

*Fifth
Edition*

Fifth Edition

Molecular Biology

Molecular Biology

CBN BRNCH
GenColl

QH 506
.W43
2012
Copy 1

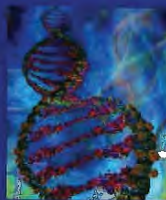
Mc
Graw
Hill

Robert F. Weaver



Fifth
Edition

Molecular Biology



About the Cover

On the cover is a beautiful artist's impression of the DNA double helix. It is not intended to depict the DNA structure accurately, but instead represents the double-helical nature of one DNA duplex against an abstract background of other DNA molecules.



Do More

**McGraw-Hill Higher Education and
Blackboard® have teamed up!**

What does this mean for you?

Now, you can do more. From within your Blackboard course, you can access McGraw-Hill Create™, select any or all content you want, and seamlessly integrate it into your Blackboard course materials. Contact your sales representative for more information.

CBN BRNCH
GenCell

QH 506
.W43
2012
Copy 1

The McGraw-Hill Companies



Molecular Biology

Fifth Edition

Robert F. Weaver

University of Kansas





MOLECULAR BIOLOGY, FIFTH EDITION

Published by McGraw-Hill, a business unit of The McGraw-Hill Companies, Inc., 1221 Avenue of the Americas, New York, NY 10020. Copyright © 2012 by The McGraw-Hill Companies, Inc. All rights reserved. Previous editions © 2008, 2005, and 2002. No part of this publication may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without the prior written consent of The McGraw-Hill Companies, Inc., including, but not limited to, in any network or other electronic storage or transmission, or broadcast for distance learning.

Some ancillaries, including electronic and print components, may not be available to customers outside the United States.

This book is printed on acid-free paper.

1 2 3 4 5 6 7 8 9 0 QDB/QDB 1 0 9 8 7 6 5 4 3 2 1

ISBN 978-0-07-352532-7

MHID 0-07-352532-4

Vice President & Editor-in-Chief: *Marty Lange*
Vice President EDP/Central Publishing Services: *Kimberly Meriwether David*
Publisher: *Janice Roerig-Blong*
Executive Marketing Manager: *Patrick F. Reidy*
Project Manager: *Robin A. Reed*
Design Coordinator: *Brenda A. Rolves*
Cover Designer: *Studio Montage, St. Louis, Missouri*
Lead Photo Research Coordinator: *Carrie K. Burger*
Cover Image: © *Getty Images RF*
Buyer: *Sandy Ludovissy*
Media Project Manager: *Balaji Sundararaman*
Compositor: *Aptara®, Inc.*
Typeface: *10/12 Sabon*
Printer: *Quad/Graphics*

All credits appearing on page or at the end of the book are considered to be an extension of the copyright page.

Library of Congress Cataloging-in-Publication Data

Weaver, Robert Franklin, 1942-
Molecular biology / Robert F. Weaver.—5th ed.
p. cm.
ISBN 978-0-07-352532-7 (hardcover : alk. paper)
1. Molecular biology. I. Title.
QH506.W43 2011
572.8—dc22

2010051759

BRIEF CONTENTS

About the Author	iv
Preface	xiii
Acknowledgments	xvii
Guide to Experimental Techniques in Molecular Biology	xix

PART I

Introduction

- 1 A Brief History 1
- 2 The Molecular Nature of Genes 12
- 3 An Introduction to Gene Function 30

PART II

Methods in Molecular Biology

- 4 Molecular Cloning Methods 49
- 5 Molecular Tools for Studying Genes and Gene Activity 75

PART III

Transcription in Bacteria

- 6 The Mechanism of Transcription in Bacteria 121
- 7 Operons: Fine Control of Bacterial Transcription 167
- 8 Major Shifts in Bacterial Transcription 196
- 9 DNA-Protein Interactions in Bacteria 222

PART IV

Transcription in Eukaryotes

- 10 Eukaryotic RNA Polymerases and Their Promoters 244
- 11 General Transcription Factors in Eukaryotes 273
- 12 Transcription Activators in Eukaryotes 314
- 13 Chromatin Structure and Its Effects on Transcription 355

PART V

Post-Transcriptional Events

- 14 RNA Processing I: Splicing 394
- 15 RNA Processing II: Capping and Polyadenylation 436
- 16 Other RNA Processing Events and Post-Transcriptional Control of Gene Expression 471

PART VI

Translation

- 17 The Mechanism of Translation I: Initiation 522
- 18 The Mechanism of Translation II: Elongation and Termination 560
- 19 Ribosomes and Transfer RNA 601

PART VII

DNA Replication, Recombination, and Transposition

- 20 DNA Replication, Damage, and Repair 636
- 21 DNA Replication II: Detailed Mechanism 677
- 22 Homologous Recombination 709
- 23 Transposition 732

PART VIII

Genomes

- 24 Introduction to Genomics: DNA Sequencing on a Genomic Scale 759
- 25 Genomics II: Functional Genomics, Proteomics, and Bioinformatics 789

Glossary 827

Index 856

C O N T E N T S

About the Author	iv
Preface	xiii
Acknowledgments	xvii
Guide to Experimental Techniques in Molecular Biology	xix

PART I

Introduction

CHAPTER 1

A Brief History 1

- 1.1 Transmission Genetics 2
 - Mendel's Laws of Inheritance 2
 - The Chromosome Theory of Inheritance 3
 - Genetic Recombination and Mapping 4
 - Physical Evidence for Recombination 5
- 1.2 Molecular Genetics 5
 - The Discovery of DNA 5
 - The Relationship Between Genes and Proteins 6
 - Activities of Genes 7
- 1.3 The Three Domains of Life 9

CHAPTER 2

The Molecular Nature of Genes 12

- 2.1 The Nature of Genetic Material 13
 - Transformation in Bacteria 13
 - The Chemical Nature of Polynucleotides 15
- 2.2 DNA Structure 18
 - Experimental Background 19
 - The Double Helix 19
- 2.3 Genes Made of RNA 22
- 2.4 Physical Chemistry of Nucleic Acids 23
 - A Variety of DNA Structures 23
 - DNAs of Various Sizes and Shapes 27

CHAPTER 3

An Introduction to Gene Function 30

- 3.1 Storing Information 31
 - Overview of Gene Expression 31
 - Protein Structure 31
 - Protein Function 35
 - Discovery of Messenger RNA 37
 - Transcription 39
 - Translation 40
- 3.2 Replication 45
- 3.3 Mutations 45
 - Sickle Cell Disease 45

PART II

Methods of Molecular Biology

CHAPTER 4

Molecular Cloning Methods 49

- 4.1 Gene Cloning 50
 - The Role of Restriction Endonucleases 50
 - Vectors 53
 - Identifying a Specific Clone with a Specific Probe 58
 - cDNA Cloning 60
 - Rapid Amplification of cDNA Ends 61
- 4.2 The Polymerase Chain Reaction 62
 - Standard PCR 62
 - Box 4.1** Jurassic Park: More than a Fantasy? 63
 - Using Reverse Transcriptase PCR (RT-PCR) in cDNA Cloning 64
 - Real-Time PCR 64
- 4.3 Methods of Expressing Cloned Genes 65
 - Expression Vectors 65
 - Other Eukaryotic Vectors 71
 - Using the Ti Plasmid to Transfer Genes to Plants 71

CHAPTER 5

Molecular Tools for Studying Genes and Gene Activity 75

- 5.1 **Molecular Separations 76**
 - Gel Electrophoresis 76
 - Two-Dimensional Gel Electrophoresis 79
 - Ion-Exchange Chromatography 80
 - Gel Filtration Chromatography 80
 - Affinity Chromatography 81
- 5.2 **Labeled Tracers 82**
 - Autoradiography 82
 - Phosphorimaging 83
 - Liquid Scintillation Counting 84
 - Nonradioactive Tracers 84
- 5.3 **Using Nucleic Acid Hybridization 85**
 - Southern Blots: Identifying Specific DNA Fragments 85
 - DNA Fingerprinting and DNA Typing 86
 - Forensic Uses of DNA Fingerprinting and DNA Typing 87
 - In Situ Hybridization: Locating Genes in Chromosomes 88
 - Immunoblots (Western Blots) 89
- 5.4 **DNA Sequencing and Physical Mapping 89**
 - The Sanger Chain-Termination Sequencing Method 90
 - Automated DNA Sequencing 91
 - High-Throughput Sequencing 93
 - Restriction Mapping 95
- 5.5 **Protein Engineering with Cloned Genes: Site-Directed Mutagenesis 97**
- 5.6 **Mapping and Quantifying Transcripts 99**
 - Northern Blots 99
 - S1 Mapping 100
 - Primer Extension 102
 - Run-Off Transcription and G-Less Cassette Transcription 103
- 5.7 **Measuring Transcription Rates in Vivo 104**
 - Nuclear Run-On Transcription 104
 - Reporter Gene Transcription 105
 - Measuring Protein Accumulation in Vivo 106
- 5.8 **Assaying DNA-Protein Interactions 108**
 - Filter Binding 108
 - Gel Mobility Shift 109
 - DNase Footprinting 109
 - DMS Footprinting and Other Footprinting Methods 109
 - Chromatin Immunoprecipitation (ChIP) 112
- 5.9 **Assaying Protein-Protein Interactions 112**

- 5.10 **Finding RNA Sequences That Interact with Other Molecules 114**
 - SELEX 114
 - Functional SELEX 114
- 5.11 **Knockouts and Transgenics 115**
 - Knockout Mice 115
 - Transgenic Mice 115

PART III

Transcription in Bacteria

CHAPTER 6

The Mechanism of Transcription in Bacteria 121

- 6.1 **RNA Polymerase Structure 122**
 - Sigma (σ) as a Specificity Factor 122
- 6.2 **Promoters 123**
 - Binding of RNA Polymerase to Promoters 123
 - Promoter Structure 125
- 6.3 **Transcription Initiation 126**
 - Sigma Stimulates Transcription Initiation 127
 - Reuse of σ 128
 - The Stochastic σ -Cycle Model 129
 - Local DNA Melting at the Promoter 132
 - Promoter Clearance 134
 - Structure and Function of σ 139
 - The Role of the α -Subunit in UP Element Recognition 142
- 6.4 **Elongation 144**
 - Core Polymerase Functions in Elongation 144
 - Structure of the Elongation Complex 146
- 6.5 **Termination of Transcription 156**
 - Rho-Independent Termination 156
 - Rho-Dependent Termination 159

CHAPTER 7

Operons: Fine Control of Bacterial Transcription 167

- 7.1 **The *lac* Operon 168**
 - Negative Control of the *lac* Operon 169
 - Discovery of the Operon 169
 - Repressor-Operator Interactions 173
 - The Mechanism of Repression 174
 - Positive Control of the *lac* Operon 177
 - The Mechanism of CAP Action 178

- 7.2 **The *ara* Operon** 182
 - The *ara* Operon Repression Loop 183
 - Evidence for the *ara* Operon Repression Loop 183
 - Autoregulation of *araC* 185
- 7.3 **The *trp* Operon** 186
 - Tryptophan's Role in Negative Control of the *trp* Operon 186
 - Control of the *trp* Operon by Attenuation 187
 - Defeating Attenuation 188
- 7.4 **Riboswitches** 190

CHAPTER 8

Major Shifts in Bacterial Transcription 196

- 8.1 **Sigma Factor Switching** 197
 - Phage Infection 197
 - Sporulation 199
 - Genes with Multiple Promoters 201
 - Other σ Switches 201
 - Anti- σ -Factors 202
- 8.2 **The RNA Polymerase Encoded in Phage T7** 202
- 8.3 **Infection of *E. coli* by Phage λ** 203
 - Lytic Reproduction of Phage λ 204
 - Establishing Lysogeny 211
 - Autoregulation of the *cI* Gene During Lysogeny 212
 - Determining the Fate of a λ Infection: Lysis or Lysogeny 217
 - Lysogen Induction 218

CHAPTER 9

DNA-Protein Interactions in Bacteria 222

- 9.1 **The λ Family of Repressors** 223
 - Probing Binding Specificity by Site-Directed Mutagenesis 223
 - Box 9.1** X-Ray Crystallography 224
 - High-Resolution Analysis of λ Repressor-Operator Interactions 229
 - High-Resolution Analysis of Phage λ Repressor-Operator Interactions 232
- 9.2 **The *trp* Repressor** 234
 - The Role of Tryptophan 234
- 9.3 **General Considerations on Protein-DNA Interactions** 235
 - Hydrogen Bonding Capabilities of the Four Different Base Pairs 235
 - The Importance of Multimeric DNA-Binding Proteins 236

- 9.4 **DNA-Binding Proteins: Action at a Distance** 237
 - The *gal* Operon 237
 - Duplicated λ Operators 237
 - Enhancers 238

PART IV

Transcription in Eukaryotes

CHAPTER 10

Eukaryotic RNA Polymerases and Their Promoters 244

- 10.1 **Multiple Forms of Eukaryotic RNA Polymerase** 245
 - Separation of the Three Nuclear Polymerases 245
 - The Roles of the Three RNA Polymerases 246
 - RNA Polymerase Subunit Structures 248
- 10.2 **Promoters** 259
 - Class II Promoters 259
 - Class I Promoters 263
 - Class III Promoters 264
- 10.3 **Enhancers and Silencers** 267
 - Enhancers 267
 - Silencers 269

CHAPTER 11

General Transcription Factors in Eukaryotes 273

- 11.1 **Class II Factors** 274
 - The Class II Preinitiation Complex 274
 - Structure and Function of TFIID 276
 - Structure and Function of TFIIB 286
 - Structure and Function of TFIIF 288
 - The Mediator Complex and the RNA Polymerase II Holoenzyme 295
 - Elongation Factors 296
- 11.2 **Class I Factors** 299
 - The Core-Binding Factor 299
 - The UPE-Binding Factor 300
 - Structure and Function of SL1 301
- 11.3 **Class III Factors** 303
 - TFIIIA 303
 - TFIIIB and C 304
 - The Role of TBP 307

CHAPTER 12

Transcription Activators in Eukaryotes 314

- 12.1 Categories of Activators 315
 - DNA-Binding Domains 315
 - Transcription-Activating Domains 315
- 12.2 Structures of the DNA-Binding Motifs of Activators 316
 - Zinc Fingers 316
 - The GAL4 Protein 318
 - The Nuclear Receptors 319
 - Homeodomains 320
 - The bZIP and bHLH Domains 321
- 12.3 Independence of the Domains of Activators 323
- 12.4 Functions of Activators 324
 - Recruitment of TFIID 324
 - Recruitment of the Holoenzyme 325
- 12.5 Interaction Among Activators 328
 - Dimerization 328
 - Action at a Distance 329
 - Box 12.1** Genomic Imprinting 332
 - Transcription Factories 334
 - Complex Enhancers 336
 - Architectural Transcription Factors 337
 - Enhanceosomes 338
 - Insulators 339
- 12.6 Regulation of Transcription Factors 343
 - Coactivators 344
 - Activator Ubiquitylation 346
 - Activator Sumoylation 347
 - Activator Acetylation 348
 - Signal Transduction Pathways 348

CHAPTER 13

Chromatin Structure and Its Effects on Transcription 355

- 13.1 Chromatin Structure 356
 - Histones 356
 - Nucleosomes 357
 - The 30-nm Fiber 360
 - Higher-Order Chromatin Folding 362
- 13.2 Chromatin Structure and Gene Activity 364
 - The Effects of Histones on Transcription of Class II Genes 365

- Nucleosome Positioning 367
- Histone Acetylation 372
- Histone Deacetylation 373
- Chromatin Remodeling 376
- Heterochromatin and Silencing 383
- Nucleosomes and Transcription Elongation 387

PART V

Post-Transcriptional Events

CHAPTER 14

RNA Processing I: Splicing 394

- 14.1 Genes in Pieces 395
 - Evidence for Split Genes 395
 - RNA Splicing 396
 - Splicing Signals 397
 - Effect of Splicing on Gene Expression 398
- 14.2 The Mechanism of Splicing of Nuclear mRNA Precursors 399
 - A Branched Intermediate 399
 - A Signal at the Branch 401
 - Spliceosomes 402
 - Spliceosome Assembly and Function 411
 - Commitment, Splice Site Selection, and Alternative Splicing 415
 - Control of Splicing 425
- 14.3 Self-Splicing RNAs 427
 - Group I Introns 427
 - Group II Introns 430

CHAPTER 15

RNA Processing II: Capping and Polyadenylation 436

- 15.1 Capping 437
 - Cap Structure 437
 - Cap Synthesis 438
 - Functions of Caps 440
- 15.2 Polyadenylation 442
 - Poly(A) 442
 - Functions of Poly(A) 443
 - Basic Mechanism of Polyadenylation 445
 - Polyadenylation Signals 446
 - Cleavage and Polyadenylation of a Pre-mRNA 448
 - Poly(A) Polymerase 454
 - Turnover of Poly(A) 454

- 15.3 Coordination of mRNA Processing Events 456**
 Binding of the CTD of Rpb1 to mRNA-Processing Proteins 457
 Changes in Association of RNA-Processing Proteins with the CTD Correlate with Changes in CTD Phosphorylation 458
 A CTD Code? 460
 Coupling Transcription Termination with mRNA 3'-End Processing 461
 Mechanism of Termination 462
 Role of Polyadenylation in mRNA Transport 466

CHAPTER 16

Other RNA Processing Events and Post-Transcriptional Control of Gene Expression 471

- 16.1 Ribosomal RNA Processing 472**
 Eukaryotic rRNA Processing 472
 Bacterial rRNA Processing 474
- 16.2 Transfer RNA Processing 475**
 Cutting Apart Polycistronic Precursors 475
 Forming Mature 5'-Ends 475
 Forming Mature 3'-Ends 476
- 16.3 Trans-Splicing 477**
 The Mechanism of Trans-Splicing 477
- 16.4 RNA Editing 479**
 Mechanism of Editing 479
 Editing by Nucleotide Deamination 482
- 16.5 Post-Transcriptional Control of Gene Expression: mRNA Stability 483**
 Casin mRNA Stability 484
 Transferrin Receptor mRNA Stability 484
- 16.6 Post-Transcriptional Control of Gene Expression: RNA Interference 488**
 Mechanism of RNAi 489
 Amplification of siRNA 494
 Role of the RNAi Machinery in Heterochromatin Formation and Gene Silencing 495
- 16.7 Piwi-Interacting RNAs and Transposon Control 501**
- 16.8 Post-Transcriptional Control of Gene Expression: MicroRNAs 502**
 Silencing of Translation by miRNAs 502
 Stimulation of Translation by miRNAs 507
- 16.9 Translation Repression, mRNA Degradation, and P-Bodies 510**
 Processing Bodies 510
 Degradation of mRNAs in P-Bodies 511
 Relief of Repression in P-Bodies 514
 Other Small RNAs 517

PART VI

Translation

CHAPTER 17

The Mechanism of Translation I: Initiation 522

- 17.1 Initiation of Translation in Bacteria 523**
 tRNA Charging 523
 Dissociation of Ribosomes 523
 Formation of the 30S Initiation Complex 525
 Formation of the 70S Initiation Complex 531
 Summary of Initiation in Bacteria 533
- 17.2 Initiation in Eukaryotes 533**
 The Scanning Model of Initiation 533
 Eukaryotic Initiation Factors 537
- 17.3 Control of Initiation 545**
 Bacterial Translational Control 545
 Eukaryotic Translational Control 548

CHAPTER 18

The Mechanism of Translation II: Elongation and Termination 560

- 18.1 The Direction of Polypeptide Synthesis and of mRNA Translation 561**
- 18.2 The Genetic Code 562**
 Nonoverlapping Codons 562
 No Gaps in the Code 563
 The Triplet Code 563
 Breaking the Code 564
 Unusual Base Pairs Between Codon and Anticodon 566
 The (Almost) Universal Code 567
- 18.3 The Elongation Cycle 569**
 Overview of Elongation 569
 A Three-Site Model of the Ribosome 570
 Elongation Step 1: Binding an Aminoacyl-tRNA to the A Site of the Ribosome 573
 Elongation Step 2: Peptide Bond Formation 577
 Elongation Step 3: Translocation 580
 G Proteins and Translation 582
 The Structures of EF-Tu and EF-G 583
- 18.4 Termination 584**
 Termination Codons 584
 Stop Codon Suppression 586
 Release Factors 586
 Dealing with Aberrant Termination 588
 Use of Stop Codons to Insert Unusual Amino Acids 593

- 18.5 **Posttranslation** 593
 Folding Nascent Proteins 594
 Release of Ribosomes from mRNA 595

CHAPTER 19

Ribosomes and Transfer RNA 601

- 19.1 **Ribosomes** 602
 Fine Structure of the 70S Ribosome 602
 Ribosome Composition 605
 Fine Structure of the 30S Subunit 606
 Fine Structure of the 50S Subunit 612
 Ribosome Structure and the Mechanism of Translation 616
 Polysomes 621
- 19.2 **Transfer RNA** 623
 The Discovery of tRNA 623
 tRNA Structure 623
 Recognition of tRNAs by Aminoacyl-tRNA Synthetase:
 The Second Genetic Code 626
 Proofreading and Editing by Aminoacyl-tRNA
 Synthetases 630

PART VII

DNA Replication, Recombination, and Transposition

CHAPTER 20

DNA Replication, Damage, and Repair 636

- 20.1 **General Features of DNA Replication** 637
 Semiconservative Replication 637
 At Least Semidiscontinuous Replication 639
 Priming of DNA Synthesis 641
 Bidirectional Replication 642
 Rolling Circle Replication 645
- 20.2 **Enzymology of DNA Replication** 646
 Three DNA Polymerases in *E. coli* 646
 Fidelity of Replication 649
 Multiple Eukaryotic DNA Polymerases 650
 Strand Separation 651
 Single-Strand DNA-Binding Proteins 651
 Topoisomerases 653
- 20.3 **DNA Damage and Repair** 656
 Damage Caused by Alkylation of Bases 657
 Damage Caused by Ultraviolet Radiation 658
 Damage Caused by Gamma and X-Rays 658
 Directly Undoing DNA Damage 659

- Excision Repair 660
 Double-Strand Break Repair in Eukaryotes 665
 Mismatch Repair 667
 Failure of Mismatch Repair in Humans 668
 Coping with DNA Damage Without Repairing It 668

CHAPTER 21

DNA Replication II: Detailed Mechanism 677

- 21.1 **Initiation** 678
 Priming in *E. coli* 678
 Priming in Eukaryotes 679
- 21.2 **Elongation** 683
 Speed of Replication 683
 The Pol III Holoenzyme and Processivity
 of Replication 683
- 21.3 **Termination** 694
 Decatenation: Disentangling Daughter DNAs 694
 Termination in Eukaryotes 695
Box 21.1 Telomeres, the Hayflick Limit, and
 Cancer 699
 Telomere Structure and Telomere-Binding Proteins in
 Lower Eukaryotes 702

CHAPTER 22

Homologous Recombination 709

- 22.1 **The RecBCD Pathway for Homologous
Recombination** 710
- 22.2 **Experimental Support for the RecBCD
Pathway** 712
 RecA 712
 RecBCD 715
 RuvA and RuvB 717
 RuvC 719
- 22.3 **Meiotic Recombination** 721
 The Mechanism of Meiotic Recombination:
 Overview 721
 The Double-Stranded DNA Break 722
 Creation of Single-Stranded Ends at DSBs 728
- 22.4 **Gene Conversion** 728

CHAPTER 23

Transposition 732

- 23.1 **Bacterial Transposons** 733
 Discovery of Bacterial Transposons 733
 Insertion Sequences: The Simplest Bacterial
 Transposons 733

- More Complex Transposons 734
- Mechanisms of Transposition 734
- 23.2 Eukaryotic Transposons 737**
 - The First Examples of Transposable Elements: *Ds* and *Ac* of Maize 737
 - P Elements 739
- 23.3 Rearrangement of Immunoglobulin Genes 740**
 - Recombination Signals 742
 - The Recombinase 743
 - Mechanism of V(D)J Recombination 743
- 23.4 Retrotransposons 745**
 - Retroviruses 745
 - Retrotransposons 749

PART VIII

Genomes

CHAPTER 24

Introduction to Genomics: DNA Sequencing on a Genomic Scale 759

- 24.1 Positional Cloning: An Introduction to Genomics 760**
 - Classical Tools of Positional Cloning 760
 - Identifying the Gene Mutated in a Human Disease 762
- 24.2 Techniques in Genomic Sequencing 765**
 - The Human Genome Project 767
 - Vectors for Large-Scale Genome Projects 769
 - The Clone-by-Clone Strategy 770

- Shotgun Sequencing 773
- Sequencing Standards 774

- 24.3 Studying and Comparing Genomic Sequences 774**
 - The Human Genome 774
 - Personal Genomics 779
 - Other Vertebrate Genomes 779
 - The Minimal Genome 782
 - The Barcode of Life 784

CHAPTER 25

Genomics II: Functional Genomics, Proteomics, and Bioinformatics 789

- 25.1 Functional Genomics: Gene Expression on a Genomic Scale 790**
 - Transcriptomics 790
 - Genomic Functional Profiling 799
 - Single-Nucleotide Polymorphisms: Pharmacogenomics 810
- 25.2 Proteomics 812**
 - Protein Separations 812
 - Protein Analysis 813
 - Quantitative Proteomics 814
 - Protein Interactions 816
- 25.3 Bioinformatics 820**
 - Finding Regulatory Motifs in Mammalian Genomes 820
 - Using the Databases Yourself 822

Glossary 827

Index 856

The Molecular Nature of Genes

Before we begin to study in detail the structure and activities of genes, and the experimental evidence underlying those concepts, we need a fuller outline of the adventure that lies before us. Thus, in this chapter and in Chapter 3, we will flesh out the brief history of molecular biology presented in Chapter 1. In this chapter we will begin this task by considering the behavior of genes as molecules.



Computer model of the DNA double helix.

© Comstock Images/Jupiter RF.

2.1 The Nature of Genetic Material

The studies that eventually revealed the chemistry of genes began in Tübingen, Germany, in 1869. There, Friedrich Miescher isolated nuclei from pus cells (white blood cells) in waste surgical bandages. He found that these nuclei contained a novel phosphorus-bearing substance that he named *nuclein*. Nuclein is mostly **chromatin**, which is a complex of deoxyribonucleic acid (DNA) and chromosomal proteins.

By the end of the nineteenth century, both DNA and ribonucleic acid (RNA) had been separated from the protein that clings to them in the cell. This allowed more detailed chemical analysis of these **nucleic acids**. (Notice that the term *nucleic acid* and its derivatives, DNA and RNA, come directly from Miescher's term *nuclein*.) By the beginning of the 1930s, P. Levene, W. Jacobs, and others had demonstrated that RNA is composed of a sugar (ribose) plus four nitrogen-containing bases, and that DNA contains a different sugar (deoxyribose) plus four bases. They discovered that each base is coupled with a sugar-phosphate to form a nucleotide. We will return to the chemical structures of DNA and RNA later in this chapter. First, let us examine the evidence that genes are made of DNA.

Transformation in Bacteria

Frederick Griffith laid the foundation for the identification of DNA as the genetic material in 1928 with his experiments on **transformation** in the bacterium pneumococcus, now known as *Streptococcus pneumoniae*. The wild-type organism is a spherical cell surrounded by a mucous coat called a capsule. The cells form large, glistening colonies, characterized as smooth (S) (Figure 2.1a). These cells are **virulent**, that is, capable of causing lethal infections upon injection into mice. A certain mutant strain of *S. pneumoniae* has lost the ability to form a capsule. As a result, it grows as small, rough (R) colonies (Figure 2.1b). More importantly, it is **avirulent**; because it has no protective coat, it is engulfed by the host's white blood cells before it can proliferate enough to do any damage.

The key finding of Griffith's work was that heat-killed virulent colonies of *S. pneumoniae* could **transform** avirulent cells to virulent ones. Neither the heat-killed virulent bacteria nor the live avirulent ones by themselves could cause a lethal infection. Together, however, they were deadly. Somehow the virulent trait passed from the dead cells to the live, avirulent ones. This transformation phenomenon is illustrated in Figure 2.2. Transformation was not transient; the ability to make a capsule and therefore to kill host animals, once conferred on the avirulent bacteria, was passed to their descendants as a heritable trait. In other words, the avirulent cells somehow gained the gene for

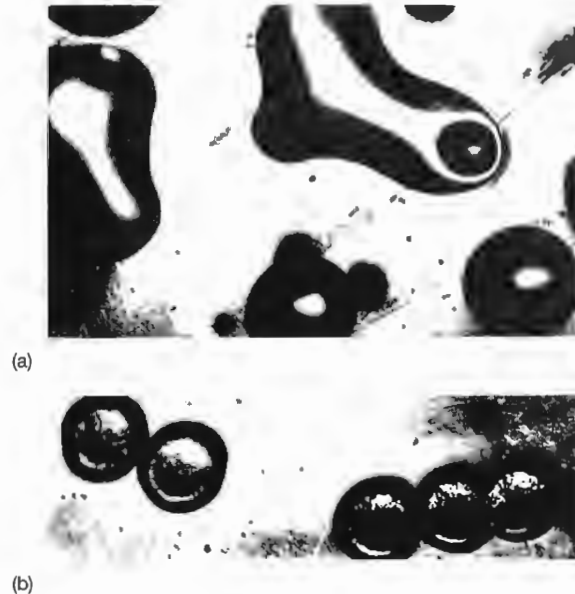


Figure 2.1 Variants of *Streptococcus pneumoniae*: (a) The large, glossy colonies contain smooth (S) virulent bacteria; (b) the small, mottled colonies are composed of rough (R) avirulent bacteria.

(Source: (a, b) Harriet Ephrussi-Taylor.)

virulence during transformation. This meant that the transforming substance in the heat-killed bacteria was probably the gene for virulence itself. The missing piece of the puzzle was the chemical nature of the transforming substance.

DNA: The Transforming Material Oswald Avery, Colin MacLeod, and Maclyn McCarty supplied the missing piece in 1944. They used a transformation test similar to the one that Griffith had introduced, and they took pains to define the chemical nature of the transforming substance from virulent cells. First, they removed the protein from the extract with organic solvents and found that the extract still transformed. Next, they subjected it to digestion with various enzymes. Trypsin and chymotrypsin, which destroy protein, had no effect on transformation. Neither did ribonuclease, which degrades RNA. These experiments ruled out protein or RNA as the transforming material. On the other hand, Avery and his coworkers found that the enzyme deoxyribonuclease (DNase), which breaks down DNA, destroyed the transforming ability of the virulent cell extract. These results suggested that the transforming substance was DNA.

Direct physical-chemical analysis supported the hypothesis that the purified transforming substance was DNA. The analytical tools Avery and his colleagues used were the following:

1. **Ultracentrifugation** They spun the transforming substance in an ultracentrifuge (a very high-speed

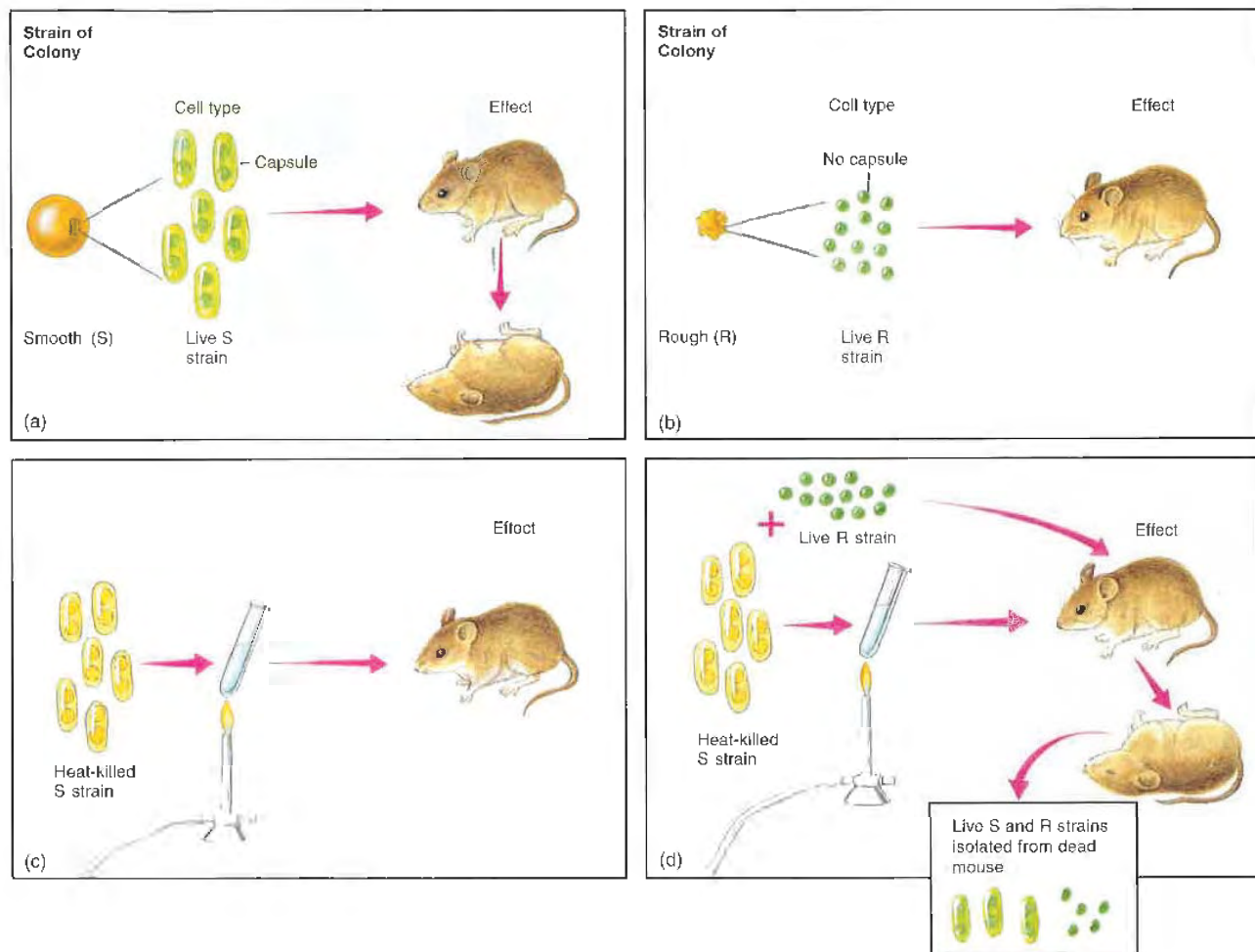


Figure 2.2 Griffith's transformation experiments. (a) Virulent strain S *S. pneumoniae* bacteria kill their host; (b) avirulent strain R bacteria cannot infect successfully, so the mouse survives; (c) strain S bacteria that are heat-killed can no longer infect; (d) a mixture of strain R and heat-killed strain S bacteria kills the mouse. The killed virulent (S) bacteria have transformed the avirulent (R) bacteria to virulent (S).

centrifuge) to estimate its size. The material with transforming activity sedimented rapidly (moved rapidly toward the bottom of the centrifuge tube), suggesting a very high molecular weight, characteristic of DNA.

2. **Electrophoresis** They placed the transforming substance in an electric field to see how rapidly it moved. The transforming activity had a relatively high mobility, also characteristic of DNA because of its high charge-to-mass ratio.
3. **Ultraviolet Absorption Spectrophotometry** They placed a solution of the transforming substance in a spectrophotometer to see what kind of ultraviolet (UV) light it absorbed most strongly. Its absorption spectrum matched that of DNA. That is, the light it

absorbed most strongly had a wavelength of about 260 nanometers (nm), in contrast to protein, which absorbs maximally at 280 nm.

4. **Elementary Chemical Analysis** This yielded an average nitrogen-to-phosphorus ratio of 1.67, about what one would expect for DNA, which is rich in both elements, but vastly lower than the value expected for protein, which is rich in nitrogen but poor in phosphorus. Even a slight protein contamination would have raised the nitrogen-to-phosphorus ratio.

Further Confirmation These findings should have settled the issue of the nature of the gene, but they had little immediate effect. The mistaken notion, from early chemical

analyses, that DNA was a monotonous repeat of a four-nucleotide sequence, such as ACTG-ACTG-ACTG, and so on, persuaded many geneticists that it could not be the genetic material. Furthermore, controversy persisted about possible protein contamination in the transforming material, whether transformation could be accomplished with other genes besides those governing R and S, and even whether bacterial genes were like the genes of higher organisms.

Yet, by 1953, when James Watson and Francis Crick published the double-helical model of DNA structure, most geneticists agreed that genes were made of DNA. What had changed? For one thing, Erwin Chargaff had shown in 1950 that the bases were not really found in equal proportions in DNA, as previous evidence had suggested, and that the base composition of DNA varied from one species to another. In fact, this is exactly what one would expect for genes, which also vary from one species to another. Furthermore, Rollin Hotchkiss had refined and extended Avery's findings. He purified the transforming substance to the point where it contained only 0.02% protein and showed that it could still change the genetic characteristics of bacterial cells. He went on to show that such highly purified DNA could transfer genetic traits other than R and S.

Finally, in 1952, A.D. Hershey and Martha Chase performed another experiment that added to the weight of evidence that genes were made of DNA. This experiment involved a bacteriophage (bacterial virus) called T2 that infects the bacterium *Escherichia coli* (Figure 2.3).

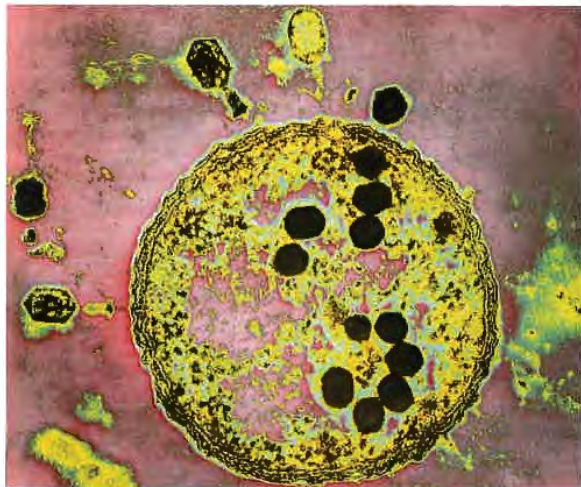


Figure 2.3 A false color transmission electron micrograph of T2 phages infecting an *E. coli* cell. Phage particles at left and top appear ready to inject their DNA into the host cell. Another T2 phage has already infected the cell, however, and progeny phage particles are being assembled. The progeny phage heads are readily discernible as dark polygons inside the host cell. (Source: © Lee Simon/Photo Researchers, Inc.)

(The term *bacteriophage* is usually shortened to *phage*.) During infection, the phage genes enter the host cell and direct the synthesis of new phage particles. The phage is composed of protein and DNA only. The question is this: Do the genes reside in the protein or in the DNA? The Hershey–Chase experiment answered this question by showing that, on infection, most of the DNA entered the bacterium, along with only a little protein. The bulk of the protein stayed on the outside (Figure 2.4). Because DNA was the major component that got into the host cells, it likely contained the genes. Of course, this conclusion was not unequivocal; the small amount of protein that entered along with the DNA could conceivably have carried the genes. But taken together with the work that had gone before, this study helped convince geneticists that DNA, and not protein, is the genetic material.

The Hershey–Chase experiment depended on radioactive labels on the DNA and protein—a different label for each. The labels used were phosphorus-32 (^{32}P) for DNA and sulfur-35 (^{35}S) for protein. These choices make sense, considering that DNA is rich in phosphorus but phage protein has none, and that protein contains sulfur but DNA does not.

Hershey and Chase allowed the labeled phages to attach by their tails to bacteria and inject their genes into their hosts. Then they removed the empty phage coats by mixing vigorously in a blender. Because they knew that the genes must go into the cell, their question was: What went in, the ^{32}P -labeled DNA or the ^{35}S -labeled protein? As we have seen, it was the DNA. In general, then, genes are made of DNA. On the other hand, as we will see later in this chapter, other experiments showed that some viral genes consist of RNA.

SUMMARY Physical-chemical experiments involving bacteria and a bacteriophage showed that their genes are made of DNA.

The Chemical Nature of Polynucleotides

By the mid-1940s, biochemists knew the fundamental chemical structures of DNA and RNA. When they broke DNA into its component parts, they found these constituents to be nitrogenous bases, phosphoric acid, and the sugar deoxyribose (hence the name *deoxyribonucleic acid*). Similarly, RNA yielded bases and phosphoric acid, plus a different sugar, ribose. The four bases found in DNA are adenine (A), cytosine (C), guanine (G), and thymine (T). RNA contains the same bases, except that uracil (U) replaces thymine. The structures of these bases,

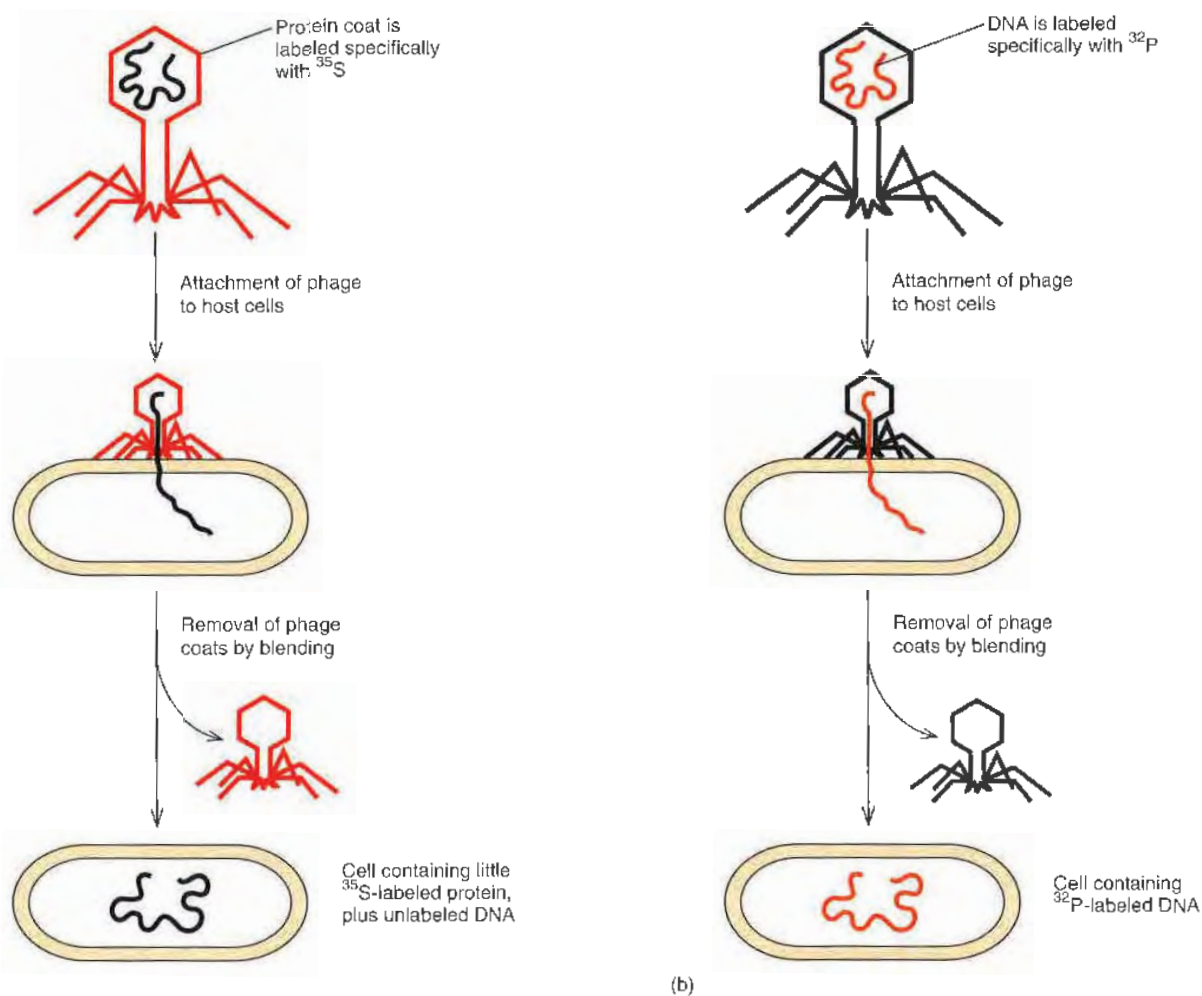


Figure 2.4 The Hershey–Chase experiment. Phage T2 contains genes that allow it to replicate in *E. coli*. Because the phage is composed of DNA and protein only, its genes must be made of one of these substances. To discover which, Hershey and Chase performed a two-part experiment. In the first part (a), they labeled the phage protein with ^{35}S (red), leaving the DNA unlabeled (black). In the second part (b), they labeled the phage DNA with ^{32}P (red),

leaving the protein unlabeled (black). Since the phage genes must enter the cell, the experimenters reasoned that the type of label found in the infected cells would indicate the nature of the genes. Most of the labeled protein remained on the outside and was stripped off the cells by use of a blender (a), whereas most of the labeled DNA entered the infected cells (b). The conclusion was that the genes of this phage are made of DNA.

shown in Figure 2.5, reveal that adenine and guanine are related to the parent molecule, purine. Therefore, we refer to these compounds as **purines**. The other bases resemble pyrimidine, so they are called **pyrimidines**. These structures constitute the alphabet of genetics.

Figure 2.6 depicts the structures of the sugars found in nucleic acids. Notice that they differ in only one place. Where ribose contains a hydroxyl (OH) group in the 2-position, deoxyribose lacks the oxygen and simply has a hydrogen (H), represented by the vertical line. Hence the name *deoxyribose*. The bases and sugars in RNA and DNA are joined together into units called **nucleosides** (Figure 2.7). The names of the nucleosides derive from the corresponding bases:

Base	Nucleoside (RNA)	Deoxynucleoside (DNA)
Adenine	Adenosine	Deoxyadenosine
Guanine	Guanosine	Deoxyguanosine
Cytosine	Cytidine	Deoxycytidine
Uracil	Uridine	Not usually found
Thymine	Not usually found	(Deoxy)thymidine

Because thymine is not usually found in RNA, the “deoxy” designation for its nucleoside is frequently assumed, and the deoxynucleoside is simply called **thymidine**. The numbering of the carbon atoms in the sugars of the nucleosides (see Figure 2.7) is important. Note that the ordinary numbers are used in the bases, so

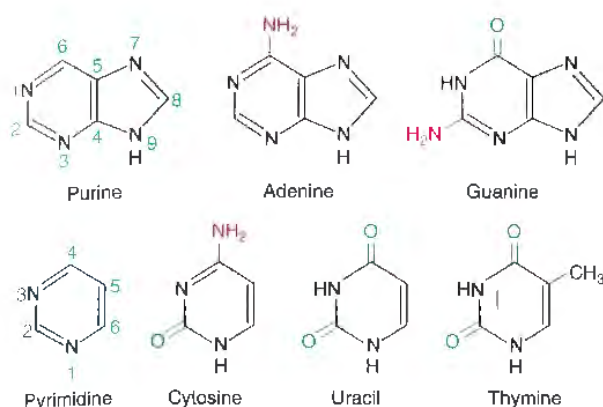


Figure 2.5 The bases of DNA and RNA. The parent bases, purine and pyrimidine, on the left, are not found in DNA and RNA. They are shown for comparison with the other five bases.

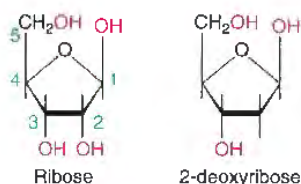


Figure 2.6 The sugars of nucleic acids. Note the OH in the 2'-position of ribose and its absence in deoxyribose.

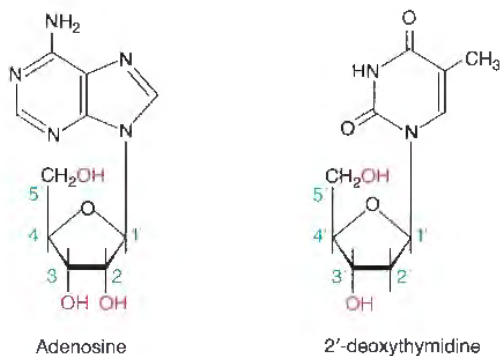


Figure 2.7 Two examples of nucleosides.

the carbons in the sugars are called by primed numbers. Thus, for example, the base is linked to the 1'-position of the sugar, the 2'-position is deoxy in deoxynucleosides, and the sugars are linked together in DNA and RNA through their 3'- and 5'-positions.

The structures in Figure 2.5 were drawn using an organic chemistry shorthand that leaves out certain atoms for simplicity's sake. Figures 2.6 and 2.7 use a slightly different convention, in which a straight line with a free end denotes a C-H bond with a hydrogen atom at the end. Figure 2.8 shows the structures of

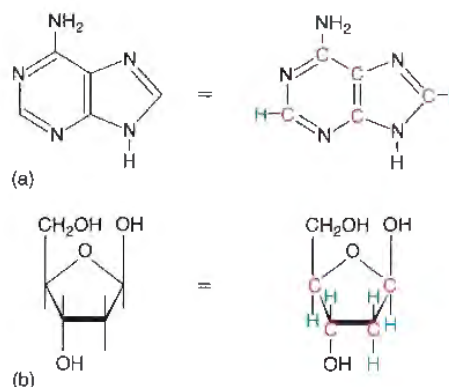


Figure 2.8 The structures of (a) adenine and (b) deoxyribose. Note that the structures on the left do not designate most or all of the carbons and some of the hydrogens. These designations are included in the structures on the right, in red and blue, respectively.

adenine and deoxyribose, first in shorthand, then with every atom included.

The subunits of DNA and RNA are **nucleotides**, which are nucleosides with a phosphate group attached through a phosphoester bond (Figure 2.9). An ester is an organic compound formed from an alcohol (bearing a hydroxyl group) and an acid. In the case of a nucleotide, the alcohol group is the 5'-hydroxyl group of the sugar, and the acid is phosphoric acid, which is why we call the ester a *phosphoester*. Figure 2.9 also shows the structure of one of the four DNA precursors, deoxyadenosine-5'-triphosphate (dATP). When synthesis of DNA takes place, two phosphate groups are removed from dATP, leaving deoxyadenosine-5'-monophosphate (dAMP). The other three nucleotides in DNA (dCMP, dGMP, and dTMP) have analogous structures and names.

We will discuss the synthesis of DNA in detail in Chapters 20 and 21. For now, notice the structure of the bonds that join nucleotides together in DNA and RNA (Figure 2.10). These are called **phosphodiester bonds** because they involve phosphoric acid linked to *two* sugars: one through a sugar 5'-group, the other through a sugar 3'-group. You will notice that the bases have been rotated in this picture, relative to their positions in previous figures. This more closely resembles their geometry in DNA or RNA. Note also that this **trinucleotide**, or string of three nucleotides, has polarity: The top of the molecule bears a free 5'-phosphate group, so it is called the **5'-end**. The bottom, with a free 3'-hydroxyl group, is called the **3'-end**.

Figure 2.11 introduces a shorthand way of representing a nucleotide or a DNA chain. This notation presents the deoxyribose sugar as a vertical line, with the base joined to the 1'-position at the top and the phosphodiester links to neighboring nucleotides through the 3'-(middle) and 5'-(bottom) positions.

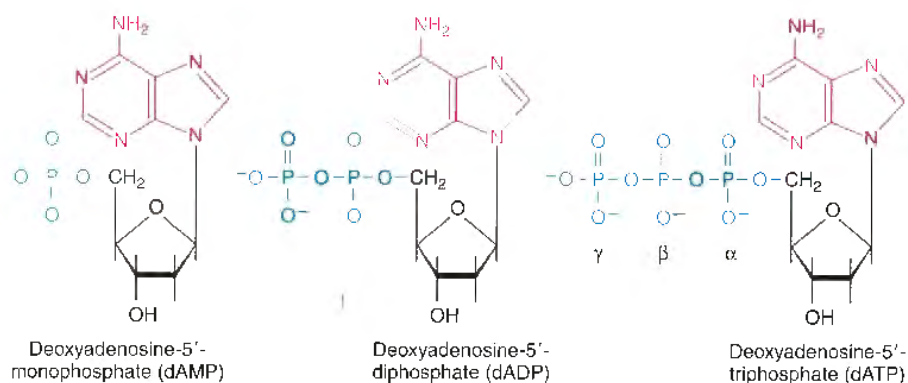


Figure 2.9 Three nucleotides. The 5'-nucleotides of deoxyadenosine are formed by phosphorylating the 5'-hydroxyl group. The addition of one phosphate results in deoxyadenosine-5'-monophosphate (dAMP). One more phosphate yields deoxyadenosine-5'-diphosphate (dADP). Three phosphates (designated α , β , γ) give deoxyadenosine-5'-triphosphate (dATP).

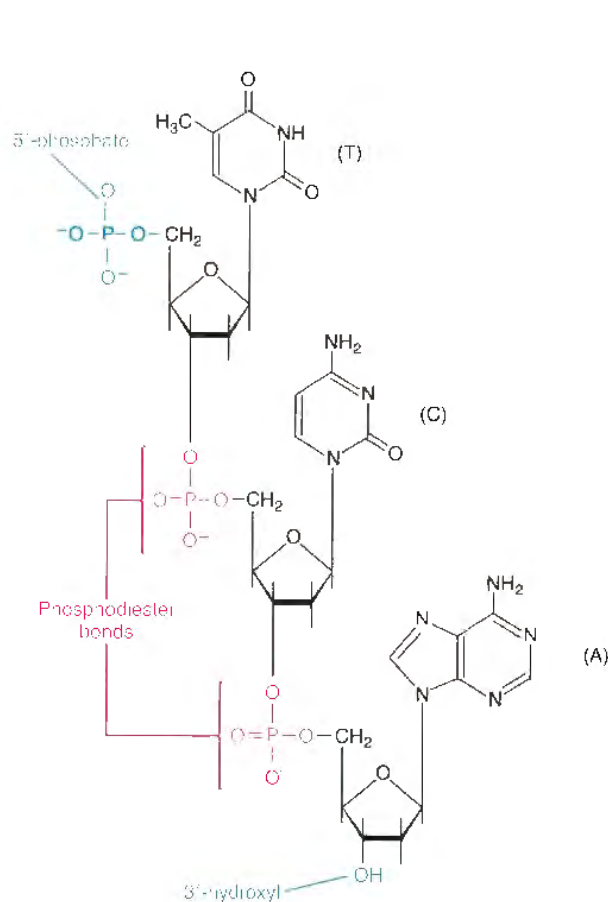


Figure 2.10 A trinucleotide. This little piece of DNA contains only three nucleotides linked together by phosphodiester bonds (red) between the 5'- and 3'-hydroxyl groups of the sugars. The 5'-end of this DNA is at the top, where a free 5'-phosphate group (blue) is located; the 3'-end is at the bottom, where a free 3'-hydroxyl group (also blue) appears. The sequence of this DNA could be read as 5'pdTpdCpdA3'. This would usually be simplified to TCA.

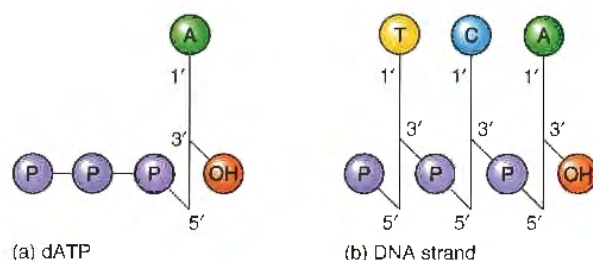


Figure 2.11 Shorthand DNA notation. (a) The nucleotide dATP. This illustration highlights four features of this DNA building block: (1) The deoxyribose sugar is represented by the vertical black line. (2) At the top, attached to the 1'-position of the sugar is the base, adenine (green). (3) In the middle, at the 3'-position of the sugar is a hydroxyl group (OH, orange). (4) At the bottom, attached to the 5'-position of the sugar is a triphosphate group (purple). (b) A short DNA strand. The same trinucleotide (TCA) illustrated in Figure 2.10 is shown here in shorthand. Note the 5'-phosphate and the phosphodiester bonds (purple), and the 3'-hydroxyl group (orange). According to convention, this little piece of DNA is written 5' to 3' left to right.

SUMMARY DNA and RNA are chain-like molecules composed of subunits called nucleotides. The nucleotides contain a base linked to the 1'-position of a sugar (ribose in RNA or deoxyribose in DNA) and a phosphate group. The phosphate joins the sugars in a DNA or RNA chain through their 5'- and 3'-hydroxyl groups by phosphodiester bonds.

2.2 DNA Structure

All the facts about DNA and RNA just mentioned were known by the end of the 1940s. By that time it was also becoming clear that DNA was the genetic material and

that it therefore stood at the very center of the study of life. Yet the three-dimensional structure of DNA was unknown. For these reasons, several researchers dedicated themselves to finding this structure.

Experimental Background

One of the scientists interested in DNA structure was Linus Pauling, a theoretical chemist at the California Institute of Technology. He was already famous for his studies on chemical bonding and for his elucidation of the α -helix, an important feature of protein structure. Indeed, the α -helix, held together by hydrogen bonds, laid the intellectual groundwork for the double-helix model of DNA proposed by Watson and Crick. Another group trying to find the structure of DNA included Maurice Wilkins, Rosalind Franklin, and their colleagues at King's College in London. They were using x-ray diffraction to analyze the three-dimensional structure of DNA. Finally, James Watson and Francis Crick entered the race. Watson, still in his early twenties, but already holding a Ph.D. degree from Indiana University, had come to the Cavendish Laboratories in Cambridge, England, to learn about DNA. There he met Crick, a physicist who at age 35 was retraining as a molecular biologist. Watson and Crick performed no experiments themselves. Their tactic was to use other groups' data to build a DNA model.

Erwin Chargaff was another very important contributor. We have already seen how his 1950 paper helped identify DNA as the genetic material, but the paper contained another piece of information that was even more significant. Chargaff's studies of the base compositions of DNAs from various sources revealed that the content of purines was always roughly equal to the content of pyrimidines. Furthermore, the amounts of adenine and thymine were always roughly equal, as were the amounts of guanine and cytosine. These findings, known as Chargaff's rules, provided a valuable foundation for Watson and Crick's model. Table 2.1 presents Chargaff's data.

You will notice some deviation from the rules due to incomplete recovery of some of the bases, but the overall pattern is clear.

Perhaps the most crucial piece of the puzzle came from an x-ray diffraction picture of DNA taken by Franklin in 1952—a picture that Wilkins shared with James Watson in London on January 30, 1953. The x-ray technique worked as follows: The experimenter made a very concentrated, viscous solution of DNA, then reached in with a needle and pulled out a fiber. This was not a single molecule, but a whole batch of DNA molecules, forced into side-by-side alignment by the pulling action. Given the right relative humidity in the surrounding air, this fiber was enough like a crystal that it diffracted x-rays in an interpretable way. In fact, the x-ray diffraction pattern in Franklin's picture (Figure 2.12) was so simple—a series of spots arranged in an X shape—that it indicated that the DNA structure itself must be very simple. By contrast, a complex, irregular molecule like a protein gives a complex x-ray diffraction pattern with many spots, rather like a surface peppered by a shotgun blast. Because DNA is very large, it can be simple only if it has a regular, repeating structure. And the simplest repeating shape that a long, thin molecule can assume is a corkscrew, or helix.

The Double Helix

Franklin's x-ray work strongly suggested that DNA was a helix. Not only that, it gave some important information about the size and shape of the helix. In particular, the spacing between adjacent bands in an arm of the X is inversely related to the overall repeat distance in the helix, 33.2 angstroms (33.2 Å), and the spacing from the top of the X to the bottom is inversely related to the spacing (3.32 Å) between the repeated elements (base pairs) in the helix. (See Chapter 9 for information on how Bragg's law explains these inverse relationships.) However, even though the Franklin picture told much about

Table 2.1 Composition of DNA in Moles of Base per Mole of Phosphate

	Human						Avian Tubercle Bacilli	Bovine				
	Sperm		Thymus	Liver Carcinoma	Yeast			Thymus			Spleen	
	#1	#2			#1	#2		#1	#2	#3	#1	#2
A:	0.29	0.27	0.28	0.27	0.24	0.30	0.12	0.26	0.28	0.30	0.25	0.26
T:	0.31	0.30	0.28	0.27	0.25	0.29	0.11	0.25	0.24	0.25	0.24	0.24
G:	0.18	0.17	0.19	0.18	0.14	0.18	0.28	0.21	0.24	0.22	0.20	0.21
C:	0.18	0.18	0.16	0.15	0.13	0.15	0.26	0.16	0.18	0.17	0.15	0.17
Recovery:	0.96	0.92	0.91	0.87	0.76	0.92	0.77	0.88	0.94	0.94	0.84	0.88

Source: E. Chargaff "Chemical Specificity of Nucleic Acids and Mechanism of Their Enzymatic Degradation," *Experientia* 6:206, 1950.

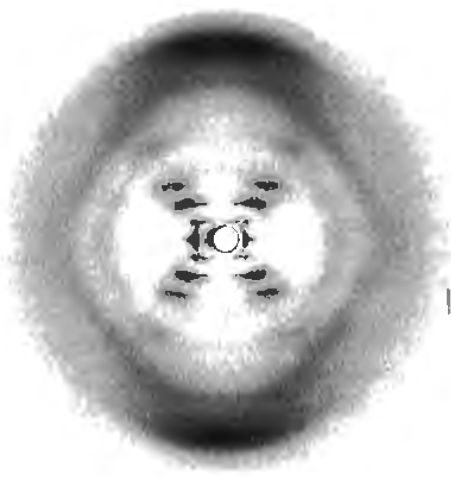


Figure 2.12 Franklin's x-ray picture of DNA. The regularity of this pattern indicated that DNA is a helix. The spacing between the bands at the top and bottom of the X gave the spacing between elements of the helix (base pairs) as 3.32 Å. The spacing between neighboring bands in the pattern gave the overall repeat of the helix (the length of one helical turn) as 33.2 Å. (Source: Courtesy Professor M.H.F. Wilkins, Biophysics Dept., King's College, London.)

DNA, it presented a paradox: DNA was a helix with a regular, repeating structure, but for DNA to serve its genetic function, it must have an *irregular* sequence of bases.

Watson and Crick saw a way to resolve this contradiction and satisfy Chargaff's rules at the same time: DNA must be a **double helix** with its sugar-phosphate backbones on the outside and its bases on the inside. Moreover, the bases must be paired, with a purine in one strand always across from a pyrimidine in the other. This way the helix would be uniform; it would not have bulges where two large purines were paired or constrictions where two small pyrimidines were paired. Watson has joked about the reason he seized on a double helix: "I had decided to build two-chain models. Francis would have to agree. Even though he was a physicist, he knew that important biological objects come in pairs."

But Chargaff's rules went further than this. They decreed that the amounts of adenine and thymine were equal and so were the amounts of guanine and cytosine. This fit very neatly with Watson and Crick's observation that an adenine-thymine base pair held together by hydrogen bonds has almost exactly the same shape as a guanine-cytosine base pair (Figure 2.13). So Watson and Crick postulated that adenine must always pair with thymine, and guanine with cytosine. This way, the double-stranded DNA will be uniform, composed of very similarly shaped base pairs, regardless of the unpredictable sequence of either DNA strand by itself.

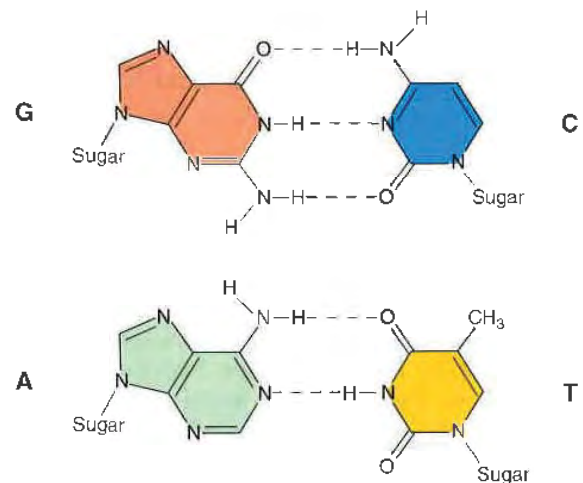


Figure 2.13 The base pairs of DNA. A guanine-cytosine pair (G-C), held together by three hydrogen bonds (dashed lines), has almost exactly the same shape as an adenine-thymine pair (A-T), held together by two hydrogen bonds.

This was their crucial insight, and the key to the structure of DNA.

The double helix, often likened to a twisted ladder, is presented in three ways in Figure 2.14. The curving sides of the ladder represent the sugar-phosphate backbones of the two DNA strands; the rungs are the base pairs. The spacing between base pairs is 3.32 Å, and the overall helix repeat distance is about 33.2 Å, meaning that there are about 10 base pairs (bp) per turn of the helix. (One angstrom [Å] is one ten-billionth of a meter or one-tenth of a nanometer [nm].) The arrows indicate that the two strands are **antiparallel**. If one has 5'→3' polarity from top to bottom, the other must have 3'→5' polarity from top to bottom. In solution, DNA has a structure very similar to the one just described, but the helix contains about 10.4 bp per turn.

Watson and Crick published the outline of their model in the journal *Nature*, back-to-back with papers by Wilkins and Franklin and their coworkers showing the x-ray data. The Watson-Crick paper is a classic of simplicity—only 900 words, barely over a page long. It was published very rapidly, less than a month after it was submitted. Actually, Crick wanted to spell out the biological implications of the model, but Watson was uncomfortable doing that. They compromised on a sentence that is one of the greatest understatements in scientific literature: "It has not escaped our notice that the specific base pairing we have proposed immediately suggests a possible copying mechanism for the genetic material."

As this provocative sentence indicates, Watson and Crick's model does indeed suggest a copying mechanism

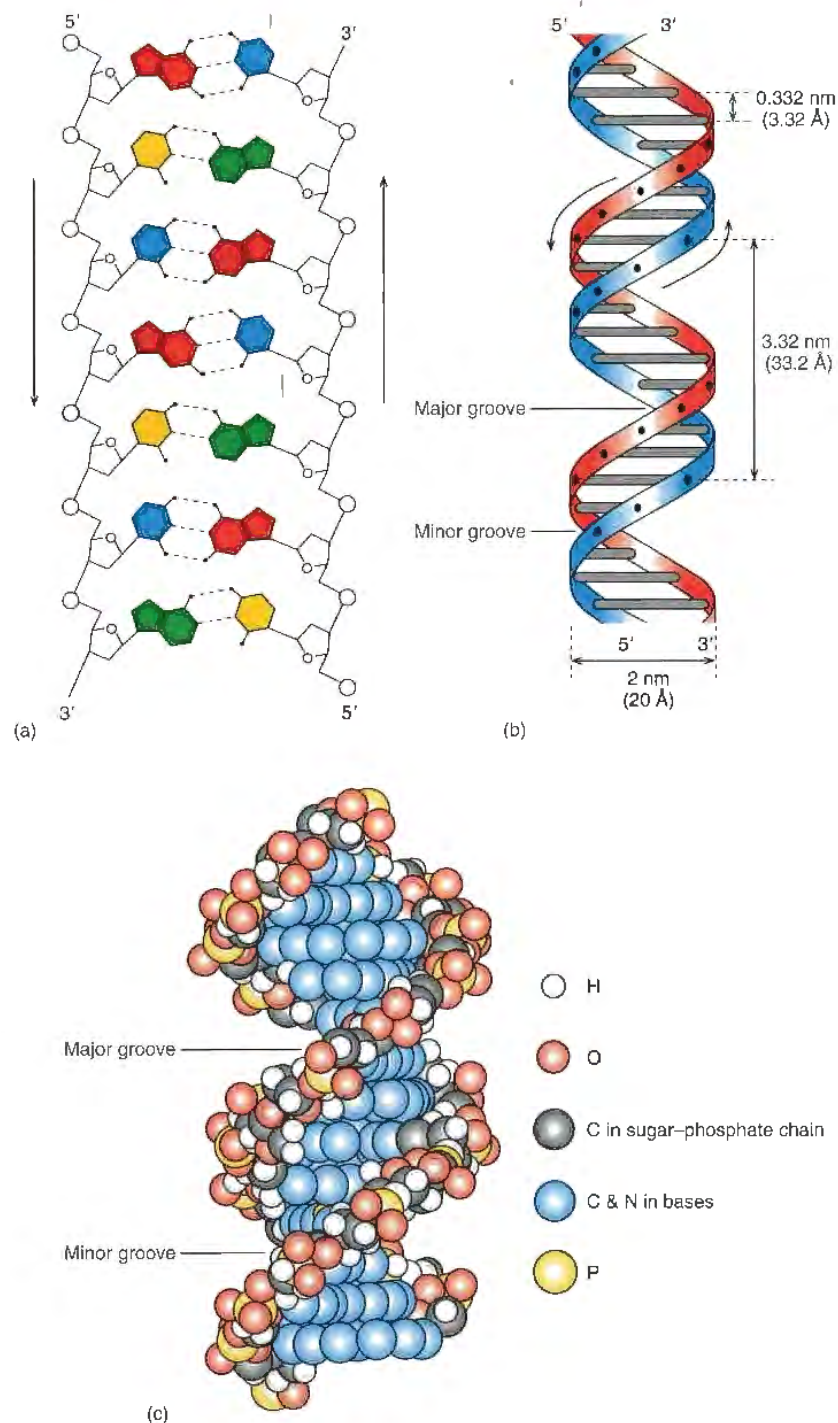


Figure 2.14 Three models of DNA structure. (a) The helix is straightened out to show the base pairing in the middle. Each type of base is represented by a different color, with the sugar-phosphate backbones in black. Note the three hydrogen bonds in the G-C pairs and the two in the A-T pairs. The vertical arrows beside each strand point in the 5'→3' direction and indicate the antiparallel nature of the two DNA strands. The left strand runs 5'→3', top to bottom; the right strand runs 5'→3', bottom to top. The deoxyribose rings (white pentagons with O representing oxygen) also show that the two strands have opposite orientations: The rings in the right strand are

inverted relative to those in the left strand. (b) The DNA double helix is presented as a twisted ladder whose sides represent the sugar-phosphate backbones of the two strands and whose rungs represent base pairs. The curved arrows beside the two strands indicate the 5'→3' orientation of each strand, further illustrating that the two strands are antiparallel. (c) A space-filling model. The sugar-phosphate backbones appear as strings of dark gray, red, light gray, and yellow spheres, whereas the base pairs are rendered as horizontal flat plates composed of blue spheres. Note the major and minor grooves in the helices depicted in parts (b) and (c).

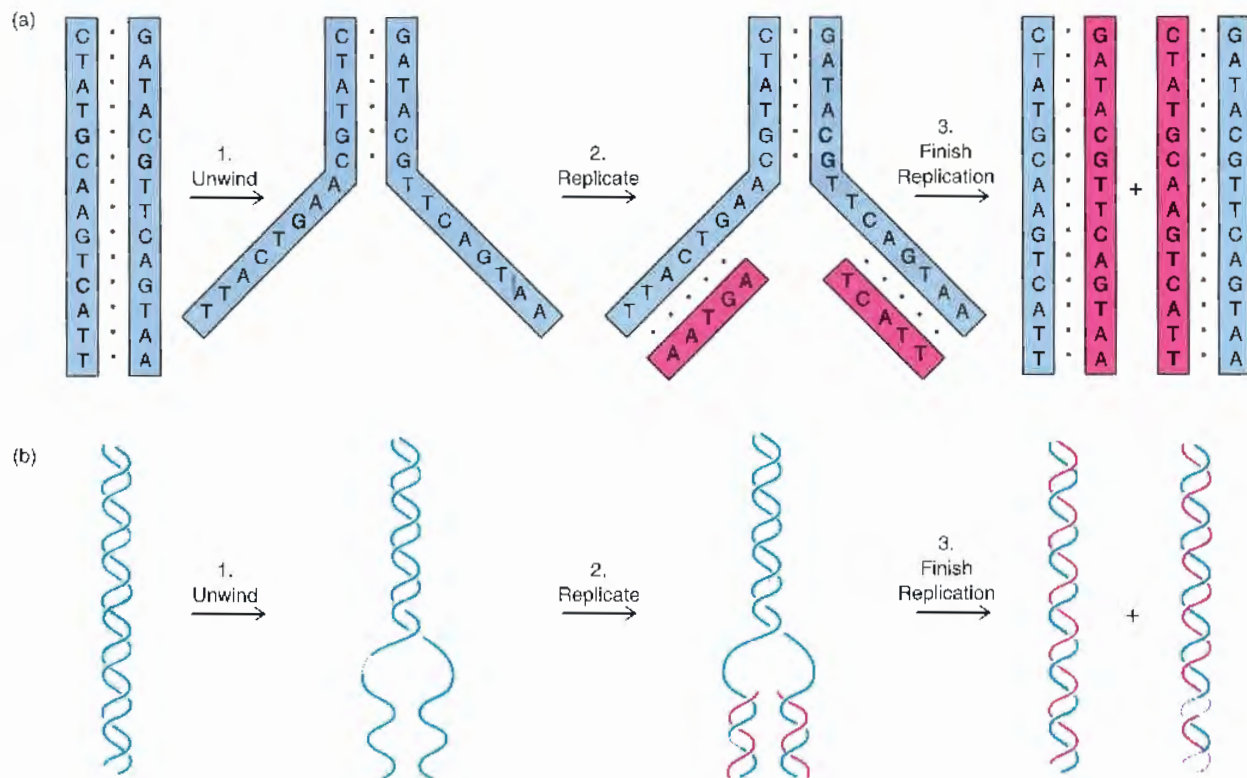


Figure 2.15 Replication of DNA. (a) For simplicity, the two parental DNA strands (blue) are represented as parallel lines. Step 1: During replication these parental strands separate, or unwind. Step 2: New strands (pink) are built with bases complementary to those of the separated parental strands. Step 3: Replication is finished, with the parental strands totally separated and the new strands completed. The end result is two double-stranded DNA duplexes identical to the original. Therefore, each daughter duplex

gets one parental strand (blue) and one new strand (pink). Because only one parental strand is conserved in each of the daughter duplexes, this mechanism of replication is called **semiconservative**. (b) A more realistic portrayal of the same process. Here the strands are shown in a double helix instead of as parallel lines. Notice again that two daughter duplexes are generated, each with one parental strand (blue) and one new strand (pink).

for DNA. Because one strand is the complement of the other, the two strands can be separated, and each can then serve as the template for building a new partner. Figure 2.15 shows schematically how this is accomplished. Notice how this mechanism, known as **semiconservative replication**, ensures that the two daughter DNA duplexes will be exactly the same as the parent, preserving the integrity of the genes as cells divide. In 1958, Matthew Meselson and Franklin Stahl demonstrated that this really is how DNA replicates (Chapter 20).

SUMMARY The DNA molecule is a double helix, with sugar-phosphate backbones on the outside and base pairs on the inside. The bases pair in a specific way: adenine (A) with thymine (T), and guanine (G) with cytosine (C). The replication of

DNA is semiconservative, with each strand serving as the template for building a complementary partner.

2.3 Genes Made of RNA

The genetic system explored by Hershey and Chase was a phage, a bacterial virus. A virus particle by itself is essentially just a package of genes. It has no life of its own, no metabolic activity; it is inert. But when the virus infects a host cell, it seems to come to life. Suddenly the host cell begins making viral proteins. Then the viral genes are replicated and the newly made genes, together with viral coat proteins, assemble into progeny virus particles. Because of

their behavior as inert particles outside, but life-like agents inside their hosts, viruses resist classification. Some scientists refer to them as “living things” or even “organisms.” Others prefer a label that, although more cumbersome, is also more descriptive of a virus’s less-than-living status: infectious agent.

All true organisms and some viruses contain genes made of DNA. But other viruses, including several phages, plant and animal viruses (e.g., HIV, the AIDS virus), have RNA genes. Sometimes viral RNA genes are double-stranded, but usually they are single-stranded.

We have already encountered one famous example of the use of viruses in molecular biology research. We will see many more in subsequent chapters. In fact, without viruses, the field of molecular biology would be immeasurably poorer.

SUMMARY Certain viruses contain genes made of RNA instead of DNA.

2.4 Physical Chemistry of Nucleic Acids

DNA and RNA molecules can assume several different structures. Let us examine these and the behavior of DNA under conditions that encourage the two strands to separate and then come together again.

A Variety of DNA Structures

The structure for DNA proposed by Watson and Crick (see Figure 2.14) represents the sodium salt of DNA in a fiber produced at very high relative humidity (92%). This is called the **B form** of DNA. Although it is probably close to the conformation of most DNA in the cell, it is not the only conformation available to double-stranded nucleic acids. If we reduce the relative humidity surrounding the DNA fiber to 75%, the sodium salt of DNA assumes the **A form** (Figure 2.16a). This differs from the B form (Figure 2.16b) in several respects. Most obviously, the plane of a base pair is no longer roughly perpendicular to the helical axis, but tilts 20 degrees away from horizontal.

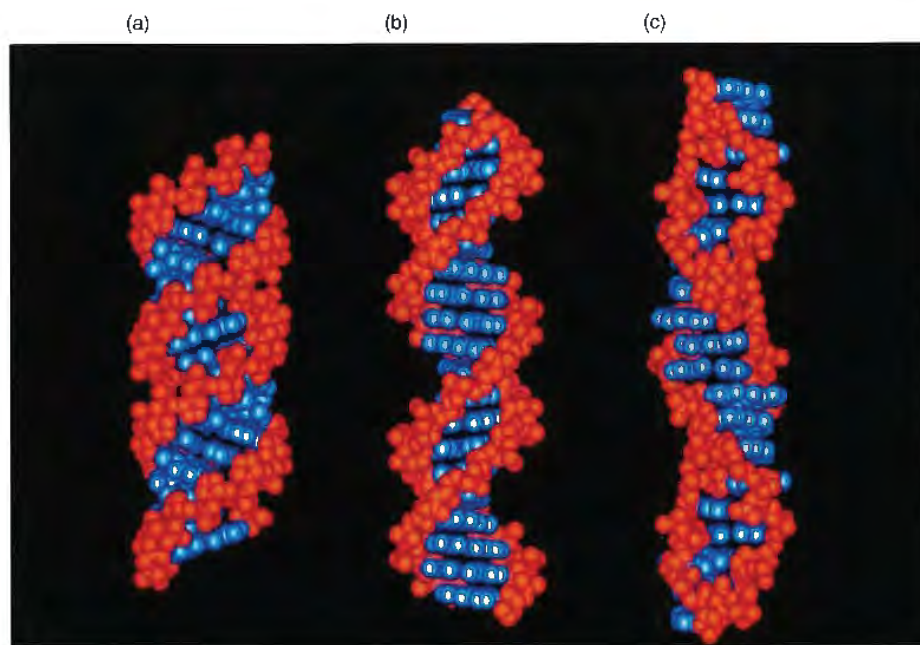


Figure 2.16 Computer graphic models of A-, B-, and Z-DNA. (a) A-DNA. Note the base pairs (blue), whose tilt up from right to left is especially apparent in the major grooves at the top and near the bottom. Note also the right-handed helix traced by the sugar–phosphate backbone (red). (b) B-DNA. Note the familiar right-

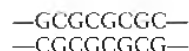
handed helix, with roughly horizontal base pairs. (c) Z-DNA. Note the left-handed helix. All these DNAs are depicted with the same number of base pairs, emphasizing the differences in compactness of the three DNA forms. (Source: Courtesy Fusao Takusagawa.)

Table 2.2 Forms of DNA

Form	Pitch Å	Residues per Turn	Inclination of Base Pair from Horizontal (degrees)
A	24.6	10.7	+19
B	33.2	~10	-1.2
Z	45.6	12	-9

Also, the A helix packs in 10.7 bp per helical turn instead of the 10 found in the B form crystal structure, and each turn occurs in only 24.6 instead of 33.2 Å. This means that the pitch, or distance required for one complete turn of the helix, is only 24.6 instead of 33.2 Å, as in B-DNA. A hybrid polynucleotide containing one DNA and one RNA strand assumes the A form in solution, as does a double-stranded RNA. Table 2.2 presents these helical parameters for A and B form DNA, and for a left-handed Z-form of DNA, discussed in the next paragraph.

Both the A and B form DNA structures are right-handed: The helix turns clockwise away from you whether you look at it from the top or the bottom. Alexander Rich and his colleagues discovered in 1979 that DNA does not always have to be right-handed. They showed that double-stranded DNA containing strands of alternating purines and pyrimidines (e.g., poly[dG-dC] · poly[dG-dC]):



can exist in an extended left-handed helical form. Because of the zigzag look of this DNA's backbone when viewed from the side, it is often called **Z-DNA**. Figure 2.16c presents a picture of Z-DNA. The helical parameters of this structure are given in Table 2.2. Although Rich discovered Z-DNA in studies of model compounds like poly[dG-dC] · poly[dG-dC], this structure seems to be more than just a laboratory curiosity. Evidence suggests that living cells contain a small proportion of Z-DNA. Moreover, Keji Zhao and colleagues discovered in 2001 that activation of at least one gene requires that a regulatory sequence switch to the Z-DNA form.

SUMMARY In the cell, DNA may exist in the common B form, with base pairs horizontal. A small fraction of the DNA may assume an extended left-handed helical form called Z-DNA (at least in eukaryotes). An RNA-DNA hybrid assumes a third helical shape, called the A form, with base pairs tilted away from the horizontal.

Separating the Two Strands of a DNA Double Helix

Although the ratios of G to C and A to T in an organism's DNA are fixed, the GC content (percentage of G + C) can vary considerably from one DNA to another. Table 2.3 lists the GC contents of DNAs from several organisms and viruses. The values range from 22–73%, and these differences are reflected in differences in the physical properties of DNA.

When a DNA solution is heated enough, the noncovalent forces that hold the two strands together weaken and finally break. When this happens, the two strands come apart in a process known as **DNA denaturation**, or **DNA melting**. The temperature at which the DNA strands are half denatured is called the melting temperature, or T_m . Figure 2.17 contains a melting curve for DNA from *Streptococcus pneumoniae*. The amount of strand separation, or melting, is measured by the absorbance of the DNA solution at 260 nm. Nucleic acids absorb light at this

Table 2.3 Relative G + C Contents of Various DNAs

Sources of DNA	Percent (G + C)
<i>Dictyostelium</i> (slime mold)	22
<i>Streptococcus pyogenes</i>	34
Vaccinia virus	36
<i>Bacillus cereus</i>	37
<i>B. megaterium</i>	38
<i>Haemophilus influenzae</i>	39
<i>Saccharomyces cerevisiae</i>	39
Calf thymus	40
Rat liver	40
Bull sperm	41
<i>Streptococcus pneumoniae</i>	42
Wheat germ	43
Chicken liver	43
Mouse spleen	44
Salmon sperm	44
<i>B. subtilis</i>	44
T1 bacteriophage	46
<i>Escherichia coli</i>	51
T7 bacteriophage	51
T3 bacteriophage	53
<i>Neurospora crassa</i>	54
<i>Pseudomonas aeruginosa</i>	68
<i>Sarcina lutea</i>	72
<i>Micrococcus lysodeikticus</i>	72
Herpes simplex virus	72
<i>Mycobacterium phlei</i>	73

Source: From Davidson, *The Biochemistry of the Nucleic Acids*, 5th ed, revised by Adams et al., Lipponcott.

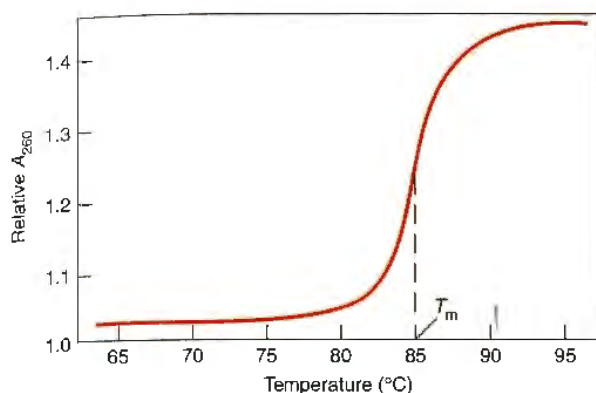


Figure 2.17 Melting curve of *Streptococcus pneumoniae* DNA. The DNA was heated, and its melting was measured by the increase in absorbance at 260 nm. The point at which the melting is half complete is the melting temperature, or T_m . The T_m for this DNA under these conditions is about 85°C. (Adapted from P. Doty, *The Harvey Lectures* 55:121, 1961.)

wavelength because of the electronic structure in their bases, but when two strands of DNA come together, the close proximity of the bases in the two strands quenches some of this absorbance. When the two strands separate, this quenching disappears and the absorbance rises 30–40%. This is called the **hyperchromic shift**. The precipitous rise in the curve shows that the strands hold fast until the temperature approaches the T_m and then rapidly let go.

The GC content of a DNA has a significant effect on its T_m . In fact, as Figure 2.18 shows, the higher a DNA's GC content, the higher its T_m . Why should this be? Recall that one of the forces holding the two strands of DNA together is hydrogen bonding. Remember also that G–C pairs form three hydrogen bonds, whereas A–T pairs have only two. It stands to reason, then, that two strands of DNA rich in G and C will hold to each other more tightly than those of AT-rich DNA. Consider two pairs of embracing centipedes. One pair has 200 legs each, the other 300. Naturally the latter pair will be harder to separate.

Heating is not the only way to denature DNA. Organic solvents such as dimethyl sulfoxide and formamide, or high pH, disrupt the hydrogen bonding between DNA strands and promote denaturation. Lowering the salt concentration of the DNA solution also aids denaturation by removing the ions that shield the negative charges on the two strands from each other. At very low ionic strength, the mutually repulsive forces of these negative charges are strong enough to denature the DNA at a relatively low temperature.

The GC content of a DNA also affects its density. Figure 2.19 shows a direct, linear relationship between GC content and density, as measured by density gradient centrifugation in a CsCl solution (see Chapter 20). Part of the reason for this dependence of density on base composition seems to be real: the larger molar volume

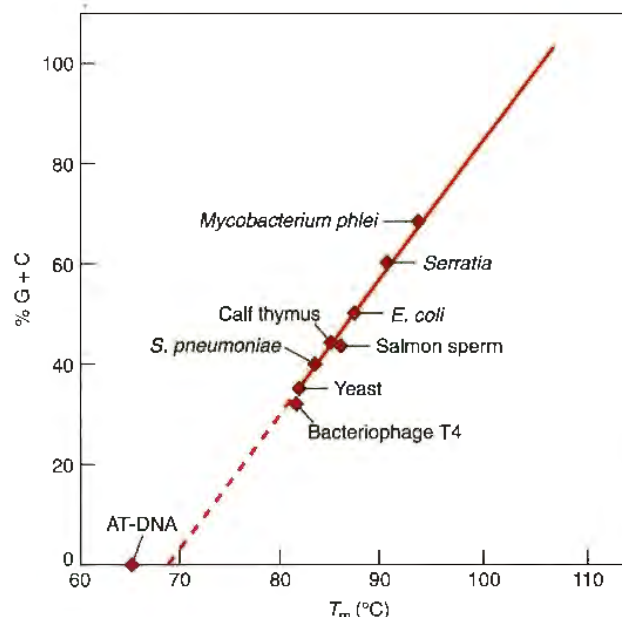


Figure 2.18 Relationship between DNA melting temperature and GC content. AT-DNA refers to synthetic DNAs composed exclusively of A and T (GC content = 0). (Adapted from P. Doty, *The Harvey Lectures* 55:121, 1961.)

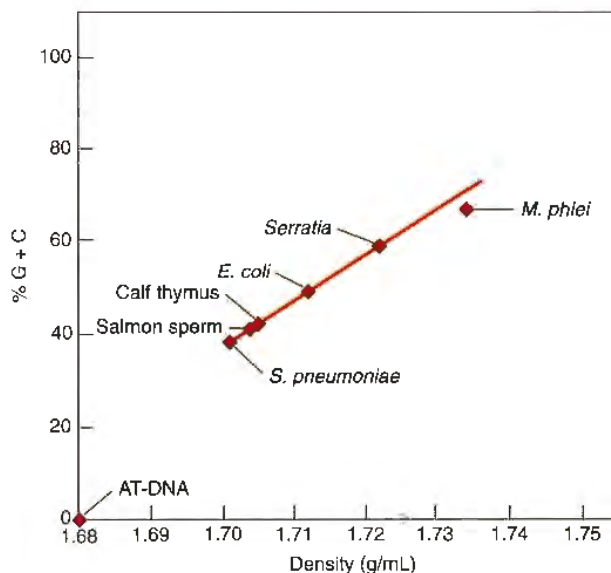


Figure 2.19 Relationship between the GC contents and densities of DNAs from various sources. AT-DNA is a synthetic DNA that is pure A + T; its GC content is therefore zero. (Adapted from P. Doty, *The Harvey Lectures* 55:121, 1961.)

of an A–T base pair, compared with a G–C base pair. But part may be an artifact of the method of measuring density using CsCl: A G–C base pair seems to have a greater tendency to bind to CsCl than does an A–T base pair. This makes its density seem even higher than it actually is.

SUMMARY The GC content of a natural DNA can vary from less than 25% to almost 75%. This can have a strong effect on the physical properties of the DNA, in particular on its melting temperature and density, each of which increases linearly with GC content. The melting temperature (T_m) of a DNA is the temperature at which the two strands are half-dissociated, or denatured. Low ionic strength, high pH, and organic solvents also promote DNA denaturation.

Reuniting the Separated DNA Strands Once the two strands of DNA separate, they can, under the proper conditions, come back together again. This is called **annealing** or **renaturation**. Several factors contribute to renaturation efficiency. Here are three of the most important:

1. **Temperature** The best temperature for renaturation of a DNA is about 25°C below its T_m . This temperature is low enough that it does not promote denaturation, but high enough to allow rapid diffusion of DNA molecules and to weaken the transient bonding between mismatched sequences and short intrastrand base-paired regions. This suggests that rapid cooling following denaturation would prevent renaturation. Indeed, a common procedure to ensure that denatured DNA stays denatured is to plunge the hot DNA solution into ice. This is called *quenching*.
2. **DNA Concentration** The concentration of DNA in the solution is also important. Within reasonable limits, the higher the concentration, the more likely it is that two complementary strands will encounter each other within a given time. In other words, the higher the concentration, the faster the annealing.
3. **Renaturation Time** Obviously, the longer the time allowed for annealing, the more will occur.

SUMMARY Separated DNA strands can be induced to renature, or anneal. Several factors influence annealing; among them are (1) temperature, (2) DNA concentration, and (3) time.

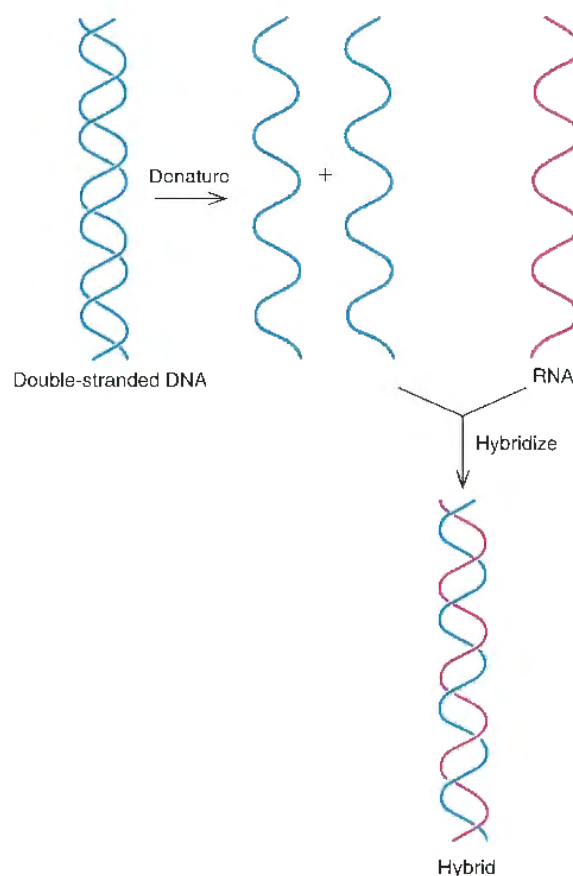


Figure 2.20 Hybridizing DNA and RNA. First, the DNA at upper left is denatured to separate the two DNA strands (blue). Then the DNA strands are mixed with a strand of RNA (red) that is complementary to one of the DNA strands. This hybridization reaction is carried out at a relatively high temperature, which favors RNA–DNA hybridization over DNA–DNA duplex formation. This hybrid has one DNA strand (blue) and one RNA strand (red).

Hybridization of Two Different Polynucleotide Chains

So far, we have dealt only with two separated DNA strands simply getting back together again, but other possibilities exist. Consider, for example, a strand of DNA and a strand of RNA getting together to form a double helix. This could happen if one separated the two strands of a gene, and placed it together with an RNA strand complementary to one of the DNA strands (Figure 2.20). We would not refer to this as annealing; instead, we would call it **hybridization** because we are putting together a **hybrid** of two different nucleic acids. The two chains do not have to be as different as DNA and RNA. If we put together two different strands of DNA having complementary, or nearly complementary, sequences we could still call it hybridization—as long as the strands are of different origin. The difference between the two complementary strands may be very subtle; for example, one may be

radioactive and the other not. As we will see later in this book, hybridization is an extremely valuable technique. In fact, it would be difficult to overestimate the importance of hybridization to molecular biology.

DNAs of Various Sizes and Shapes

Table 2.4 shows the sizes of the haploid genomes of several organisms and viruses. The sizes are expressed three ways: molecular weight, number of base pairs, and length. These are all related, of course. We already know how to convert number of base pairs to length, because about 10.4 bp occur per helical turn, which is 33.2 Å long. To convert base pairs to molecular weight, we simply need to multiply by 660, which is the approximate molecular weight of one average nucleotide pair.

How do we measure these sizes? For small DNAs, this is fairly easy. For example, consider phage PM2 DNA, which contains a double-stranded, circular DNA. How do we know it is circular? The most straightforward way to find out is simply by looking at it. We can do this using an electron microscope, but first we have to treat the DNA so that it stops electrons and will show up in a micrograph just as bones stop x-rays and therefore show up in an x-ray picture. The most common way of doing this is by *shadowing* the DNA with a heavy metal such as platinum.

One places the DNA on an electron microscope grid and bombards it with minute droplets of metal from a shallow angle. This makes the metal pile up beside the DNA like snow behind a fence. One rotates the DNA on the grid so it becomes shadowed all around. Now the metal will stop the electrons in the electron microscope and make the DNA appear as light strings against a darker background. Printing reverses this image to give a picture such as Figure 2.21, which is an electron micrograph of PM2 DNA in two forms: an open circle (lower left) and a supercoil (upper right), in which the DNA coils around itself rather like a twisted rubber band. We can also use pictures like these to measure the length of the DNA. This is more accurate if we include a standard DNA of known length in the same picture.

The size of a DNA can also be estimated by gel electrophoresis, a topic we will discuss in Chapter 5.

SUMMARY Natural DNAs come in sizes ranging from several kilobases to thousands of megabases. The size of a small DNA can be estimated by electron microscopy. This technique can also reveal whether a DNA is circular or linear, and whether it is supercoiled.

Table 2.4 Sizes of Various DNAs

Source	Molecular Weight	Base Pairs (bp)	Length
Viruses and Mitochondria:			
SV40 (mammalian tumor virus)	3.5×10^6	5226	1.7 μm
Bacteriophage ϕ X174 (double-stranded form)	3.2×10^6	5386	1.8 μm
Bacteriophage λ	3.3×10^7	4.85×10^4	13 μm
Bacteriophage T2 or T4	1.3×10^8	2×10^5	50 μm
Human mitochondria	9.5×10^6	16,596	5 μm
Bacteria:			
<i>Haemophilus influenzae</i>	1.2×10^9	1.83×10^6	620 μm
<i>Escherichia coli</i>	3.1×10^9	4.64×10^6	1.6 mm
<i>Salmonella typhimurium</i>	8×10^9	1.1×10^7	3.8 mm
Eukaryotes (content per haploid nucleus):			
<i>Saccharomyces cerevisiae</i> (yeast)	7.9×10^9	1.2×10^7	4.1 mm
<i>Neurospora crassa</i> (pink bread mold)	$\approx 1.9 \times 10^{10}$	$\approx 2.7 \times 10^7$	≈ 9.2 mm
<i>Drosophila melanogaster</i> (fruit fly)	$\approx 1.2 \times 10^{11}$	$\approx 1.8 \times 10^8$	≈ 6.0 cm
<i>Mus musculus</i> (mouse)	$\approx 1.5 \times 10^{12}$	$\approx 2.2 \times 10^9$	≈ 750 cm
<i>Homo sapiens</i> (human)	$\approx 2.3 \times 10^{12}$	$\approx 3.2 \times 10^9$	≈ 1.1 m
<i>Zea mays</i> (corn, or maize)	$\approx 4.4 \times 10^{12}$	$\approx 6.6 \times 10^9$	≈ 2.2 m
<i>Rana pipiens</i> (frog)	$\approx 1.4 \times 10^{13}$	$\approx 2.3 \times 10^{10}$	≈ 7.7 m
<i>Lilium longiflorum</i> (lily)	$\approx 2 \times 10^{14}$	$\approx 3 \times 10^{11}$	≈ 100 m

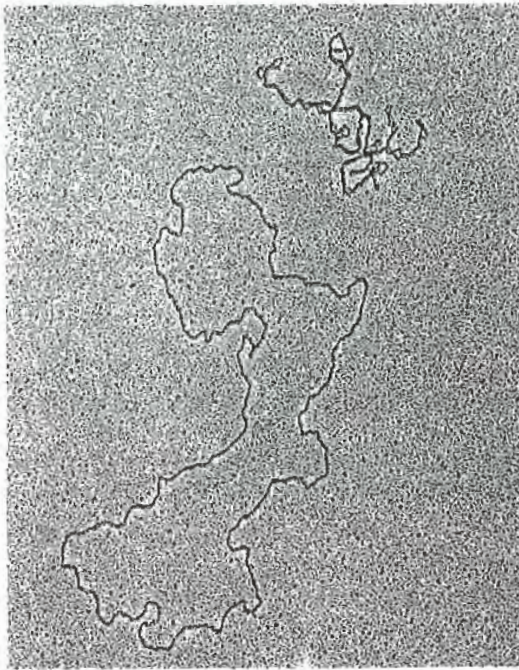


Figure 2.21 Electron micrograph of phage PM2 DNA. The open circular form is shown on the lower left and the supercoiled form is shown at the upper right. (Source: © Jack Griffith)

The Relationship Between DNA Size and Genetic Capacity

How many genes are in a given DNA? It is impossible to tell just from the size of the DNA, because we do not know how much of a given DNA is devoted to genes and how much is space between genes, or even intervening sequences within genes. We can, however, estimate an upper limit on the number of genes a DNA can hold. We start with the assumption that the genes we are discussing here are those that encode proteins. In Chapter 3 and other chapters, we will see that many genes simply encode RNAs, but we are ignoring them here. We also assume that an average protein has a molecular mass of about 40,000 D. How many amino acids does this represent? The molecular masses of amino acids vary, but they average about 110 D. To simplify our calculation, let us assume that the average is 110. That means our average protein contains $40,000/110$, or about 364 amino acids. Because each amino acid requires 3 bp of DNA to code for it, a protein containing 364 amino acids needs a gene of about 1092 bp.

Consider a few of the DNAs listed in Table 2.4. The *E. coli* chromosome contains 4.6×10^6 bp, so it could encode about 4200 average proteins. Phage λ , which infects *E. coli*, has only 4.85×10^4 bp, so it can code for only about 44 proteins. One of the smallest double-stranded DNAs on the list, belonging to the phage ϕ X174, has a mere 5375 bp. In principle, that is only enough to code

for about five proteins, but the phage squeezes in some extra information by overlapping its genes.

DNA Content and the C-Value Paradox You would probably predict that complex organisms such as vertebrates need more genes than simple organisms like yeast. Therefore, they should have higher C-values, or DNA content per haploid cell. In general, your prediction would be right; mouse and human haploid cells contain more than 100 times more DNA than yeast haploid cells. Furthermore, yeast cells have about five times more DNA than *E. coli* cells, which are even simpler. However, this correspondence between an organism's physical complexity and the DNA content of its cells is not perfect. Consider, for example, the frog. Intuitively, you would not suspect that an amphibian would have a higher C-value than a human, yet the frog has seven times more DNA per cell. Even more dramatic is the fact that the lily has 100 times more DNA per cell than a human.

This perplexing situation is called the C-value paradox. It becomes even more difficult to explain when we look at organisms within a group. For example, some amphibian species have C-values 100 times higher than those of others, and the C-values of flowering plants vary even more widely. Does this mean that one kind of higher plant has 100 times more genes than another? That is simply unbelievable. It would raise questions about what all those extra genes are good for and why we do not notice tremendous differences in physical complexity among these organisms. The more plausible explanation of the C-value paradox is that organisms with extraordinarily high C-values simply have a great deal of extra, noncoding DNA. The function, if any, of this extra DNA is still mysterious.

In fact, even mammals have much more DNA than they need for genes. Applying our simple rule (dividing the number of base pairs by 1090) to the human genome yields an estimate of about 3 million for the maximum number of genes, which is far too high.

In fact, the finished version of the human genome suggests that there are only about 20–25,000 genes. This means that human cells contain more than 100 times more DNA than they apparently need. Much of this extra DNA is found in intervening sequences within eukaryotic genes (Chapter 14). The rest is in noncoding regions outside of genes.

SUMMARY There is a rough correlation between the DNA content and the number of genes in a cell or virus. However, this correlation breaks down in several cases of closely related organisms where the DNA content per haploid cell (C-value) varies widely. This C-value paradox is probably explained, not by extra genes, but by extra noncoding DNA in some organisms.

SUMMARY

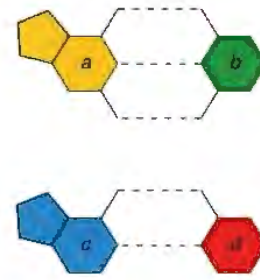
Genes of all true organisms are made of DNA; certain viruses have genes made of RNA. DNA and RNA are chain-like molecules composed of subunits called nucleotides. DNA has a double-helical structure with sugar-phosphate backbones on the outside and base pairs on the inside. The bases pair in a specific way: adenine (A) with thymine (T) and guanine (G) with cytosine (C). When DNA replicates, the parental strands separate; each then serves as the template for making a new, complementary strand.

The G + C content of a natural DNA can vary from 22–73%, and this can have a strong effect on the physical properties of DNA, particularly its melting temperature. The melting temperature (T_m) of a DNA is the temperature at which the two strands are half-dissociated, or denatured. Separated DNA strands can be induced to renature, or anneal. Complementary strands of polynucleotides (either RNA or DNA) from different sources can form a double helix in a process called hybridization. Natural DNAs vary widely in length. The size of a small DNA can be estimated by electron microscopy.

A rough correlation occurs between the DNA content and the number of genes in a cell or virus. However, this correlation does not hold in several cases of closely related organisms in which the DNA content per haploid cell (C-value) varies widely. This C-value paradox is probably explained by extra noncoding DNA in some organisms.

REVIEW QUESTIONS

1. Compare and contrast the experimental approaches used by Avery and colleagues, and by Hershey and Chase, to demonstrate that DNA is the genetic material.
2. Draw the general structure of a deoxynucleoside monophosphate. Show the sugar structure in detail and indicate the positions of attachment of the base and the phosphate. Also indicate the deoxy position.
3. Draw the structure of a phosphodiester bond linking two nucleotides. Show enough of the two sugars that the sugar positions involved in the phosphodiester bond are clear.
4. Which DNA purine forms three H bonds with its partner in the other DNA strand? Which forms two H bonds? Which DNA pyrimidine forms three H bonds with its partner? Which forms two H bonds?
5. The following drawings are the outlines of two DNA base pairs, with the bases identified as *a*, *b*, *c*, and *d*. What are the real identities of these bases?



6. Draw a typical DNA melting curve. Label the axes and point out the melting temperature.
7. Use a graph to illustrate the relationship between the GC content of a DNA and its melting temperature. What is the explanation for this relationship?
8. Use a drawing to illustrate the principle of nucleic acid hybridization.

ANALYTICAL QUESTIONS

1. The double-stranded DNA genome of human herpes simplex virus 1 has a molecular mass of about 1.0×10^5 kD. (a) How many base pairs does this virus contain? (b) How many full double-helical turns does this DNA contain? (c) How long is this DNA in microns?
2. How many proteins of average size could be encoded in a virus with a DNA genome having 12,000 bp, assuming no overlap of genes?

SUGGESTED READINGS

- Adams, R.I.P., R.H. Burdon, A.M. Campbell, and R.M.S. Smellie, eds. 1976. *Davidson's The Biochemistry of the Nucleic Acids*, 8th ed. The structure of DNA, chapter 5. New York: Academic Press.
- Avery, O.T., C.M. McLeod, and M. McCarty. 1944. Studies on the chemical nature of the substance-inducing transformation of pneumococcal types. *Journal of Experimental Medicine* 79:137–58.
- Chargaff, E. 1950. Chemical specificity of the nucleic acids and their enzymatic degradation. *Experientia* 6:201–9.
- Dickerson, R.E. 1983. The DNA helix and how it reads. *Scientific American* 249 (December): 94–111.
- Hershey, A.D., and M. Chase. 1952. Independent functions of viral protein and nucleic acid in growth of bacteriophage. *Journal of General Physiology* 36:39–56.
- Watson, J.D., and F.H.C. Crick. 1953. Genetical implications of the structure of deoxyribonucleic acid. *Nature* 171:964–67.
- Watson, J.D., and F.H.C. Crick. 1953. Molecular structure of the nucleic acids: A structure for deoxyribose nucleic acid. *Nature* 171:737–38.