and operons are regulated, the current challenge is to explore how multiple genes and operons are controlled in a way that increases the fitness of bacterial cells. Answering these questions may be important to our understanding of how certain bacteria are able to sustain infections that threaten human health and how some bacterial species are able to detoxify pollutants by using them as food. Chapter 19 introduces some techniques that biologists use to study coordinated changes in gene expression.

ESSAY Control of Gene Expression in Bacterial Pathogens

A staggering number of bacteria are living in and on your body right now. To give just one example, the number of bacterial cells in your mouth at this moment far exceeds the number of people who have ever lived, including the world's current population of more than 6.4 billion. Virtually all of these bacteria are **commensal**—meaning they live in association with their host without causing harm.

As you know, however, a few bacterial species can cause disease. How are certain bacteria able to invade and harm us? How are disease-causing bacteria, or **pathogens**, transmitted from person to person? And how can bacterial infections be controlled or prevented? In many cases, the answers depend on an understanding of how the expression of key genes is controlled. To drive this point home, consider *Vibrio cholerae*, the causative agent of cholera (**Figure 17.15**).

Cholera is a disease characterized by watery diarrhea that, if untreated, can lead to dehydration and death. The disease is

currently most prevalent in Africa and Asia, where 5 million cases occur each year. Because *V. cholerae* is transmitted from person to person when diarrhea from an infected individual contaminates drinking water, cholera occurs primarily in regions with poor sanitation.

A staggering number of bacteria are living in and on your body right now.

Vibrio cholerae is extraordinarily flexible regarding the habitats it occupies. For example, near the confluence of the Ganges and Brahmaputra Rivers in the Bengal region of India and Bangladesh, *V. cholerae* can be found as free-swimming cells, inside two different eukaryotes, in dense layers coating rocks and other underwater surfaces, and inside humans. In the Bengal, most incidences of human infection occur during an outbreak that coincides with the end of



FIGURE 17.15 *Vibrio cholerae*, the Bacterium That Causes Cholera
[©Dr. Dennis Kunkel/Visuals Unlimited].

the monsoon rains in October and November each year. This observation suggests that environmental signals trigger *V. cholerae*'s yearly emergence as a human pathogen.

What genes allow *V. cholerae* to act as a human pathogen, and how are these genes regulated? Once it is ingested by a person, *V. cholerae* confronts an environment that is markedly different from its other habitats. In fact, most of the ingested bacteria are killed in the harsh, acidic environment of the stomach. The few that survive enter the small intestine, where they encounter a thick layer of mucus that protects the intestinal cells. Then, in response to a still-unknown signal in the environment, major changes in *V. cholerae* gene expression occur. The new proteins that are produced allow *V. cholerae* cells to move through the mucus layer and make contact with the intestinal cells. Each invading cell then produces a short, hairlike projection called a pilus, which allows the *V. cholerae* cell to attach to an intestinal cell. The proteins required to create this structure result from the coordinately regulated transcription of pilus-specific genes.

Once the bacterial cells have attached to intestinal cells, *V. cholerae* begins to manufacture and secrete a toxin composed

of two different polypeptide chains. When intestinal cells take up the toxin, a devastating chain of events occurs. The toxin activates a host protein that stimulates human adenylyl cyclase. Massive amounts of cAMP are produced in response, resulting in a flow of Cl^- ions from the affected intestinal cells into the center (lumen) of the small intestine. Water follows by osmosis. The result is the watery diarrhea symptomatic of cholera.

Although the diarrhea can be fatal to the host, it increases the fitness of *V. cholerae* cells in environments where human sanitation is bad. Because *V. cholerae* cells are shed in diarrhea, they can enter drinking-water supplies if systems for transporting and processing human sewage are in poor repair or nonexistent. In this way, *V. cholerae* cells that induce diarrhea are efficiently transmitted to new hosts. *Vibrio cholerae* cells that have alleles responsible for diarrhea have higher fitness than do *V. cholerae* cells that lack these alleles.

The genes that encode the pilus proteins and cholera toxin are under the control of regulatory genes called the *Tox* (for toxin) loci. As **Figure 17.16** shows, *ToxR* produces a transcription factor that responds to a signal in the intestinal environment. When the *ToxR* protein is activated, it stimulates transcription of the *ToxT* gene. The *ToxT* product, ToxT, is a DNA-binding protein that activates transcription of several genes required for infection, including the genes that encode the cholera toxin subunits. These genes are called *ctxA* and *ctxB*. As you might suspect, they are part of a polycistronic operon.

Biologists are currently trying to translate research on gene regulation into better tools for fighting cholera. For example, blocking the activity of the *ToxR* or *ToxT* proteins should prevent disease. This hypothesis has been supported by the observation that *ToxR*[−] mutants are harmless. Researchers are now attempting to develop drugs that prevent *ToxR* from binding the signal molecule, that stop *ToxR* from activating *ToxT*, or that prevent *ToxT* from contacting DNA.

The axiom “knowledge is power” applies here: Although drugs like those proposed earlier are not yet available, research on gene regulation has provided several novel and promising drug targets.

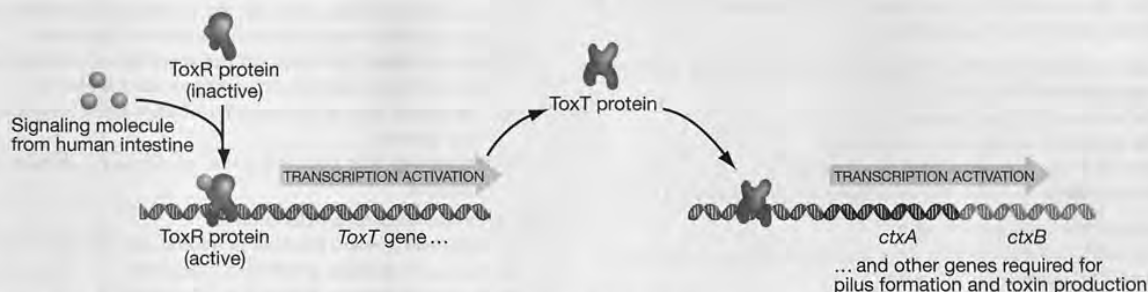


FIGURE 17.16 In the Cholera Bacterium, the Genes Required for Infection Are Regulated

QUESTION Biologists have not been able to identify the signaling molecule that triggers the cascade of events shown here. What properties must this molecule have?

CHAPTER REVIEW

Summary of Key Concepts

- Changes in gene expression allow bacterial cells to respond to environmental changes.

Transcription in bacteria is constitutive or inducible. Constitutive expression occurs in genes whose products are required at all times, such as genes that encode glycolytic enzymes. Expression of most genes is induced by environmental signals, however, so gene products are produced on demand.

Precise regulation over gene expression is important to bacterial cells. For example, some species switch between different food sources to take advantage of the nutrients currently available in their environment. Efficient use of nutrients is beneficial because cells are usually competing for access to resources. In the intestinal bacterium *E. coli*, for example, changes in gene expression allow cells to use lactose as a food source when glucose is not available.

- Gene expression can be controlled at three levels: transcription, translation, or post-translation (protein activation).

Among the three levels of gene regulation, there is a trade-off between the speed of response to changed conditions and the efficient use of resources. In post-translational control, for example, the change in gene expression is extremely fast but demands a considerable investment of resources to make and regulate the protein in question.

- Transcriptional control can be positive or negative. In the case of the genes for lactose metabolism in *E. coli*, positive control occurs when a regulatory protein binds to DNA and increases

the transcription rate. Negative control occurs because another regulatory protein must be released before transcription can occur.

Protein-coding genes involved in lactose metabolism are normally transcribed at low levels because a repressor protein binds to a DNA sequence—the operator—near their promoter. But when present, lactose binds to the repressor and induces a conformational change that causes the repressor to fall off the operator. Rapid transcription of the *lac* loci does not begin, however, unless glucose is also lacking. When glucose is scarce, cAMP levels rise in the cell. This is important because the regulatory protein CAP undergoes a conformational change when it is bound to cAMP. The CAP-cAMP complex binds to a control sequence near the *lac* promoter and increases transcription by facilitating binding by RNA polymerase.

Web Tutorial 17.1 The *lac* Operon

- Many regulatory proteins bind to specific sites in DNA. Because different regulatory proteins have different amino acid sequences, they bind to different DNA sequences.

The regulatory proteins involved in negative control and positive control are often allosterically regulated, meaning that they have the ability to switch between two different conformations in response to binding a small molecule—often a nutrient or product affected by the gene being regulated. The change in conformation affects the protein's ability to bind to DNA.

Questions

Content Review

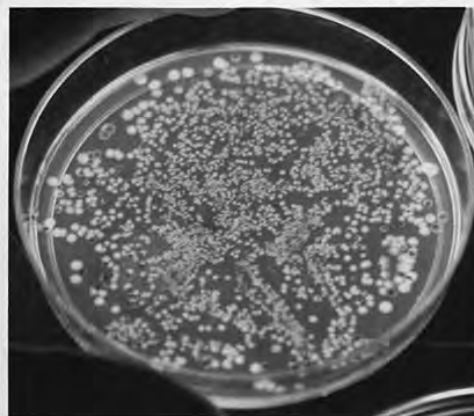
1. Genes for enzymes in glycolysis are expressed constitutively. Why?
 - a. Their expression is controlled at three levels: transcriptional, translational, and post-translational.
 - b. Glycolysis occurs at all times, so they need to synthesize ATP-producing enzymes constantly.
 - c. Transcription is activated only in response to signals indicating that ATP supplies are low.
 - d. They do not need to be expressed when fermentation instead of cellular respiration is occurring.
2. Why do researchers frequently begin a search for mutants by exposing organisms to UV light or X-rays?
 - a. The treatment causes mutants to turn yellow.
 - b. The treatment triggers gene expression.
 - c. The treatment exposes constitutive mutants.
 - d. The treatment increases the mutation rate by damaging DNA.
3. What is the goal of a genetic screen?
 - a. to identify individuals that have mutations in genes that are required for a particular process
 - b. to produce a large number of mutant individuals, which can later be analyzed in detail
 - c. to determine whether the mutations carried by particular individuals are in protein-coding genes or regulatory elements
 - d. to determine the fitness or viability of mutant individuals
4. Why are the genes involved in lactose metabolism considered to be an operon?
 - a. They occupy adjacent locations on the *E. coli* chromosome.
 - b. They have a similar function.
 - c. They are all required for normal cell function.
 - d. They are under the control of the same promoter.
5. What is catabolite repression?
 - a. a mechanism that turns off the synthesis of enzymes responsible for catabolic reactions when the product is present
 - b. a mechanism that turns off the synthesis of enzymes responsible for catabolic reactions when the product is absent
 - c. repression that occurs because of allosteric changes in a regulatory protein
 - d. repression that occurs because of allosteric changes in a DNA sequence
6. What is a helix-turn-helix motif?
 - a. a protein domain involved in folding into the active conformation
 - b. a protein domain involved in induction
 - c. a protein domain involved in DNA binding
 - d. a protein domain involved in catalysis

Conceptual Review

1. Explain the difference between positive and negative control over transcription. Why is it advantageous for the *lac* operon in *E. coli* to be under both positive control and negative control? What would happen if only negative control occurred? What would happen if only positive control occurred?
2. In *E. coli*, rising levels of cAMP can be considered a starvation signal. Explain.
3. Explain the role of post-translational control in the *lac* operon.
4. CAP is also known as the cAMP-receptor protein. Why?
5. Explain how the *lac* repressor binds to DNA. Then explain the relationship between the two-fold rotational symmetry of the operator and the four-part structure of the repressor protein.
6. The galactose released when β -galactosidase cleaves lactose enters the glycolytic pathway in *E. coli* once a series of enzyme-catalyzed reactions has converted the galactose to glucose-6-phosphate. Why, then, is glucose, not lactose, the preferred sugar in *E. coli*?

Group Discussion Problems

1. You are interested in using bacteria to metabolize wastes at an old chemical plant and convert them into harmless compounds. You find bacteria that are able to tolerate high levels of the toxic compounds toluene and benzene, and you suspect that it is due to the ability of the bacteria to break these compounds into less-toxic products. If that is true, these toluene- and benzene-resistant strains will be valuable for cleaning up toxic sites. How could you find out whether these bacteria have enzymes that allow them to metabolize toluene?
2. Assuming that the bacteria you examined in Group Discussion Problem 1 do have an enzymatic pathway to break down toluene, would you predict that the genes involved are constitutively expressed, under positive control, or under negative control? Why? What experiments could you conduct to test your hypothesis?
3. The *lacI* mutants are constitutive because the repressor fails to recognize and bind to the operator region. Other repressor mutants have been isolated that are called *LacI^S* mutants. These repressor proteins continue to bind to the operator, even in the presence of the inducer. How would this mutation affect the function of the *lac* operon? Specifically, how well would *LacI^S* mutants do in an environment that has lactose as its sole sugar?
4. X-gal is a colorless, lactose-like molecule that can be split into two fragments by β -galactosidase. One of these product molecules is blue. The following photograph is a close-up of *E. coli* colonies growing in a medium that contains lactose.



Draw a line to three colonies whose cells have functioning copies of β -galactosidase. Draw a line to three colonies whose cells have mutations in the *lacZ* locus or in one of the genes involved in regulation of the *lacZ*. Suppose you could analyze the sequence of the β -galactosidase gene from each of the mutant colonies. How would these data help you distinguish which cells are structural mutants and which are regulatory mutants?

Answers to Multiple-Choice Questions 1. b; 2. d; 3. a; 4. d;
5. a; 6. c

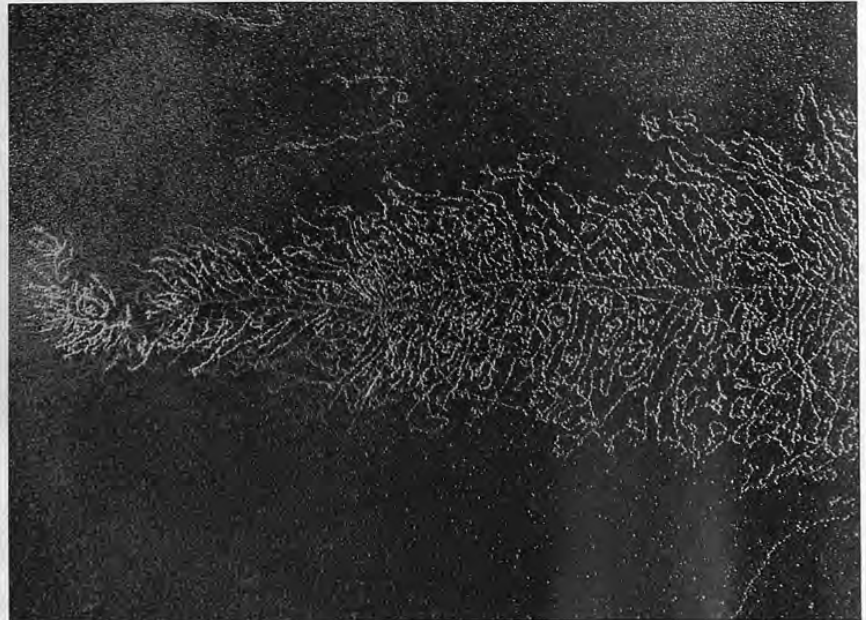
www.prenhall.com/freeman is your resource for the following:
Web Tutorials; Online Quizzes and other Online Study Guide materials; Answers to Conceptual Review Questions; Solutions to Group Discussion Problems; Answers to Figure Caption Questions and Exercises; and Additional Readings and Research.

18

Control of Gene Expression in Eukaryotes

KEY CONCEPTS

- Changes in gene expression allow eukaryotic cells to respond to changes in the environment and cause distinct cell types to develop.
- In eukaryotes, DNA is packaged with proteins into complex structures that must be opened before transcription can occur.
- In eukaryotes, transcription can be initiated only when specific proteins bind to the promoter and to regulatory sequences that may be close to the promoter or far from it.
- Alternative splicing allows a single gene to code for several different products.
- Cancer can develop when mutations disable genes that regulate cell-cycle control genes.



A eukaryotic cell's DNA being transcribed. The horizontal strand in the middle of the micrograph is DNA; the yellow strands above and below it are mRNAs.

In bacteria, precise regulation of genes is essential for cells to respond to changes in their environment, as Chapter 17 indicated. In *Escherichia coli*, the genes that are required to import and cleave lactose are rapidly expressed only when the cells have to rely on lactose as a source of energy—meaning glucose is absent and lactose is present. *Vibrio cholerae* can live in a wide array of environments, but the genes that allow the cells to infect a human are expressed only when the cells find themselves inside someone's small intestine.

Unicellular eukaryotes face similar challenges. Consider the yeast *Saccharomyces cerevisiae*, which is used extensively in the production of beer, wine, and bread. In nature, the cells live on the skins of grapes and other fruits. In the wild, variation in temperature and humidity can be high. In addition, the sugars that are available to the cells change dramatically in type and concentration as the fruit that they are living on ripens, falls, and rots. For yeast cells to grow and reproduce efficiently, gene expression has to be modified in response to these environmen-

tal changes. Further, the offspring of a cell that lived on a grape may be blown to a nearby orchard and take up residence on an apple. The genes that are expressed in the cells of grapes versus those expressed in the cells of apples vary markedly.

The cells that make up multicellular eukaryotes face an even greater challenge. Consider that, at one time, your body consisted of a single cell—a fertilized egg. Now you have trillions of cells, each of which has a specialized structure and function. You have heart muscle cells, lung cells, nerve cells, skin cells, and so on. Even though these cells look and behave very differently, each of them contains the same genes. Your bone cells are different from your blood cells not because they contain different genes, but because they express different genes. Your bone cells have blood cell genes—they just don't transcribe them.

How is it possible for the cells inside your body to express unique sets of genes? The answer is that your cells respond to changes in their environment, just as bacteria and unicellular eukaryotes do. But instead of responding to the presence of mole-

cules such as glucose or lactose in the external environment, your cells respond to the presence of signals from other cells—from an internal environment. As a human being or an oak tree develops, cells that are located in different parts of the organism are exposed to different signals. As a result, they express different genes. Differential gene expression is responsible for creating different cell types, arranging them into tissues, and coordinating their activity to form the multicellular society we call an individual.

How does all this happen? Unit 4 explores the nature of the signals that trigger the formation of muscle, bone, leaf, and flower cells in multicellular organisms. Here we need to focus on what happens once a eukaryotic cell receives a signal. Let's begin by surveying the various ways that gene expression can be controlled and then plunge into a detailed look at key points in gene regulation. The chapter closes with a look at how breakdowns in the control of gene expression can help trigger cancer.

18.1 Mechanisms of Gene Regulation— an Overview

Like bacteria, eukaryotes can control gene expression at the levels of transcription, translation, and post-translation (protein activation or inactivation). But as **Figure 18.1** shows, two additional levels of control occur in eukaryotes as genetic information flows from DNA to proteins. The first level involves the DNA-protein complex at the top of the figure. In eukaryotes, DNA is wrapped around proteins to create a protein-DNA complex called **chromatin**. Eukaryotic genes have promoters, just as bacterial genes do. Before transcription can begin in eukaryotes, DNA near the promoter must be released from tight interactions with proteins so that RNA polymerase can make contact with the promoter. To capture this point, biologists say that **chromatin remodeling** must occur prior to transcription.

The second level of regulation that is unique to eukaryotes involves **RNA processing**—the steps required to produce a mature, processed mRNA from a **primary RNA transcript**, the preliminary result of transcription. Recall from Chapter 16 that introns have to be spliced out of primary transcripts. In some cases, carefully orchestrated alterations in the splicing pattern occur. When the splicing events in a particular primary RNA transcript change, a different message emerges. The altered message, in turn, leads to the production of a different product.

How many of the five potential control points shown in **Figure 18.1** are actually used? In all cells, every one of them is employed at certain times. The primary focus of this chapter is to explore three of these key transition points: (1) control of packing DNA into chromatin (control over chromatin remodeling), (2) control of transcription, and (3) control of RNA processing.

To appreciate the breadth and complexity of gene regulation in eukaryotes, let's follow the series of events that occurs as an embryonic cell responds to a developmental signal. Suppose a molecule arrives that specifies the production of a muscle-specific protein. What happens next?

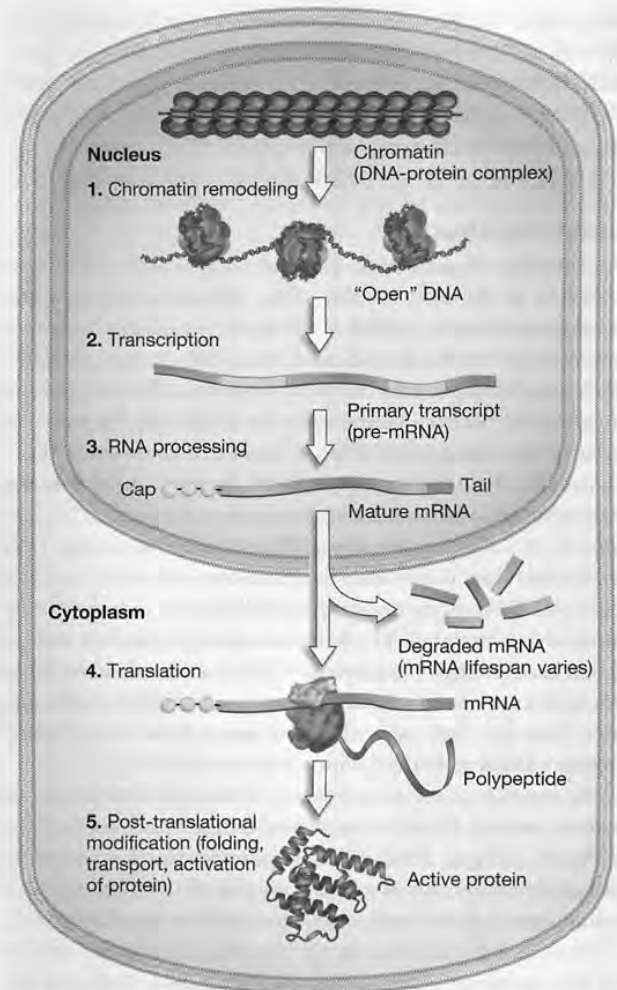


FIGURE 18.1 In Eukaryotes, Gene Expression Can Be Controlled at Many Different Steps

Numerous steps occur during protein synthesis in eukaryotes. Each step provides an opportunity for regulation of gene expression.

EXERCISE Place and label a horizontal line such that mechanisms of transcriptional control fall above the line and mechanisms of post-transcriptional control fall below it.

18.2 Eukaryotic DNA and the Regulation of Gene Expression

If the arrival of a signaling molecule from outside the cell is going to result in the transcription of a particular gene, the DNA around the target gene must be drastically remodeled. To appreciate why, consider that a typical cell in your body contains about six billion base pairs of DNA. Lined up end to end, these nucleotide pairs would form a double helix about 2 m (6.5 feet) long. But the nucleus that holds this DNA is thinner than a sheet of paper. In eukaryotes, DNA is packed inside the nucleus so tightly that RNA polymerase can't access it. Part of this packing is done by supercoiling—meaning the DNA double helix is twisted on itself

many times, just as in bacterial chromosomes (see Chapter 7). The supercoiled DNA found in bacteria does not need to be altered extensively before it can be transcribed, however. Eukaryotic DNA is different—it has to undergo a series of significant physical changes before transcription can take place. The reason is that supercoiling is just part of the packing system found in eukaryotes.

Chromatin Structure

The first data on the physical nature of eukaryotic DNA were published in the early 1900s, when chemical analyses established that eukaryotic DNA is intimately associated with proteins. Later work documented that the most abundant DNA-associated proteins belong to a group called the **histones**. In the 1970s electron micrographs like the one in **Figure 18.2a** revealed that the protein-DNA complex, or chromatin, has a regular structure. In some preparations for electron microscopy, chromatin actually looked like beads on a string. The “beads” came to be called **nucleosomes**. More details emerged in 1984 when researchers determined the three-dimensional structure of eukaryotic DNA by using X-ray crystallography (a technique introduced in Chapter 4). The X-ray crystallographic data indicated that each nucleosome consists of DNA wrapped almost twice around a core of eight histone proteins. As **Figure 18.2b** indicates, between each pair of nucleosomes there is a “linker” stretch of DNA associated with a histone called **H1**.

The intimate association between DNA and histones occurs in part because DNA is negatively charged and histones are positively charged. DNA has a negative charge because of its phosphate groups; histones are positively charged because they contain many lysine and/or arginine residues (see Chapter 3).

More recent work has shown that there is another layer of complexity in eukaryotic DNA. As **Figure 18.2c** indicates, H1 histones interact with each other and with histones in other nucleosomes to produce a tightly packed structure. Based on its width, this structure is called the **30-nanometer fiber**. (Recall that a nanometer is one-billionth of a meter and is abbreviated nm.) Often, 30-nm fibers are packed into still larger structures, which organize DNA inside the nucleus.

The elaborate structure of chromatin does more than just package DNA into the nucleus. Chromatin structure has profound implications for the control of gene expression. To drive this point home, consider the 30-nanometer fiber illustrated in **Figure 18.2c**. If this tightly packed stretch of DNA contains a promoter, how can RNA polymerase bind to it and initiate transcription?

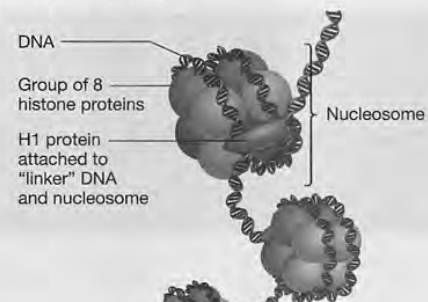
Evidence that Chromatin Structure Is Altered in Active Genes

Once the nucleosome-based structure of chromatin was established, biologists hypothesized that the close physical interaction between DNA and histones must be altered for RNA polymerase to make contact with DNA. More specifically, biologists hypothesized that a gene could not be transcribed until the chromatin near its promoter was remodeled. The idea was that the chromatin must be relaxed or decondensed for RNA

(a) Nucleosomes in chromatin



(b) Nucleosome structure



(c) 30-Nanometer fiber



FIGURE 18.2 Chromatin Has Several Levels of Structure

(a) Electron micrograph of chromatin, showing that bead-like structures, called nucleosomes, are linked by short stretches of DNA. (b) X-ray crystallographic data revealed that nucleosomes consist of DNA wrapped around a core of eight histone proteins. The linker DNA is associated with a particular histone called H1. (c) A model of the structure of the 30-nanometer fiber. It proposes that H1 proteins form a core in the center of the fiber. Each ball here represents a group of eight histones, to make the overall structure of the 30-nanometer fiber clearer.

polymerase to bind to the promoter. If so, then chromatin remodeling would represent the first step in the control of eukaryotic gene expression. Two types of studies have provided strong support for this hypothesis.

The first type of evidence comes from studies with the enzyme DNase, which cuts DNA at random locations. DNase cannot cut DNA efficiently if the molecule is tightly complexed with his-

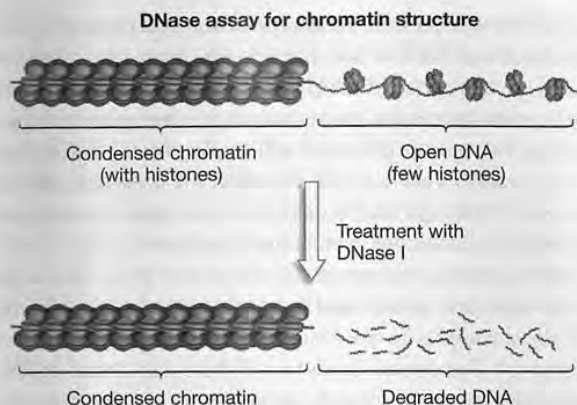


FIGURE 18.3 DNase Cannot Cut Intact Condensed Chromatin
DNase is an enzyme that cuts DNA at random locations.

tones. As **Figure 18.3** shows, the enzyme works effectively only if DNA is in the “open” configuration. Harold Weintraub and Mark Groudine used this observation to test the hypothesis that the DNA of actively transcribed genes is in an open configuration. They performed the test by comparing the structure of chromatin in two genes of the same cell type: the β -globin and ovalbumin genes of chicken blood cells. β -Globin is a protein that is part of the hemoglobin found in red blood cells; ovalbumin is a major protein of egg white. In blood cells, the β -globin gene is transcribed at high levels. The ovalbumin gene, in contrast, is not transcribed at all. After treating blood cells with DNase and then analyzing the state of the β -globin and ovalbumin genes, the researchers found that DNase cut up the β -globin gene much more readily than the ovalbumin gene. They interpreted this finding as evidence that chromatin in blood cells was in an open configuration at the β -globin gene but closed at the ovalbumin gene. Analogous studies using DNase on different genes and in different cell types yielded similar results.

The second type of evidence in support of the chromatin-remodeling hypothesis comes from studies of mutant brewer's yeast cells that do not produce the usual complement of histones. Researchers found that many yeast genes that are normally never transcribed are instead transcribed at high levels at all times in these mutant cells. To interpret this finding, biologists hypothesized that the lack of histone proteins prevented the assembly of normal chromatin. If the absence of normal histone-DNA interactions promotes transcription, then the presence of normal histone-DNA interactions must prevent it.

Taken together, the data suggest that the normal or default state of eukaryotic genes is to be turned off. If so, then gene expression depends on chromatin being opened up in the promoter region.

How Is Chromatin Altered?

Although much remains to be learned about how chromatin is opened up prior to transcription, researchers have made impor-

tant progress recently. One key finding is that two major types of protein are involved in modifying chromatin structure. One group of proteins creates structures called **chromatin-remodeling complexes**, which reshape chromatin through a series of reactions that are dependent on ATP. It is not known exactly how these ATP-dependent enzymes work, however. The second type of chromatin-modifying protein works by adding small molecules such as acetyl (CH_3COOH) or methyl (CH_3) groups to histones. These processes are known as **acetylation** and **methylation**, respectively.

Figure 18.4 shows how the enzymes called **histone acetyltransferases (HATs)** affect chromatin structure. HATs acetylate the positively charged lysine residues in histones. When a HAT adds an acetyl group to selected histones, the number of positive charges on the histones is reduced. The result is less electrostatic attraction between histones and the negatively charged DNA. The association between nucleosomes and DNA is loosened and the chromatin decondenses.

Figure 18.4 also shows how chromatin is “recondensed” by a group of enzymes called **histone deacetylases (HDACs)**, which remove the acetyl groups added by HATs. Histone deacetylase activity reverses the effects of acetylation, returning chromatin to its default condensed state. If HATs are an on switch for transcription, HDACs are the off switch.

What happens once a section of DNA is opened up and exposed to RNA polymerase? In bacteria, the protein sigma binds to RNA polymerase and allows it to contact the promoter. Then transcription begins if any required positive regulators, such as CAP, are in place and negative regulators, such as the repressor, are absent. Does anything similar happen in eukaryotes?

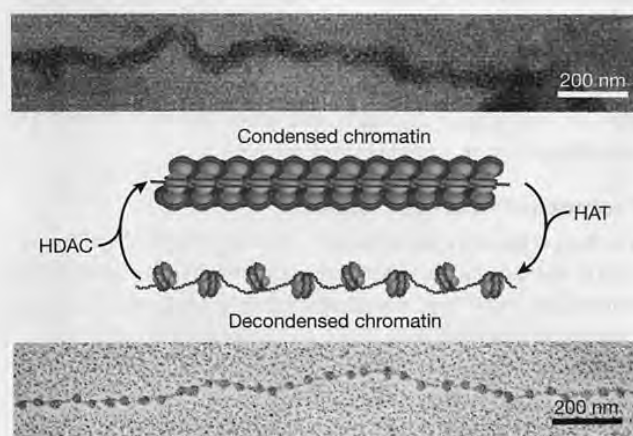


FIGURE 18.4 Chromatin Decondensation and Condensation Are Driven by Histone Acetyl Transferases and Histone Deacetylases
Histone acetyl transferases (HATs) cause chromatin to decondense, and histone deacetylases (HDACs) cause it to condense. **EXERCISE** In decondensed chromatin, each nucleosome has an acetyl group attached to it. Indicate this by adding “Ac” to several nucleosomes in the bottom part of the drawing.

✓ CHECK YOUR UNDERSTANDING

Eukaryotic DNA is wrapped tightly around histones, forming nucleosomes, which are then coiled into dense structures called 30-nm fibers. Before transcription can begin, the chromatin (DNA-protein complex) must be relaxed by chromatin-remodeling complexes and HATs so that RNA polymerase can contact the promoter. You should be able to (1) draw and label the structure of a nucleosome and a 30-nm fiber and explain their function, and (2) compare and contrast the structure of DNA in bacteria versus that in eukaryotes.

18.3 Regulatory Sequences and Regulatory Proteins

Chapter 16 introduced the **promoter**—the site in DNA where RNA polymerase binds to initiate transcription. Eukaryotic promoters are located about 30 nucleotides upstream of the point where RNA polymerase begins transcription. Many contain a specific base sequence, known as a **TATA box**, where a sigma-like protein binds and allows RNA polymerase to contact DNA. Based on their position and function, eukaryotic promoters are much like bacterial promoters. Recall from Chapter 16 that bacterial promoters may vary in sequence and bind different types of sigma proteins. Eukaryotic genes that are transcribed by RNA polymerase II also have promoters that vary in sequence, but all eukaryotic promoters are bound by the same protein: the **TATA-binding protein (TBP)**.

If eukaryotic genes have promoters that interact with the same promoter-binding protein, how can transcription be controlled? The answer lies in interactions between regulatory proteins and regulatory sequences. Eukaryotic regulatory proteins are analogous to *E. coli*'s CAP and the repressor protein analyzed in Chapter 17. **Regulatory sequences** are sections of DNA that are involved in controlling the activity of genes, similar to the CAP site and operators introduced in Chapter 17.

As it turns out, some regulatory sequences found in eukaryotes are very similar to those observed in bacteria. Others are radically different. Let's take a closer look.

Promoter-Proximal Elements

The first regulatory sequences in eukaryotic DNA were discovered in the late 1970s, when Yasuji Oshima and co-workers developed an important model system to study control of gene expression in eukaryotes. Much as François Jacob and Jacques Monod had focused on lactose metabolism in *E. coli*, Oshima's team set out to understand how yeast cells control the metabolism of the sugar galactose. When galactose is absent, yeast cells produce tiny quantities of the enzymes required to metabolize it. But when galactose is present, transcription of the genes encoding these enzymes increases by a factor of 1000.

Oshima and co-workers focused their early work on mutants that were unable to use galactose. The team's first major result was the discovery of mutant cells that failed to produce any of the five enzymes required for galactose metabolism, even

if galactose was present. To interpret this observation, they hypothesized that the cells had a knock-out (loss-of-function) mutation that completely disabled a regulatory protein. Like CAP, this hypothesized protein was thought to exert positive control over the five genes. The team called the gene *GAL4* and the protein product GAL4. (In *S. cerevisiae*, the names of genes are written in full caps and italicized. As in other species, gene names are italicized but protein names are not.)

This hypothesis was supported when other researchers isolated the regulatory protein and found that it has a DNA-binding domain, analogous to the helix-turn-helix motif introduced in Chapter 17. The domain binds to a 20-base-pair stretch of DNA located just upstream from the promoter for the five genes that GAL4 regulates. The location and structure of this regulatory sequence are comparable to those of the CAP binding site in the *lac* operon of *E. coli*. Similar regulatory sequences have now been found in a wide array of eukaryotic genes and species. Because such sequences are located close to the promoter and bind regulatory proteins, they are termed **promoter-proximal elements**.

As Figure 18.5 shows, promoter-proximal elements are just upstream from the promoter and the gene's start site; the exons and introns that make up the actual gene are farther downstream. Unlike the promoter itself, promoter-proximal elements have sequences that are unique to specific genes. In this way, they furnish a mechanism for eukaryotic cells to exert precise control over transcription.

The discovery of positive control and promoter-proximal elements provided a satisfying parallel between gene regulation in bacteria and in eukaryotes. This picture changed, however, when researchers discovered a new class of eukaryotic DNA regulatory sequences—sequences unlike anything in bacteria.

Enhancers

Susumu Tonegawa and colleagues made a discovery that may rank as the most startling in the history of research on gene expression. Tonegawa's group was exploring how human immune system cells regulate the genes involved in the production of antibodies. **Antibodies** are proteins that bind to specific sites on other molecules. In your immune system, antibodies bind to viruses and bacteria and mark them for destruction by other

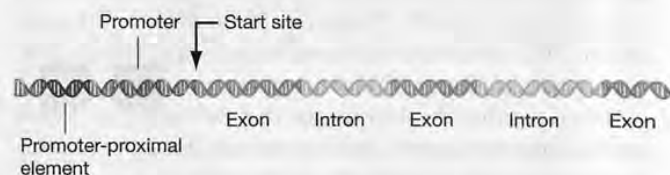


FIGURE 18.5 Promoter-Proximal Elements Regulate the Expression of Some Eukaryotic Genes

All eukaryotic genes have promoters, the sites where RNA polymerase initially is in contact with the gene. Exons and introns are not drawn to scale throughout this chapter. They are typically very large compared with promoters and promoter-proximal (regulatory) elements.

cells. The antibody gene that the researchers were working with is broken into many introns and exons. Introns are DNA sequences spliced out of the primary mRNA transcript; exons are regions of eukaryotic genes that are included in the mature RNA once splicing is complete. The biologists used techniques that will be introduced in Chapter 19 to place copies of an intron in new locations and had found that, when they placed the intron close to a gene, the gene's transcription rate increased. Based on this observation, Tonegawa and co-workers hypothesized that the intron contained some sort of regulatory sequence.

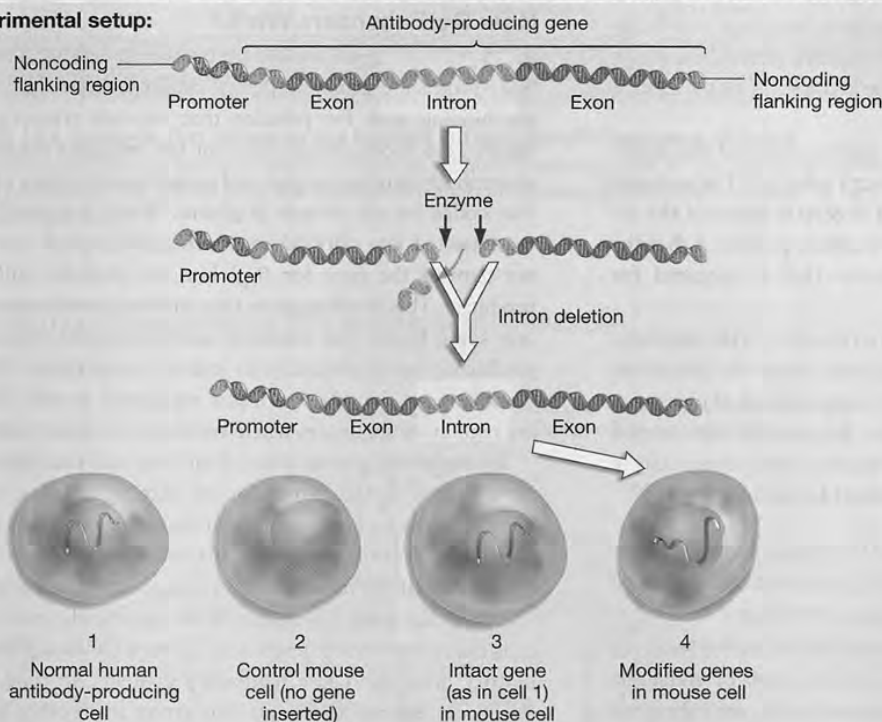
To test this hypothesis, the biologists performed what is now considered a classic experiment (Figure 18.6). The protocol was simple in concept: Starting with a human antibody-producing gene that included an intron flanked by two exons, as shown in step 1 of Figure 18.6, the team removed different portions of the intron and observed whether transcription of the gene still took place. The researchers used enzymes to cut out specific pieces of the intron in multiple copies of the gene (step 2) and to ligate the resulting sections back together (step 3). Each of the modified genes was missing a different section of the intron. Then the team

Question: Can a regulatory sequence in DNA be located far from the promoter?

Hypothesis: A regulatory sequence exists far from the promoter of an antibody-producing gene—in an intron.

Alternative hypothesis: Regulatory sequences are located close to the promoter—not in introns or in other distant sequences.

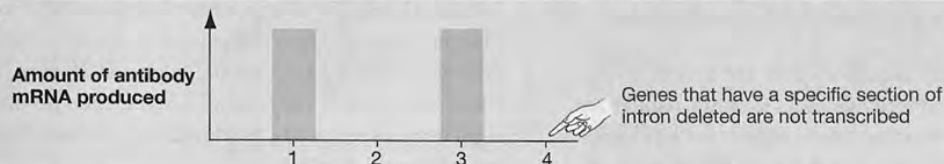
Experimental setup:



Prediction: Transcription will be reduced when part of the intron is deleted.

Prediction of alternative hypothesis: Transcription will not be affected by deleting part of the intron.

Results:



Conclusion: The deleted part of the intron must contain a regulatory sequence that is required for transcription. Thus, regulatory sequences in DNA may be located far from the promoter—in this case, in an intron.

FIGURE 18.6 Evidence That Enhancers Are Required for Transcription

QUESTION Why did the researchers assess mRNA production in human cells that did not receive a modified or intact gene?

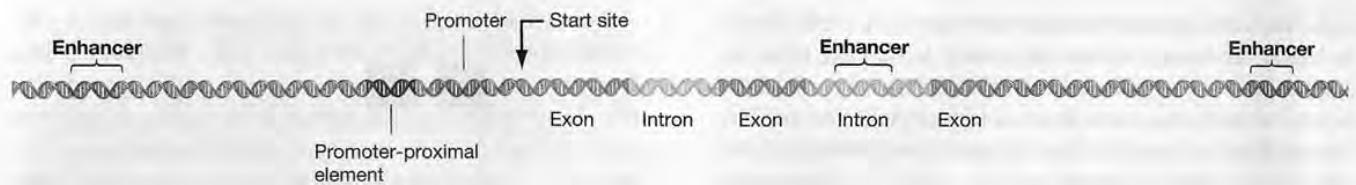


FIGURE 18.7 Enhancers Are Far from the Genes They Regulate

Eukaryotic genes often have more than one enhancer. Enhancers can be located either 5' or 3' to the gene they regulate or in introns and can be located tens of thousands of base pairs away from the promoter.

inserted different versions of the gene into mouse cells, which do not normally contain the antibody-producing gene. Some mouse cells received normal copies of the gene; some received copies of the gene missing different portions of the intron; some received no gene (step 4). Finally, the researchers analyzed the mRNAs that were produced and compared them with mRNA from a normal human cell and from mouse cells that received the normal gene. If their hypothesis was correct and a regulatory sequence was located inside the intron, then some of the modified genes should lack that sequence. Transcription should be reduced in genes lacking the regulatory sequence.

As the “Results” of Figure 18.6 shows, some of the modified copies of the gene were not transcribed at all. The message is clear: The gene is not transcribed if certain parts of the intron are missing. Tonegawa and co-workers proposed that the intron contains a regulatory sequence that is required for transcription to be activated.

This result was remarkable for two reasons: (1) the regulatory sequence was thousands of bases away from the promoter, and (2) it was downstream of the promoter instead of upstream. Regulatory elements that are far from the promoter are termed **enhancers**. Follow-up work has shown that enhancers occur in all eukaryotes and that they have several key characteristics:

- Enhancers can be more than 100,000 bases away from the promoter. They can be located in introns or in untranscribed 5' or 3' sequences flanking the gene (see Figure 18.7). Researchers have yet to find enhancers located in exons.
- Like promoter-proximal elements, many types of enhancers exist. Different genes are associated with different enhancers.
- Enhancers can work even if their normal 5' → 3' orientation is flipped.
- Enhancers can work even if they are moved to a new location in the vicinity of the gene, on the same chromosome.

Enhancers are regulatory sequences that are unique to eukaryotes. In addition, eukaryotic genomes contain regulatory sequences that are similar in structure to enhancers but opposite in function. These sequences are **silencers**. When silencers are active, they shut down transcription.

Once enhancers and silencers had been characterized, it was clear that they represented a type of regulatory sequence that worked very differently than promoter-proximal elements did.

For example, there was good evidence that GAL4 and other types of regulatory proteins bind to promoter-proximal elements and interact directly with TBP or RNA polymerase. These interactions stabilize binding and promote transcription. But how could a regulatory sequence that is distant from the promoter help initiate transcription?

How Do Enhancers Work?

An experiment that followed up on Tonegawa's work with antibody-producing genes supported the hypothesis that enhancers are binding sites for proteins that regulate transcription. As Figure 18.8 shows, researchers cut the enhancer sequence from an antibody-producing gene and spliced it into copies of the gene that codes for the protein β -globin. When this modified gene was inserted into antibody-producing cells, which normally do not express the gene for β -globin, the β -globin mRNA was produced. This result suggests that antibody-producing cells contain some factor that interacts specifically with the antibody-producing gene's enhancer to induce transcription. Follow-up work showed that the factor is a regulatory protein. Enhancers are regulatory sequences where regulatory proteins bind.

By analyzing mutant yeast, fruit flies, and roundworms that have defects in the expression of particular genes, biologists have identified a large number of regulatory proteins that bind to enhancers and silencers. This research program has supported one of the most general conclusions about gene regulation in eukaryotes: In multicellular species, different types of cells express different genes because they contain different regulatory proteins. These regulatory proteins, in turn, are produced in response to signals that arrive from other cells early in embryonic development. For example, a signaling molecule might arrive at a cell early in development and trigger the production of regulatory proteins that are specific to muscle cells. Because the regulatory proteins bind to specific enhancers and silencers and promoter-proximal elements, they trigger the production of muscle-specific proteins (Figure 18.9). If no “become a muscle cell” signal arrives, then no muscle-specific regulatory proteins are produced and no muscle-specific gene expression takes place.

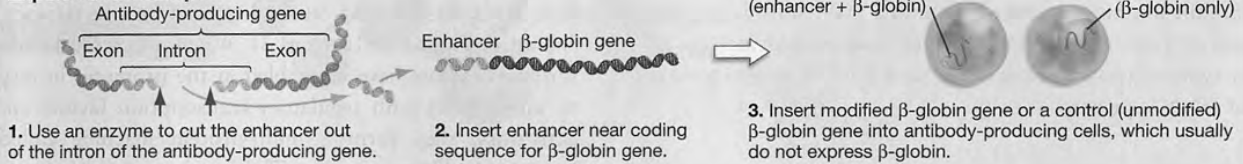
To summarize, differential gene expression is based on the production of specific regulatory proteins. Eukaryotic genes are turned on when specific regulatory proteins bind to enhancers and promoter-proximal elements; the genes are turned off

Question: What is the role of enhancers in gene expression?

Hypothesis: Enhancers are binding sites for regulatory proteins that activate transcription.

Null hypothesis: Enhancers are not binding sites for regulatory proteins that activate transcription.

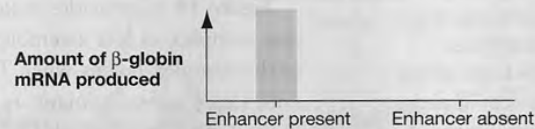
Experimental setup:



Prediction: β -Globin mRNAs will be produced in the cells with enhancer present but not in the control.

Prediction of null hypothesis: β -Globin mRNAs will not be produced in either group of cells.

Results:



Conclusion: Antibody-producing cells must contain a factor that interacts specifically with the enhancer for the antibody-producing gene and results in transcription.

FIGURE 18.8 Evidence That Enhancers Are Involved in Tissue-Specific Expression of Genes

Researchers attached the enhancer sequence from an antibody-producing gene to the coding sequence for the β -globin gene. The altered gene was expressed in antibody-producing cells, which normally do not produce β -globin.

EXTRACELLULAR SIGNALS TRIGGER CELL-SPECIFIC GENE EXPRESSION.

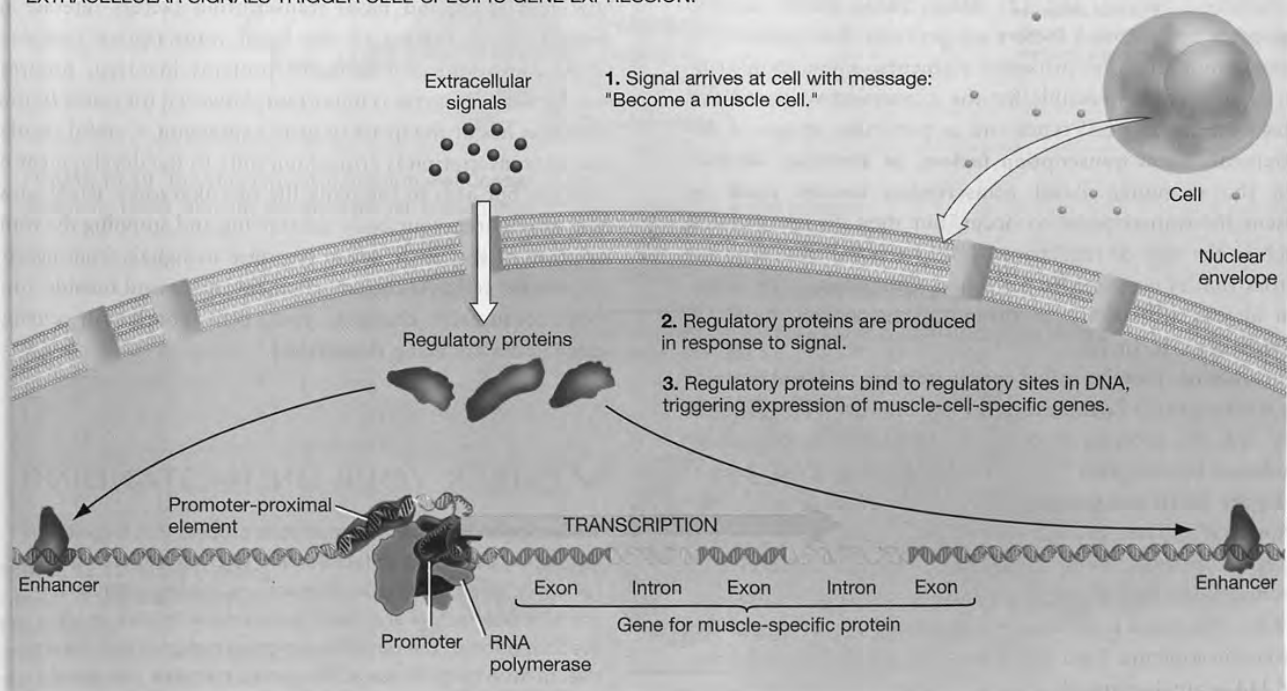


FIGURE 18.9 Why Are Certain Proteins Produced Only in Certain Types of Cells?

Signaling molecules from outside the cell trigger the production or activation of cell-specific regulatory proteins, which then influence transcription by binding to enhancers, silencers, or promoter-proximal elements in DNA.

when regulatory proteins bind to silencers. Distinctive regulatory proteins are what make a muscle cell a muscle cell and a bone cell a bone cell.

How do the interactions between regulatory proteins and DNA sequences control transcription? As we'll see in the next section, the initiation of transcription is a marvelously complex process in eukaryotes. It requires the coordinated activity of many types of proteins interacting with each other and a variety of DNA sequences.

✓CHECK YOUR UNDERSTANDING

Eukaryotic genes have regulatory sequences called promoter-proximal elements close to their promoters and regulatory sequences called enhancers or silencers far from their promoters. You should be able to (1) explain the logic of the experiment that led to the discovery of an enhancer in an intron, and (2) compare and contrast regulatory sequences in the DNA of bacteria versus those in eukaryotes.

18.4 Transcription Initiation

After several decades of research, many questions remain about how transcription is initiated in eukaryotes. It is clear, however, that two broad classes of proteins interact with regulatory sequences at the start of transcription: (1) regulatory transcription factors and (2) basal transcription factors. **Regulatory transcription factors** are proteins that bind to enhancers and promoter-proximal elements. These transcription factors are responsible for the expression of particular genes in particular cell types and at particular stages of development. **Basal transcription factors**, in contrast, interact with the promoter. Basal transcription factors must be present for transcription to occur, but they do not provide much in the way of regulation. TBP is a basal transcription factor that is common to all genes. Other basal transcription factors are specific to promoters recognized by RNA polymerase I, II, or III.

In addition, proteins called **coactivators** are involved in starting transcription. Coactivators do not bind to DNA. Instead they link the proteins involved in initiating transcription—regulatory transcription factors and basal transcription factors.

Figure 18.10 summarizes the current model for how transcription is initiated in eukaryotes. The first step is the binding of regulatory transcription factors to DNA; these factors recruit chromatin-remodeling complexes and histone acetyl transferases (HATs). The result is chromatin remodeling, a loosening of the chromatin structure. Once the chromatin-remodeling complexes and HATs are in place, they open a broad swath of chromatin that includes the promoter region (step 2).

Once chromatin's grip is relaxed, other regulatory transcription factors can bind enhancers and promoter-proximal

elements. When these regulatory transcription factors are bound to DNA, they interact with basal transcription factors. Because enhancers are located far from the promoter, the regulatory transcription factors are normally far from the basal transcription factors. For the two sets of proteins to make contact, DNA has to loop out and away from the promoter, as shown in Figure 18.10 (step 3). When all of the basal transcription factors have assembled at the promoter in response to interactions with regulatory transcription factors and coactivators, they form a multi-protein machine called the **basal transcription complex**. The basal transcription complex then "recruits" RNA polymerase II so that transcription can begin (step 4).

Figure 18.11 provides a closer look at the basal transcription complex as it is assembled at the promoter. Construction of the complex begins when TBP binds to the TATA box. In a multistep process, as many as 60 other proteins then assemble around the DNA-bound TBP. The assembly process depends on interactions with regulatory transcription factors that are bound to enhancers, silencers, and promoter-proximal elements; the result is the formation of a basal transcription complex that can position RNA polymerase II in a way that initiates transcription. Compared with the situation in bacteria, where just 3 to 5 proteins may interact at the promoter to initiate transcription, the state of affairs in eukaryotes is remarkably complicated.

Currently, biologists are focused on understanding exactly how regulatory and basal transcription factors interact to control the formation of the basal transcription complex. Given the number of different proteins involved, progress may be slow. Progress is important, however, for transcription initiation lies at the heart of gene expression. Careful regulation of transcription is critical not only to the development of embryos but also to the daily life of eukaryotes. Right now, cells throughout your body are starting and stopping the transcription of specific genes in response to signals from nearby and distant cells. As the environment inside and outside your body continually changes, your cells continually change which genes are being transcribed.

✓CHECK YOUR UNDERSTANDING

Transcription initiation is a multistep process that begins when regulatory transcription factors bind to DNA and recruit proteins that open chromatin. Interactions between regulatory transcription factors and basal transcription factors result in the formation of the basal transcription complex and the arrival of RNA polymerase at the gene's start site. You should be able to compare and contrast the sigma proteins and regulatory proteins observed in bacteria with the basal and regulatory transcription factors found in eukaryotes.

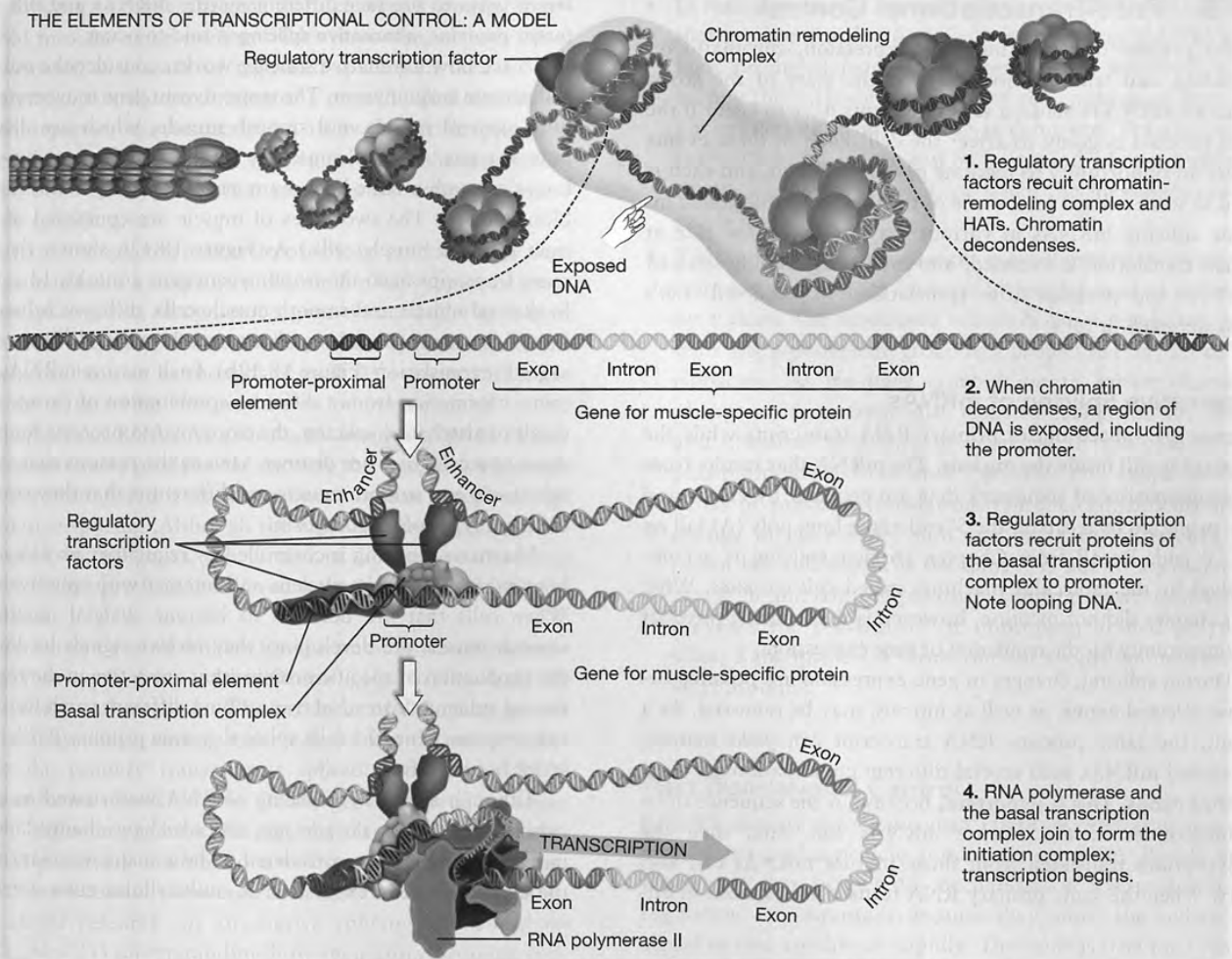


FIGURE 18.10 The Elements of Transcriptional Control: A Model

According to the currently accepted model, transcription is initiated through a series of steps.

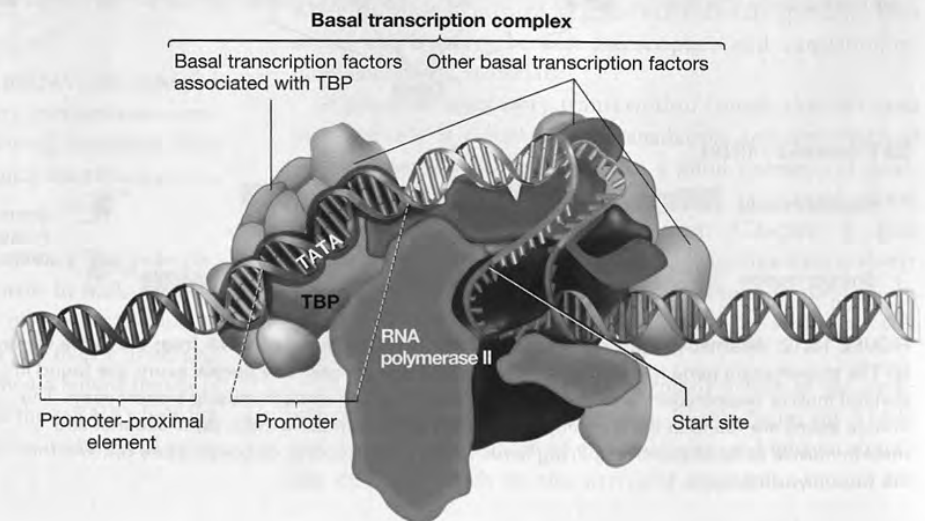


FIGURE 18.11 Many Proteins Interact at the Start Site for Transcription

Basal transcription factors associate with TBP when it binds to the promoter. The basal transcription complex also includes other basal transcription factors and RNA polymerase II.

18.5 Post-Transcriptional Control

In the process of regulating gene expression, chromatin remodeling and transcription are just the start of the story. Once an mRNA is made, a series of events have to occur if the final product is going to affect the cell. Each of these events offers an opportunity to regulate gene expression, and each is used in some cells at least some of the time. Control points include splicing mRNAs in various ways, altering the rate at which translation is initiated, and modifying the life span of mRNAs and proteins after translation has occurred. Let's consider each in turn.

Alternative Splicing of mRNAs

Introns are spliced out of primary RNA transcripts while the message is still inside the nucleus. The mRNA that results from splicing consists of sequences that are encoded by exons and are protected by a cap on the 5' end and a long poly (A) tail on the 3' end. Recall from Chapter 16 that splicing is accomplished by the molecular machines called *spliceosomes*. What that chapter did not mention, however, is that splicing provides an opportunity for the regulation of gene expression.

During splicing, changes in gene expression are possible because selected exons, as well as introns, may be removed. As a result, the same primary RNA transcript can yield mature, processed mRNAs with several different combinations of transcribed exons. This is important, because, if the sequence of ribonucleotides in the mature mRNAs can vary, then the polypeptides translated from those mature mRNAs can also vary. When the same primary RNA transcript is spliced in dif-

ferent ways to produce different mature mRNAs and thus different proteins, **alternative splicing** is said to occur.

To see how alternative splicing works, consider the muscle-cell protein tropomyosin. The tropomyosin gene is expressed in both skeletal muscle and smooth muscle, which are distinct muscle types. (Skeletal muscle is responsible for moving your bones; smooth muscle lines many parts of your gut and certain blood vessels. The two types of muscle are composed of distinct types of muscle cells.) As **Figure 18.12a** shows, the primary transcript from the tropomyosin gene contains 14 exons. In skeletal muscle and smooth muscle cells, different subsets of the 14 exons are spliced together to produce two different messages for translation (**Figure 18.12b**). Each mature mRNA contains information from a different combination of exons. As a result of alternative splicing, the tropomyosin proteins found in these two cell types are distinct. One of the reasons that skeletal muscle and smooth muscle are different is that they contain different types of tropomyosin.

Alternative splicing is controlled by regulatory proteins that bind to mRNAs in the nucleus and interact with spliceosomes. When cells that are destined to become skeletal muscle or smooth muscle are developing, they receive signals leading to the production of specific proteins that are active in the regulation of splicing. Instead of transcribing different versions of the tropomyosin gene, the cells splice the same primary RNA transcript but in different ways.

Although alternative splicing of RNA was viewed as a curiosity until about a decade ago, attitudes have changed. Alternative splicing is now considered to be a major mechanism in the control of gene expression in multicellular eukaryotes. In

(a) Tropomyosin gene



(b) Processed mRNAs



FIGURE 18.12 Alternative Splicing Produces More than One Mature mRNA from the Same Gene

(a) The tropomyosin gene has a large series of exons and introns. The purple exons are found in skeletal muscle tropomyosin, and the blue exons are found in smooth muscle tropomyosin. The orange exons are found in the tropomyosin of both types of muscle cells. (b) In skeletal and smooth muscle cells, alternative splicing leads to the production of distinct mature mRNAs for the tropomyosin protein.

fruit flies, for example, the gene that determines gender is transcribed into the same primary RNA transcript in both males and females. A cascade of regulated splicing in females ultimately results in the production of a protein called *doublesex*, which represses the transcription of male-specific genes. In males, an alternative series of splicing events results in a version of the doublesex protein that represses the transcription of female-specific genes.

Alternative splicing makes it possible to produce more than one protein from a single gene. Before the importance of alternative splicing was recognized, a gene was considered to be a nucleotide sequence that encodes a single protein or RNA. Now genes have to be thought of as DNA sequences that are capable of directing the production of one or more related polypeptides or RNAs. As a result, the number of proteins that an organism can produce far exceeds the number of genes it has. For example, at least 35 percent of human genes undergo alternative splicing. Although the human genome contains an estimated 40,000 genes, the number of different proteins that your cells can produce is believed to be between 100,000 and 1 million.

The current record holder for the number of mRNAs derived from one gene is the *Dscam* gene in the fruit fly *Drosophila melanogaster*. The products of this gene are involved in guiding growing nerve cells within the embryo. Because the primary transcript is spliced into about 38,000 distinct forms of mRNA, the *Dscam* gene can produce about 38,000 different products. Researchers are testing the hypothesis that particular combinations of Dscam proteins allow particular nerve cells to connect to each other in the fly's brain.

Current research on alternative splicing is focused on two topics: (1) understanding how regulatory proteins control the process and (2) documenting the extent to which alternative splicing enriches the genetic information stored in a genome. Stay tuned.

Translational Control

Once splicing is complete and processed mRNAs are exported to the cytoplasm, several new regulatory mechanisms come into play prior to translation. The following examples illustrate the array of mechanisms that influence the timing or extent of translation:

- The life span of an mRNA in the cell can vary. For example, the mRNA for casein—the major protein in milk—is produced in the mammary gland tissue of female mammals. Normally this mRNA has a half-life of just over 1 hour, and little casein protein is produced. But when a female mouse is lactating, regulatory molecules increase the mRNA's half-life to 28.5 hours—leading to a huge increase in the production of casein protein.

- In many species of animals, eggs are loaded with mRNAs that are not translated until fertilization occurs. The proteins produced from these messages are involved in directing the early development of the embryo, so they are not needed until fertilization is complete. Translation of the mRNAs in the egg can be prevented when a regulatory protein binds to them, or when the message's cap or tail is modified.
- The overall rate of translation in a cell may slow or stop in response to a sudden increase in temperature or infection by a virus. The slowdown occurs because regulatory proteins add a phosphate group to a protein that is part of the ribosome. As we have pointed out in earlier chapters, phosphorylation frequently leads to changes in the shape and chemical reactivity of proteins. In the case of the phosphorylated ribosomal protein, the shape change slows or prevents translation. For the cell, this dramatic change in gene expression can mean the difference between life and death. If the danger is due to a sudden increase in ambient temperature, shutting down translation prevents the production of improperly folded polypeptides; if the insult is a virus, the cell avoids manufacturing viral proteins.

Post-Translational Control

Control of gene expression may continue even after translation occurs and a protein product is complete. Recall from Chapter 17 that, in bacteria, mechanisms of post-translational regulation are important because they allow the cell to respond to new conditions rapidly. The same is true for eukaryotes. Instead of waiting for transcription and translation to occur, the cell can respond to altered conditions by quickly activating or inactivating existing proteins. Regulatory mechanisms occurring late in the flow of information from DNA to RNA to protein involve a trade-off between speed and resource use, however, because transcription and translation require energy and materials.

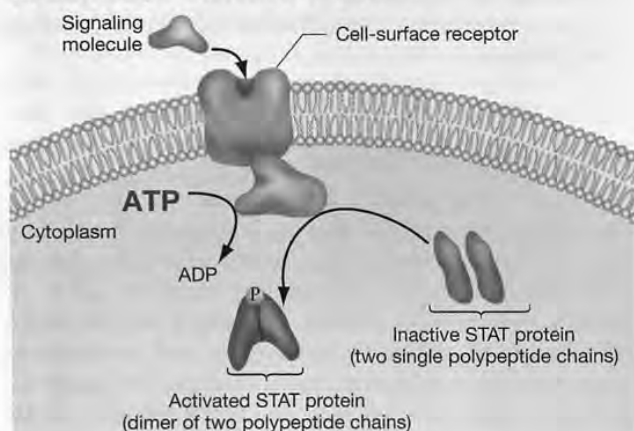
A group of regulatory transcription factors that is found in mammals is called **signal transducers and activators of transcription (STATs)** and provides a good example of post-translational control. STATs participate in several signal transduction pathways. Recall from Chapter 8 that signal transduction pathways are interactions that convert a signal at the cell surface into a response by the cell. In white blood cells, STAT activity results in increased cell growth and cell division.

In white blood cells, STATs normally reside in the cytoplasm as single polypeptide chains. In this form, the STATs are inactive. But a chain of events is triggered when a signaling molecule binds to and activates a receptor on the cell

surface (Figure 18.13a). The activated receptor adds a phosphate to a STAT polypeptide chain, triggering the formation of a **dimer** (a unit made of two parts). The activated STAT dimer then moves into the nucleus, binds to an enhancer, and activates transcription (Figure 18.13b). The STAT remains active until the phosphate is removed by another regulatory protein. At this point, the two polypeptide chains come apart and are transported back to the cytoplasm.

Phosphorylation is a very common mechanism of post-translational control over gene expression, particularly in signal transduction pathways. It is only one of several post-translational control mechanisms that have been documented, however. The activity of a protein may also be modified by folding or by enzymes that cleave off a portion of the molecule. But no matter what the mechanism, post-translational control is correlated with particularly rapid changes in gene expression.

(a) Growth signal activates STAT protein.



(b) Activated STAT protein stimulates transcription.

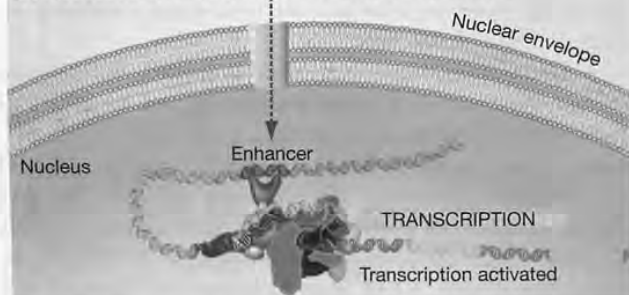


FIGURE 18.13 STATs Activate Transcription after They Are Phosphorylated

(a) A series of events, beginning with the binding of a signaling molecule to a cell-surface receptor, leads to the phosphorylation of STAT proteins. This causes two STAT polypeptide chains to join, forming an activated dimer. (b) Activated STAT proteins move to the nucleus, bind to DNA, and activate the transcription of selected target genes.

A 50-Year Retrospective: How Does Gene Expression in Bacteria and in Eukaryotes Compare?

For about five decades, biologists have been studying how gene expression is controlled. Almost as soon as they knew that information in DNA is transcribed into RNA and then translated into proteins, researchers began asking questions about how that flow of information is regulated. Early studies on bacteria established the existence of promoters, the structure and function of regulatory proteins and regulatory sequences, and the principles of positive control and negative control. More recent work has highlighted the importance of chromatin structure and alternative splicing in eukaryotes.

What are the fundamental differences in how gene expression is regulated in bacteria and eukaryotes? Biologists point to four primary differences, two of which involve levels of control that exist in eukaryotes but not in bacteria:

1. **Packaging** The chromatin structure of eukaryotic DNA must be opened in order for TBP, the basal transcription complex, and RNA polymerase to gain access to genes and initiate transcription. A key insight here is that, because eukaryotic DNA is packaged so tightly, the default state of transcription in eukaryotes is off. In contrast, the default state of transcription in bacteria, with their supercoiled DNA and freely accessible promoters, is on.
2. **Alternative splicing** Prior to translation, primary transcripts in eukaryotes must be spliced—an event that does not occur in bacteria. The fundamental consequence is that the one-to-one correspondence between the number of genes and the number of gene products observed in bacteria is not seen in eukaryotes. Instead, each eukaryotic gene may code for one to thousands of distinct products.
3. **Complexity** Transcriptional control is much more complex in eukaryotes than in bacteria. The function of sigma proteins in bacteria is analogous to the role of the basal transcription complex in eukaryotes. Likewise, the function of CAP, the repressor, and other regulatory proteins is analogous to the role of regulatory transcription factors in eukaryotes. But the sheer number of eukaryotic proteins involved in regulating transcription—and the complexity of their interactions—dwarfs those in bacteria.
4. **Coordinated expression** In bacteria, genes that are involved in the same cellular response are organized into operons controlled by a single promoter. Because their polycistronic mRNAs are translated together, several proteins are produced in a coordinated fashion. In contrast, operons are rare in eukaryotes. In these organisms, genes that are physically scattered can be expressed at the same time because a single set of regulatory transcription factors can trigger the transcription of several genes. For example, muscle-specific genes found on several different chromosomes can be transcribed in response to the same muscle-specific regulatory transcrip-

tion factor. In this way, eukaryotes coordinate the expression of functionally related genes.

To explain why these differences exist, biologists point to a fundamental contrast between unicellular and multicellular individuals. In both types of organisms, cells have to respond to environmental changes in an appropriate way. Bacteria and yeast cells change gene expression to respond to changes in the availability of sugars and other nutrients; your muscle cells change gene expression in response to changes in temperature, the availability of oxygen and sugars, and general activity. But in addition, the cells of multicellular organisms have to differentiate as an individual develops. Changes in gene expression are responsible for the differentiation of muscle cells, bone cells, leaf cells, and flower cells in response to signals from other cells. The need for each cell type to have a unique pattern of gene expression is thought to be the reason why control of gene expression is so much more complex in eukaryotes than in bacteria.

One of the two great frontiers in research on gene expression is to understand how developmental signals produce cell-specific gene expression in multicellular organisms. The other major frontier is to understand how certain defects in gene regulation result in uncontrolled cell growth and the suite of diseases called cancer.

18.6 Linking Cancer with Defects in Gene Regulation

Normal regulation of gene expression results in the orderly development of an embryo and appropriate responses to environmental changes. Abnormal regulation of gene expression, in contrast, can lead to developmental abnormalities and diseases such as cancer. Hundreds of distinct cancers exist. These diseases are enormously variable regarding the tissues they affect, their rate of progression, and their outcome. Because the underlying defects, symptoms, and consequences are so diverse, cancer is not a single disease but a family of related diseases. Cancers are related because they all result from uncontrolled cell growth.

From this perspective, the fundamental question in cancer biology is a simple one: What causes uncontrolled cell growth? The short answer is that each type of cancer is caused by a different set of defects that lead to uncontrolled cell growth. The longer and more accurate answer is based on two key observations. First, recall from Chapter 11 that cancer results from defects in the proteins that control the cell cycle. Second, cancer is associated with mutations that knock out key genes. For example, Chapter 14 linked the cancer-causing disorder xeroderma pigmentosum and hereditary nonpolyposis colorectal cancer with defects in DNA repair. This association supports the hypothesis that increased mutation rates are involved in triggering cancer. The hypothesis is also supported by the observation that people

who have been exposed to **mutagens**—that is, radiation or chemicals that induce mutation—have an increased risk of developing cancer. For example, cancer rates were extremely high in the tens of thousands of people who were exposed to massive doses of radiation from the atomic bombs dropped on Hiroshima and Nagasaki, Japan, at the end of World War II. The link between mutation and cancer is so strong that suspected **chemical carcinogens** (cancer-causing compounds) are first tested by their ability to cause mutations in bacteria.

The fundamental question of cancer biology comes into sharper focus once the role of cell-cycle regulators and mutation is acknowledged. Now we can ask, Which genes, when mutated, disrupt the cell cycle and trigger uncontrolled cell growth?

Intensive research over the past two decades has shown that many cancers are associated with mutations in regulatory transcription factors or regulatory sequences in DNA. These mutations lead to cancer when they affect one of two classes of genes: (1) genes that stop or slow the cell cycle, and (2) genes that trigger cell growth and division by initiating specific phases in the cell cycle.

Genes that stop or slow the cell cycle are called **tumor suppressor genes**. Their products prevent the cell cycle from progressing unless specific signals indicate that conditions are right for moving forward with mitosis and cell division. If a mutation disrupts normal function of a tumor suppressor gene, then a key “brake” on the cell cycle is eliminated.

Genes that encourage cell growth by triggering specific phases in the cell cycle are called **proto-oncogenes** (“first-cancer-genes”). In normal cells, proto-oncogenes are required to initiate each phase in the cell cycle. They are active only when conditions are appropriate for growth, however. In cancerous cells, defects in the regulation of proto-oncogenes cause these genes to stimulate growth at all times. In cases such as this, a mutation has converted the proto-oncogene into an **oncogene**—an allele that promotes cancer development.

To gain a deeper understanding of how defects in gene expression can lead to cancer, let us consider two examples from recent research. The first, the *p53* gene, illustrates how problems with regulating tumor suppressor genes can lead to uncontrolled cell growth. The second, STAT mutations, demonstrates how defects in regulatory transcription factors can lead to the constant activation of proto-oncogenes, constant cell growth, and tumor formation.

p53—Guardian of the Genome

The gene that is most often defective in human cancers codes for a regulatory transcription factor. This gene is called **p53**, because the protein it codes for has a molecular weight of approximately 53 kilodaltons (53,000 daltons, or 53,000 amu—see Chapter 2). DNA sequencing studies have revealed that mutant, nonfunctional forms of *p53* are found in over half of all human cancers.

What is the link between a loss of p53 protein activity and cancer? A key observation suggested an answer to this question: When researchers exposed normal, noncancerous human cells to UV radiation, levels of p53 protein increased markedly. Recall from Chapter 14 that UV radiation damages DNA. Follow-up studies confirmed that there is a close correlation between DNA damage and the amount of p53 in a cell. In addition, analyses of the protein's structure showed that it contains a DNA-binding domain.

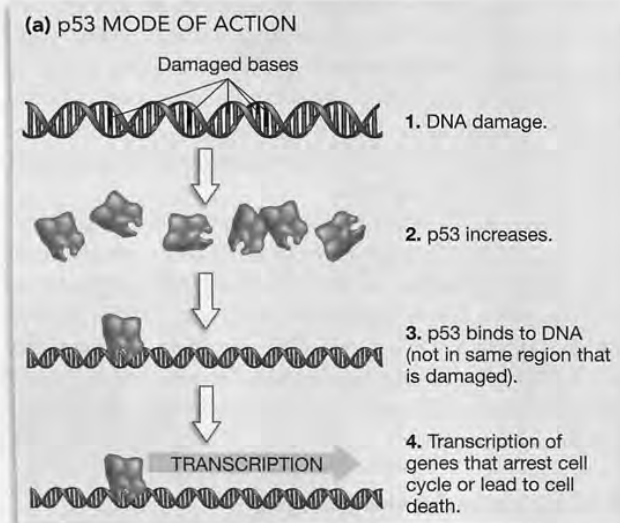
These observations inspired the hypothesis that p53 is a transcription factor that serves as the master brake on the cell cycle (Figure 18.14a). In this model, p53 is activated after DNA damage occurs. The hypothesis contends that the activated protein binds to the enhancers of genes that arrest the cell cycle. Once the cell cycle is arrested, the cell has time to repair its DNA before continuing to grow and divide.

Recent research has shown that this model of p53 function is correct in almost every detail. For example, consider the results of studies based on X-ray crystallography. The three-dimensional models generated by this technique confirmed that p53 binds directly to DNA (Figure 18.14b). In addition, researchers who mapped the location of mutations that make p53 defective found that virtually all of the cancer-causing mutations were located in p53's DNA-binding site (see the gray regions of Figure 18.14b). This observation supports the hypothesis that defective forms of the protein can't bind to en-

hancers. Further, investigators have documented that one of the genes induced by p53 encodes a protein that binds to the cyclin-Cdk complexes introduced in Chapter 11. Binding by this protein prevents the cyclin-Cdk complexes from triggering M (mitosis) phase in the cell cycle.

Recent research has also shown that when a cell's DNA is extensively damaged and cannot be repaired, p53 activates the transcription of genes that cause the cell to take its own life via apoptosis (see Chapter 11). But if mutations in the *p53* gene make the protein product inactive, then damaged cells are not shut down or killed. They continue to move through the cell cycle, except now they are likely to contain many mutations because of the DNA damage they have left unrepaired. If these mutations create oncogenes or inactivate genes for tumor suppressors, the cells have taken a key step on the road to cancer.

To summarize, *p53* functions as a tumor suppressor gene. It prevents cancer onset by stopping the cell cycle when DNA is damaged. When *p53* is functioning normally, mutations that produce oncogenes are repaired or eliminated before uncontrolled cell growth can begin. The role of *p53* in preventing cancer is so fundamental that biologists call this gene "the guardian of the genome." Currently, research on the p53 protein is forging ahead on two fronts: Biologists are striving to identify more of the genes that are regulated by this protein and to find molecules that could act as anticancer drugs by mimicking p53's shape and activity.



(b) DNA-binding domain of p53

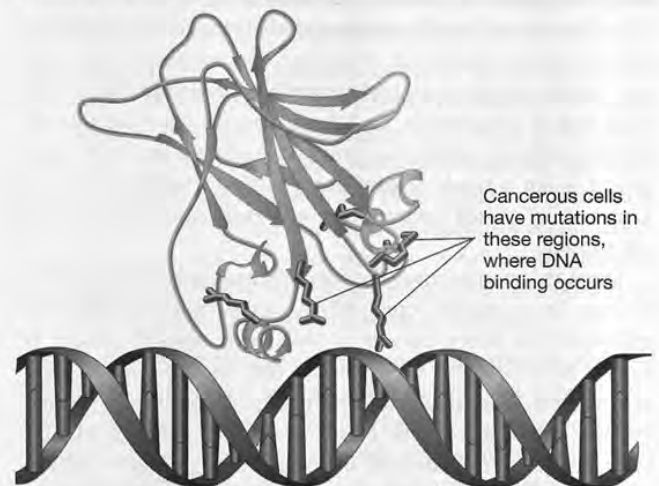


FIGURE 18.14 p53 Is a DNA-Binding Protein That Acts as a Tumor Suppressor

(a) Summary of the current model for p53's mode of action. (b) Ribbon model showing the DNA-binding domain of p53. In cancer patients, the amino acids in the gray portions of the molecule are often different than the normal versions. **QUESTION** Why would substituting one amino acid for another at one of the locations highlighted in part (b) affect a protein's ability to bind to DNA? If the model in part (a) is correct, what happens when p53 cannot bind to DNA efficiently?

STAT Mutations in Cancer

The family of transcription activators called STATs is another type of protein implicated in cancer. Recall from Section 18.5 that, in several types of cells, STATs activate genes that promote cell growth in response to signals from outside the cell. In normal cells, STATs must be phosphorylated before they can bind to regulatory sequences in DNA and activate transcription. But in some human cancers, STATs bind to DNA all the time, or constitutively—not just when they are phosphorylated. When this happens, STATs constantly stimulate cell growth and division. These observations suggest that mutations can convert STATs into oncogenes.

What type of mutations could turn STATs into oncogenes? To answer this question, recall that the activated form of a STAT is a dimer—a combination of two polypeptides. For a dimer to form, an amino acid residue at the point of contact between the two STAT proteins must be phosphorylated (Figure 18.15a).

These observations inspired a hypothesis that STATs turn into oncogenes if mutations cause them to form dimers without

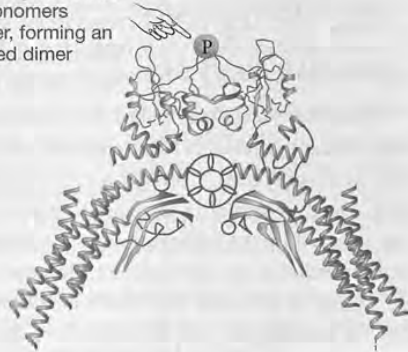
being phosphorylated—that is, without being stimulated by a normal growth signal. More specifically, researchers suggested that if mutations caused cysteine residues to be substituted in the portion of the protein where dimer formation takes place, then disulfide bonds might form between STAT3 monomers, creating constitutively activated dimers (Figure 18.15b).

To test this hypothesis, researchers created mutant versions of the STAT3 protein with cysteines in the appropriate location. The researchers expressed the protein in normal (control) cells and analyzed the amount of protein produced by a gene regulated by STAT3. Compared with cells with normal STAT3 proteins, the cysteine-containing mutant cells produced 10 times as much of the target gene product (Figure 18.15c). This observation was consistent with the biologists' prediction that cysteine mutants form dimers and bind to DNA constitutively.

To confirm their result, the biologists injected mice either with cells that expressed cysteine-mutant STAT3 or with cells that expressed normal STAT3. After just four weeks, all of the

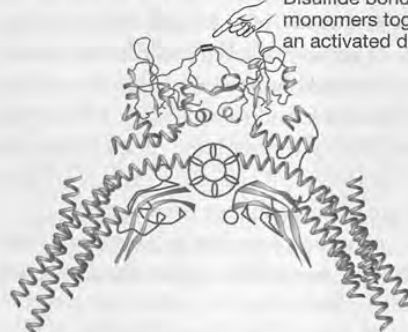
(a) Normal STAT3, phosphorylated

Phosphorylation links two monomers together, forming an activated dimer

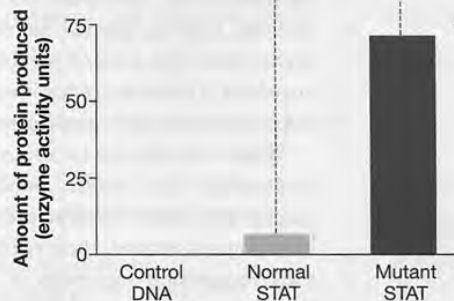


(b) Experimental STAT3, with cysteines at binding site for dimer formation

Disulfide bonds link two monomers together, forming an activated dimer



(c) Consequences for transcription rate



Cysteine mutants form dimers spontaneously and bind to DNA constitutively, causing an abnormally high transcription rate

FIGURE 18.15 Generating a Cancer-Causing Mutation Experimentally

(a) Phosphorylated sections of STAT3 proteins bind to one another, forming a dimer. (b) Substitution of cysteines at the locations shown in red result in the formation of disulfide bonds between STAT3 proteins. As a result, the STAT proteins are activated at all times. (c) Histogram showing the transcriptional activity of a gene that is normally activated by STAT3. The data come from cells that received either a control DNA (a gene that does not encode a transcription activator), a normal STAT3 gene, or a copy of the cysteine-mutant version of the STAT3 gene.

mice that had been injected with mutant cells had developed tumors at the site of injection. In contrast, none of the mice that had been injected with normal cells showed evidence of tumors, even after eight weeks.

These experiments have an important message: Any mutation that constitutively forms dimers of STAT proteins causes constitutive activation of STAT target genes. As a result, these mutations set the stage for cancer. An even more global message is that constant activation of a regulatory transcription factor can lead to cancer. Cancers originate with mutations that disrupt the normal regulation of cell-cycle control genes.

✓ CHECK YOUR UNDERSTANDING

Cancer is associated with mutations and with loss of control over the cell cycle. These two observations are connected, because mutations can cause cancer when they occur in genes that are involved in regulating cell-cycle control genes. For example, when a protein that activates the cell cycle is activated constitutively because of a mutation in a regulatory gene, then uncontrolled cell growth may occur. You should be able to explain why knock-out (loss-of-function) mutations in *p53* occur in so many cancers.

ESSAY Fly Eyes and Other Strange Tales of Gene Regulation

Imagine a 1950s-vintage science-fiction movie. As thunder cracks and lightning flashes, a mad-eyed scientist working in the bowels of a sinister-looking laboratory inserts a mouse gene into a fruit fly's genome. The next day, the scientist cackles with delight as she discovers eyes sprouting in bizarre locations on the fly. But then the fly escapes. It mates with a normal fly and sires offspring with hundreds of eyes. Thanks to their remarkable 360° vision, no predator or flyswatter can stop the little monsters. Clouds of them soar off, bent on global domination.

Preposterous? A team of not-so-mad-eyed scientists has inserted a mouse gene into fruit flies and observed the formation of eyes on the adult fly's legs, antennae, and torso (Figure 18.16). The gene in question is a regulatory transcription factor. When the gene product binds to enhancers, it sets off a cascade of gene expression events that are required for eye development.

Walter Gehring and colleagues were first to isolate and characterize the gene involved. It was called *eyeless*, because it was

originally discovered as a mutant allele that causes flies to develop without eyes. When Gehring's team created *Drosophila* embryos that expressed the normal form of the gene in locations that naturally give rise to appendages, eyes developed on the legs

When eyeless is activated in leg cells, the leg cells give rise to eyes.

and antennae. This result underscores an important conclusion: During development, the fate of a cell depends on the activity of regulatory transcription factors. When *eyeless* is activated in leg cells, the leg cells give rise to eyes.

When the researchers looked for genes in other organisms that were similar to *eyeless* in DNA sequence, the team found them in a wide array of species. Consider a mouse gene called *Small Eye* or *pax6*, which is similar in sequence to *eyeless*. The name *Small Eye* was inspired by the observation that a mutant allele of this gene leads to reduced eye development. Because *eyeless* and *Small Eye* are similar in structure and function, Gehring hypothesized that they are **homologous**—that the genes are descended from a similar gene in a common ancestor of flies and mice, which, according to the fossil record, lived some 550 million years ago. To test this hypothesis, Gehring inserted the mouse gene into *Drosophila* embryos. *Small Eye* induced the development of fly eyes essentially anywhere in the embryo that it was expressed. He got the same result when he inserted a squid gene that was similar to *eyeless*.

What does this tell us about biology? First, it is reasonable to conclude that *eyeless*, *Small Eye*, and like genes in other species are similar in structure and function because they are homologous. Second, these regulatory transcription factors and their recognition sequences in DNA have been so conserved that the proteins are nearly interchangeable among organisms separated by 550 million years of evolution. Third, one key regulatory transcription factor can unleash a cascade of responses that results in a structure as complex as an eye. Fourth, the same starting point—expression of the *eyeless* gene—can lead to very different ends, depending on the context. In flies, the mouse gene *Small Eye* leads to a fly eye, not a mouse eye.



FIGURE 18.16 *Eyeless* Encodes a Transcription Factor That Induces Eye Development

When *Drosophila* embryos are made to express the *eyeless* gene in cells that gave rise to the legs, adults develop with eyes on their legs.

CHAPTER REVIEW

Summary of Key Concepts

- **Changes in gene expression allow eukaryotic cells to respond to changes in the environment and cause distinct cell types to develop.**

In a multicellular eukaryote, different types of cells are radically different in size, shape, and function—even though they have the same DNA. Cells are different not because they have different genes but because they express different genes.

In embryos, cells begin expressing specific genes in response to signals from other cells. Because different cells receive different signals, one cell might begin expressing muscle-specific genes while a nearby cell expresses bone-specific genes. Once cells mature, they continue to receive signals that are released by other cells and that carry information about changes in the environment. In response, changes in gene expression may occur.

In both embryonic cells and mature cells, several types of changes in gene expression may occur. The transcription of specific genes may be initiated or repressed, mRNAs may be spliced in different ways to produce a different product, the life span of specific mRNAs may be extended or shortened, and the life span or activity of particular proteins may be altered.

- **In eukaryotes, DNA is packaged with proteins into complex structures that must be opened before transcription can occur.**

Eukaryotic DNA is wrapped around histone proteins to form a bead-like nucleosome, which is then coiled into 30-nm fibers and higher-order chromatin structures. Transcription cannot be initiated until the interaction between DNA and histones in chromatin is relaxed. These changes depend on the acetylation of histones and the action of chromatin-remodeling complexes.

- **In eukaryotes, transcription can be initiated only when specific proteins bind to the promoter and to regulatory sequences that may be close to the promoter or far from it.**

Regulatory transcription factors are proteins that bind to regulatory sequences called enhancers and silencers, which

are often located at a distance from the gene in question, or to promoter-proximal sequences near the start of the coding sequence. The first regulatory transcription factors that bind to DNA recruit proteins that loosen the histones' grip on the gene, making the promoter accessible to basal transcription factors. Interactions between regulatory and basal transcription factors lead to the formation of the basal transcription complex. Once this large, multi-protein complex is intact, RNA polymerase is recruited to the site and transcription begins.

Web Tutorial 18.1 Transcription Initiation in Eukaryotes

- **Alternative splicing allows a single gene to code for several different products.**

Once a message has been transcribed, several other regulatory events come into play. Alternative splicing allows a single locus to produce more than one mRNA and more than one protein. Mechanisms of post-translational control, such as the activation of transcription factors by phosphorylation, trigger rapid responses to signals from outside the cell.

- **Cancer can develop when mutations disable genes that regulate cell-cycle control genes.**

The cell cycle is controlled by specific genes. But if mutations alter the regulatory sequences of cell-cycle control genes or the genes for their regulatory transcription factors, then uncontrolled cell growth and tumor formation may result. For example, the cell-cycle regulator *p53* is responsible for stopping the cell cycle when DNA is damaged. If *p53* is mutated in a way that prevents its protein product from binding to DNA, then the cell cycle is not arrested and the damaged DNA is not repaired—leading to mutations. If the mutations alter genes that control the cell cycle, such as STAT genes, cancer may occur. Mutant STAT proteins that promote cancer bind to DNA at all times and trigger the expression of genes that activate the cell cycle, leading to uncontrolled cell growth.

Questions

Content Review

1. What is chromatin?
 - a. the protein core of the nucleosome, which consists of histones
 - b. the 30-nm fiber
 - c. the DNA-protein complex found in eukaryotes
 - d. the histone and non-histone proteins in eukaryotic nuclei
2. What is a tumor suppressor?
 - a. a gene associated with tumor formation when its product does not function
 - b. a gene associated with tumor formation when its product functions normally
 - c. a gene that accelerates the cell cycle and leads to uncontrolled cell growth
 - d. a gene that codes for a transcription factor involved in tumor formation
3. Which of the following statements about enhancers is correct?
 - a. They contain a unique base sequence called a TATA box.
 - b. They are located only in 5'-flanking regions.
 - c. They are located only in introns.
 - d. They are found in a variety of locations and are functional in any orientation.
4. In eukaryotes, why are certain genes expressed only in certain types of cells?
 - a. Regulatory transcription factors vary from cell to cell.
 - b. The promoter sequence varies from cell to cell.
 - c. The location of enhancers varies from cell to cell.
 - d. The type of RNA polymerase varies from cell to cell.

400 Unit 3 Gene Structure and Expression

5. What is alternative splicing?
 - a. the phosphorylation events that lead to different types of post-translational regulation
 - b. mRNA processing events that lead to different combinations of exons being spliced together
 - c. folding events that lead to proteins with alternative conformations
 - d. action by regulatory proteins that leads to changes in the life span of an mRNA
6. What types of proteins bind to promoter-proximal elements?
 - a. the basal transcription complex
 - b. the basal transcription complex plus RNA polymerase
 - c. basal transcription factors
 - d. regulatory transcription factors

Conceptual Review

1. Compare and contrast (a) enhancers and the CAP site; (b) promoter-proximal elements and the *lac* operon operator; (c) basal transcription factors and sigma; and (d) eukaryotic promoters and bacterial promoters.
2. Why does chromatin need to be “remodeled” for transcription to occur? Why was it logical to observe that DNA-protein interactions can be loosened by the addition of acetyl groups to histones?
3. Compare and contrast (a) enhancers and silencers; (b) promoter-proximal elements and enhancers; (c) transcription factors and coactivators; and (d) regulatory transcription factors and basal transcription factors.
4. Explain the logic behind the experiment by Tonegawa and colleagues that was responsible for the discovery of enhancers. Which treatments in the experiment served as controls, and what did they control for?
5. When the human genome was first sequenced, researchers estimated that only about 30,000 genes were present. Previously, most investigators expected that the human genome held at least 100,000 genes. Explain how alternative splicing can explain the discrepancy between the observed and predicted numbers of genes.
6. Explain why mutations in regulatory proteins and regulatory sequences can lead to loss of control over the cell cycle and the development of cancer. Include a specific example.

Group Discussion Problems

1. Histone proteins have been extremely highly conserved during evolution. The histones found in fruit flies and humans, for example, are nearly identical in amino acid sequence. Offer an explanation for this observation. (Hint: What are the consequences of a mutation in a histone?)
2. Cancers are most common in tissues where cell division is common, such as blood cells and cells in the lining of the lungs or gut. Why is this observation logical?
3. Levels of p53 protein in the cytoplasm increase after DNA damage. Design an experiment to determine whether this increase is due to increased transcription of the *p53* gene or to activation of preexisting p53 proteins via a post-translational mechanism such as phosphorylation.
4. The frequency of cancer was elevated in individuals who were exposed to the nuclear bombs that were dropped on Japan in World War II. How are these data related to the observation that individuals who experience severe sunburn or deep tanning are more likely to develop skin cancer?

Answers to Multiple-Choice Questions 1. c; 2. a; 3. d; 4. a; 5. b; 6. d

www.prenhall.com/freeman is your resource for the following: Web Tutorials; Online Quizzes and other Online Study Guide materials; Answers to Conceptual Review Questions; Solutions to Group Discussion Problems; Answers to Figure Caption Questions and Exercises; and Additional Readings and Research.