

MOLECULAR BIOLOGY OF
THE CELL

fourth edition

Bruce Alberts

Alexander Johnson

Julian Lewis

Martin Raff

Keith Roberts

Peter Walter

GS Garland Science
Taylor & Francis Group

YANKEE 32100 0336
S030 00000 00000

Garland
Vice President: Denise Schanck
Managing Editor: Sarah Gibbs
Senior Editorial Assistant: Kirsten Jenner
Managing Production Editor: Emma Hunt
Proofreader and Layout: Emma Hunt
Production Assistant: Angela Bennett
Text Editors: Marjorie Singer Anderson and Betsy Dilernia
Copy Editor: Bruce Goatly
Word Processors: Fran Dependahl, Misty Landers and Carol Winter
Designer: Blink Studio, London
Illustrator: Nigel Orme
Indexer: Janine Ross and Sherry Granum
Manufacturing: Nigel Eyre and Marion Morrow

Cell Biology Interactive
Artistic and Scientific Direction: Peter Walter
Narrated by: Julie Theriot
Production, Design, and Development: Mike Morales

Oversize
QH581.2 .M718 2002
c.2
Molecular biology of the
cell

Bruce Alberts received his Ph.D. from Harvard University and is President of the National Academy of Sciences and Professor of Biochemistry and Biophysics at the University of California, San Francisco. **Alexander Johnson** received his Ph.D. from Harvard University and is a Professor of Microbiology and Immunology at the University of California, San Francisco. **Julian Lewis** received his D.Phil. from the University of Oxford and is a Principal Scientist at the Imperial Cancer Research Fund, London.

Martin Raff received his M.D. from McGill University and is at the Medical Research Council Laboratory for Molecular Cell Biology and Cell Biology Unit and in the Biology Department at University College London. **Keith Roberts** received his Ph.D. from the University of Cambridge and is Associate Research Director at the John Innes Centre, Norwich. **Peter Walter** received his Ph.D. from The Rockefeller University in New York and is Professor and Chairman of the Department of Biochemistry and Biophysics at the University of California, San Francisco, and an Investigator of the Howard Hughes Medical Institute.

© 2002 by Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter.
© 1983, 1989, 1994 by Bruce Alberts, Dennis Bray, Julian Lewis, Martin Raff, Keith Roberts, and James D. Watson.

All rights reserved. No part of this book covered by the copyright hereon may be reproduced or used in any format in any form or by any means—graphic, electronic, or mechanical, including photocopying, recording, taping, or information storage and retrieval systems—without permission of the publisher.

Library of Congress Cataloging-in-Publication Data
Molecular biology of the cell / Bruce Alberts ... [et al.].-- 4th ed.
p. cm
Includes bibliographical references and index.
ISBN 0-8153-3218-1 (hardbound) -- ISBN 0-8153-4072-9 (pbk.)
1. Cytology. 2. Molecular biology. I. Alberts, Bruce.
[DNLM: 1. Cells. 2. Molecular Biology.]

QH581.2 .M64 2002

571.6--dc21

2001054471 CIP

Published by Garland Science, a member of the Taylor & Francis Group, 29 West 35th Street, New York, NY 10001-2299

Printed in the United States of America

15 14 13 12 11 10 9 8 7 6 5 4 3 2 1

Preface

"There is a paradox in the growth of scientific knowledge. As information accumulates in ever more intimidating quantities, disconnected facts and impenetrable mysteries give way to rational explanations, and simplicity emerges from chaos."

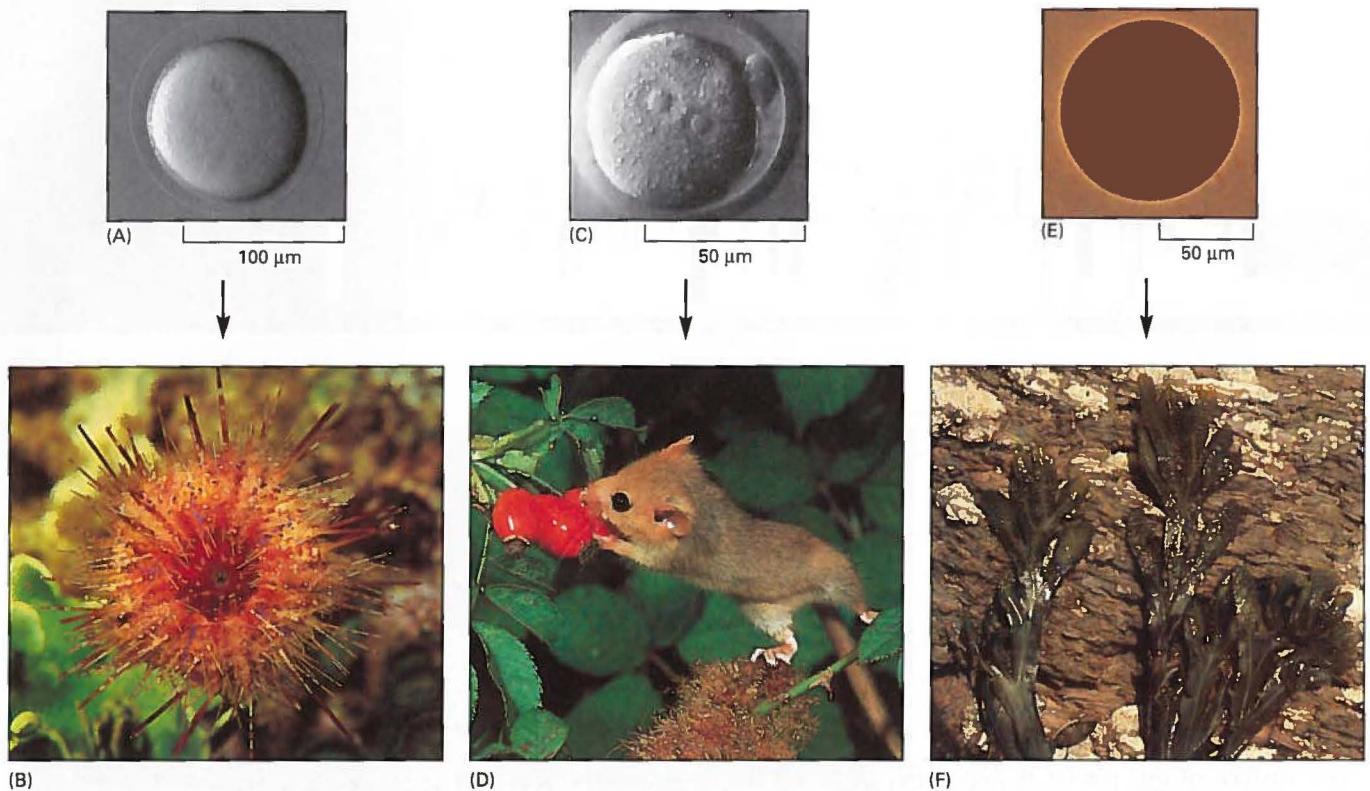
Thus began the preface of our first edition, written 18 years ago. Much of what we wrote in that preface holds for the present edition too. Our goals have not changed: we want to make cell biology comprehensible. We aim, as before, to give readers a perspective both on what is known, and on what is unknown. We have written the book for a wide range of students, but it should also prove useful for scientists wishing to follow progress outside their own specialized fields. As in past editions, each chapter is a joint composition of multiple authors and has been reviewed by a number of experts. This helps to explain why this edition, like the first, has been "a long time in gestation—three times longer than an elephant, five times longer than a whale."

Although information in the life sciences is expanding rapidly, human brain capacity is not. This discrepancy creates an increasing challenge for textbook writers, teachers, and especially students. We have been forced to think harder than ever in deciding what facts and concepts are essential. We have had to summarize—omitting much detail that, in overabundance, can prevent understanding. But specific examples are still needed to bring the subject to life. Thus, the most difficult part of the revision has been deciding what to leave out, while the most rewarding has been the opportunity that new discoveries provide to strengthen the conceptual framework.

What is new in the 4th edition?

Genomics has given us a new perspective that has demanded a complete recasting and expansion of the material on molecular genetics (Chapter 1 and Chapters 4 through 8). These six chapters can now be used as a stand-alone text in molecular biology. As before, the book ends with a major section on Cells in Their Social Context, but we have added a new chapter on Pathogens, Infection, and Innate Immunity. This reflects the remarkable advances in the understanding of the cell biology of infection, as well as the renewed awareness that infectious disease remains one of the greatest unconquered dangers in our world.

As our book makes clear, the complete sequencing of the genomes of hundreds of organisms, from bacteria to humans, has revolutionized our understanding of living things and the relationships between them. At last we can see what is there: the set of genes and proteins is finite, and we can list them. But we also recognize that these components are combined for use in marvelously subtle and complex ways, even in the simplest of organisms. Therefore, the traditional explanatory cartoons that we show on nearly every page of the book generally represent only the primitive first step toward an explanation. These drawings cannot capture the enormous complexity of the networks of protein-protein interactions that are responsible for most intracellular processes, whose understanding will require new and more quantitative forms of analysis. Thus, we are no longer as confident as we were 18 years ago that simplicity will eventually



its organization; but the free energy drives a hugely complex system of chemical processes that is specified by the hereditary information.

Most living organisms are single cells; others, such as ourselves, are vast multicellular cities in which groups of cells perform specialized functions and are linked by intricate systems of communication. But in all cases, whether we discuss the solitary bacterium or the aggregate of more than 10^{13} cells that form a human body, the whole organism has been generated by cell divisions from a single cell. The single cell, therefore, is the vehicle for the hereditary information that defines the species (Figure 1-1). And specified by this information, the cell includes the machinery to gather raw materials from the environment, and to construct out of them a new cell in its own image, complete with a new copy of the hereditary information. Nothing less than a cell has this capability.

All Cells Store Their Hereditary Information in the Same Linear Chemical Code (DNA)

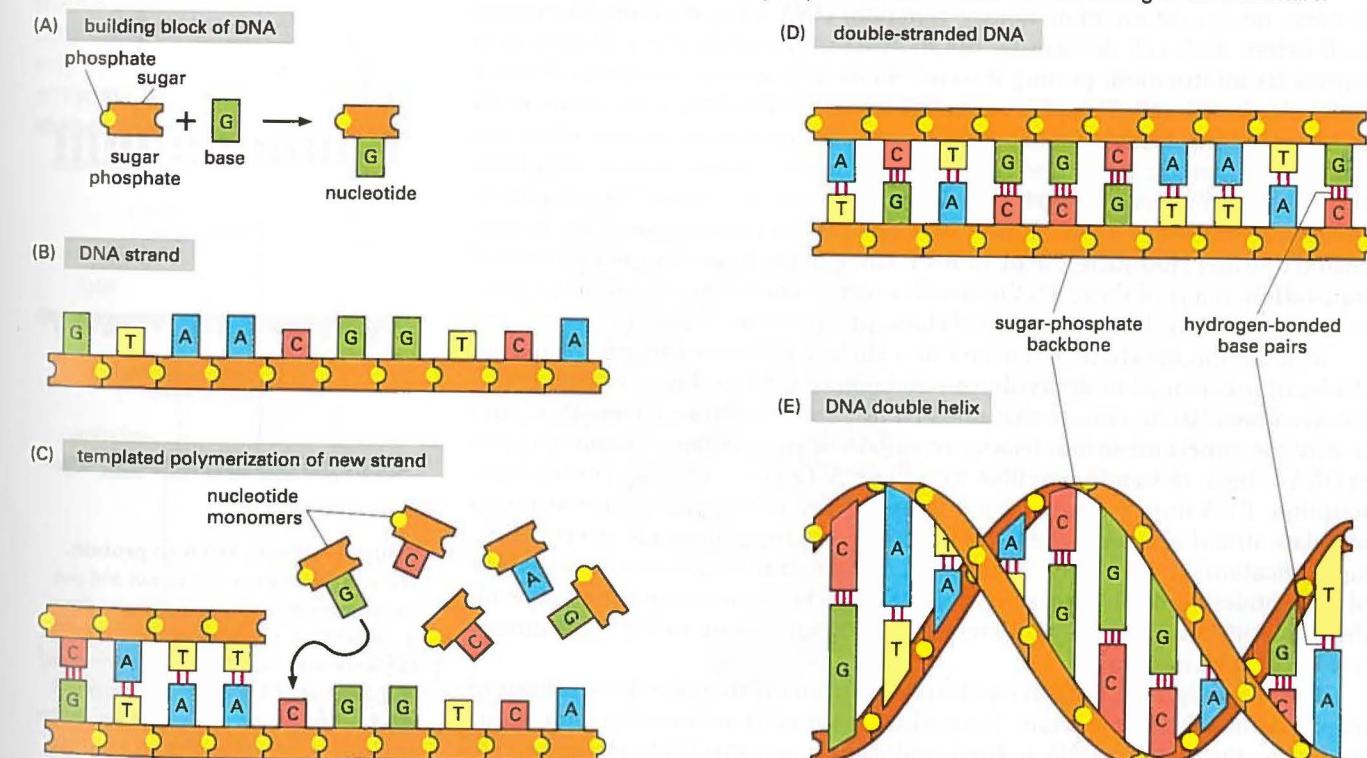
Computers have made us familiar with the concept of information as a measurable quantity—a million bytes (corresponding to about 200 pages of text) on a floppy disk, 600 million on a CD-ROM, and so on. They have also made us uncomfortably aware that the same information can be recorded in many different physical forms. A document that is written on one type of computer may be unreadable on another. As the computer world has evolved, the discs and tapes that we used 10 years ago for our electronic archives have become unreadable on present-day machines. Living cells, like computers, deal in information, and it is estimated that they have been evolving and diversifying for over 3.5 billion years. It is scarcely to be expected that they should all store their information in the same form, or that the archives of one type of cell should be readable by the information-handling machinery of another. And yet it is so. All living cells on Earth, without any known exception, store their hereditary information in the form of double-stranded molecules of DNA—long unbranched paired polymer chains, formed always of the same four types of monomers—A, T, C, G. These monomers are strung together in a long linear sequence that encodes the genetic information, just as the sequence of 1s and 0s encodes the information in a computer file. We can take a piece of DNA from a human cell and insert it

into a bacterium, or a piece of bacterial DNA and insert it into a human cell, and the information will be successfully read, interpreted, and copied. Using chemical methods, scientists can read out the complete sequence of monomers in any DNA molecule—extending for millions of nucleotides—and thereby decipher the hereditary information that each organism contains.

All Cells Replicate Their Hereditary Information by Templated Polymerization

To understand the mechanisms that make life possible, one must understand the structure of the double-stranded DNA molecule. Each monomer in a single DNA strand—that is, each **nucleotide**—consists of two parts: a sugar (deoxyribose) with a phosphate group attached to it, and a *base*, which may be either adenine (A), guanine (G), cytosine (C) or thymine (T) (Figure 1-2). Each sugar is linked to the next via the phosphate group, creating a polymer chain composed of a repetitive sugar-phosphate backbone with a series of bases protruding from it. The DNA polymer is extended by adding monomers at one end. For a single isolated strand, these can, in principle, be added in any order, because each one links to the next in the same way, through the part of the molecule that is the same for all of them. In the living cell, however, there is a constraint: DNA is not

Figure 1-2 DNA and its building blocks. (A) DNA is made from simple subunits, called nucleotides, each consisting of a sugar-phosphate molecule with a nitrogen-containing sidegroup, or base, attached to it. The bases are of four types (adenine, guanine, cytosine, and thymine), corresponding to four distinct nucleotides, labeled A, G, C, and T. (B) A single strand of DNA consists of nucleotides joined together by sugar-phosphate linkages. Note that the individual sugar-phosphate units are asymmetric, giving the backbone of the strand a definite directionality, or polarity. This directionality guides the molecular processes by which the information in DNA is interpreted and copied in cells: the information is always “read” in a consistent order, just as written English text is read from left to right. (C) Through templated polymerization, the sequence of nucleotides in an existing DNA strand controls the sequence in which nucleotides are joined together in a new DNA strand; T in one strand pairs with A in the other, and G in one strand with C in the other. The new strand has a nucleotide sequence complementary to that of the old strand, and a backbone with opposite directionality: corresponding to the GTAA... of the original strand, it has ...TTAC. (D) A normal DNA molecule consists of two such complementary strands. The nucleotides within each strand are linked by strong (covalent) chemical bonds; the complementary nucleotides on opposite strands are held together more weakly, by hydrogen bonds. (E) The two strands twist around each other to form a double helix—a robust structure that can accommodate any sequence of nucleotides without altering its basic structure.



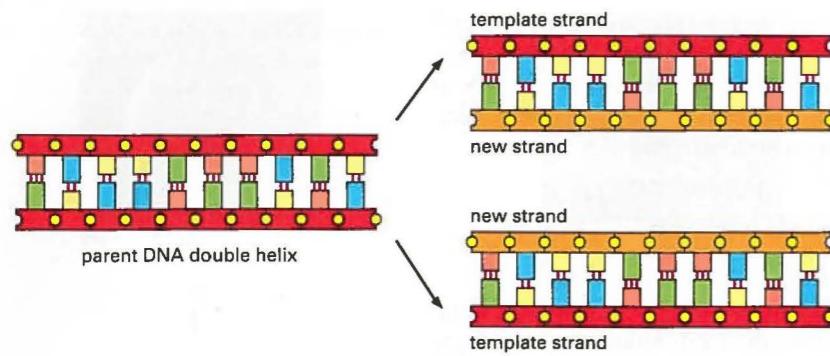


Figure 1–3 The duplication of genetic information by DNA replication. In this process, the two strands of a DNA double helix are pulled apart, and each serves as a template for synthesis of a new complementary strand.

synthesized as a free strand in isolation, but on a template formed by a preexisting DNA strand. The bases protruding from the existing strand bind to bases of the strand being synthesized, according to a strict rule defined by the complementary structures of the bases: A binds to T, and C binds to G. This base-pairing holds fresh monomers in place and thereby controls the selection of which one of the four monomers shall be added to the growing strand next. In this way, a double-stranded structure is created, consisting of two exactly complementary sequences of As, Cs, Ts, and Gs. The two strands twist around each other, forming a double helix (Figure 1–2E).

The bonds between the base pairs are weak compared with the sugar-phosphate links, and this allows the two DNA strands to be pulled apart without breakage of their backbones. Each strand then can serve as a template, in the way just described, for the synthesis of a fresh DNA strand complementary to itself—a fresh copy, that is, of the hereditary information (Figure 1–3). In different types of cells, this process of **DNA replication** occurs at different rates, with different controls to start it or stop it, and different auxiliary molecules to help it along. But the basics are universal: DNA is the information store, and *templated polymerization* is the way in which this information is copied throughout the living world.

All Cells Transcribe Portions of Their Hereditary Information into the Same Intermediary Form (RNA)

To carry out its information-storage function, DNA must do more than copy itself before each cell division by the mechanism just described. It must also express its information, putting it to use so as to guide the synthesis of other molecules in the cell. This also occurs by a mechanism that is the same in all living organisms, leading first and foremost to the production of two other key classes of polymers: RNAs and proteins. The process begins with a templated polymerization called **transcription**, in which segments of the DNA sequence are used as templates to guide the synthesis of shorter molecules of the closely related polymer **ribonucleic acid**, or RNA. Later, in the more complex process of **translation**, many of these RNA molecules serve to direct the synthesis of polymers of a radically different chemical class—the *proteins* (Figure 1–4).

In RNA, the backbone is formed of a slightly different sugar from that of DNA—ribose instead of deoxyribose—and one of the four bases is slightly different—uracil (U) in place of thymine (T); but the other three bases—A, C, and G—are the same, and all four bases pair with their complementary counterparts in DNA—the A, U, C, and G of RNA with the T, A, G, and C of DNA. During transcription, RNA monomers are lined up and selected for polymerization on a template strand of DNA in the same way that DNA monomers are selected during replication. The outcome is therefore a polymer molecule whose sequence of nucleotides faithfully represents a part of the cell's genetic information, even though written in a slightly different alphabet, consisting of RNA monomers instead of DNA monomers.

The same segment of DNA can be used repeatedly to guide the synthesis of many identical RNA transcripts. Thus, whereas the cell's archive of genetic information in the form of DNA is fixed and sacrosanct, the RNA transcripts are

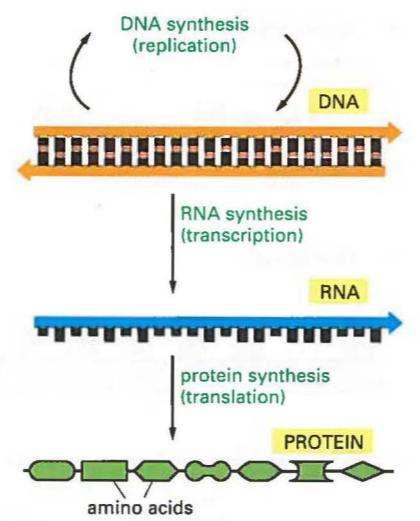
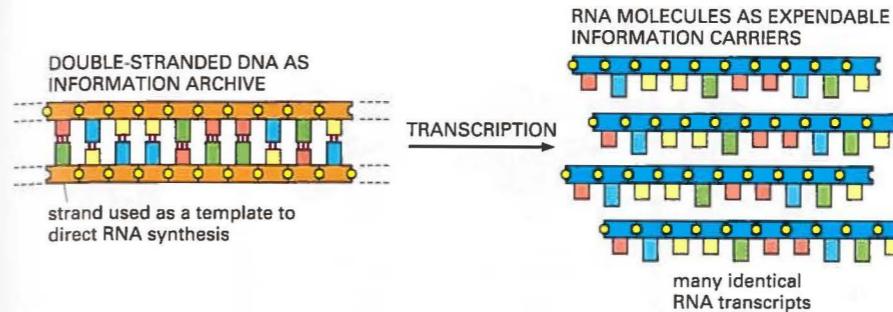


Figure 1–4 From DNA to protein. Genetic information is read out and put to use through a two-step process. First, in **transcription**, segments of the DNA sequence are used to guide the synthesis of molecules of RNA. Then, in **translation**, the RNA molecules are used to guide the synthesis of molecules of protein.



mass-produced and disposable (Figure 1–5). As we shall see, the primary role of most of these transcripts is to serve as intermediates in the transfer of genetic information: they serve as **messenger RNA (mRNA)** to guide the synthesis of proteins according to the genetic instructions stored in the DNA.

RNA molecules have distinctive structures that can also give them other specialized chemical capabilities. Being single-stranded, their backbone is flexible, so that the polymer chain can bend back on itself to allow one part of the molecule to form weak bonds with another part of the same molecule. This occurs when segments of the sequence are locally complementary: a ...GGGG... segment, for example, will tend to associate with a ...CCCC... segment. These types of internal associations can cause an RNA chain to fold up into a specific shape that is dictated by its sequence (Figure 1–6). The shape of the RNA molecule, in turn, may enable it to recognize other molecules by binding to them selectively—and even, in certain cases, to catalyze chemical changes in the molecules that are bound. As we see later in this book, a few chemical reactions catalyzed by RNA molecules are crucial for several of the most ancient and fundamental processes in living cells, and it has been suggested that more extensive catalysis by RNA played a central part in the early evolution of life (discussed in Chapter 6).

All Cells Use Proteins as Catalysts

Protein molecules, like DNA and RNA molecules, are long unbranched polymer chains, formed by the stringing together of monomeric building blocks drawn from a standard repertoire that is the same for all living cells. Like DNA and RNA, they carry information in the form of a linear sequence of symbols, in the same way as a human message written in an alphabetic script. There are many different protein molecules in each cell, and—leaving out the water—they form most of the cell's mass.

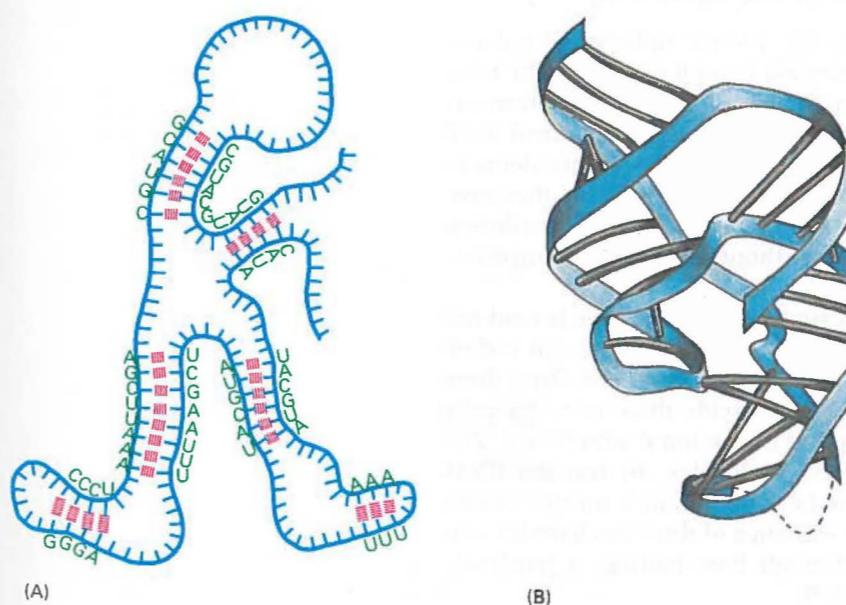
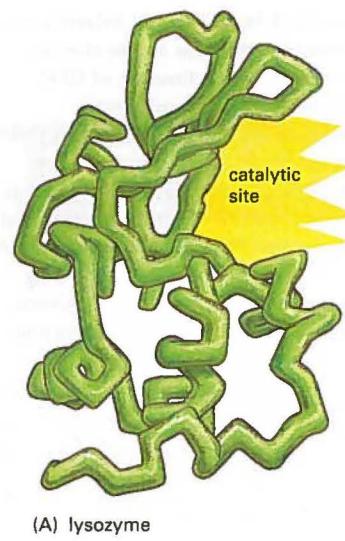


Figure 1–5 How genetic information is broadcast for use inside the cell. Each cell contains a fixed set of DNA molecules—its archive of genetic information. A given segment of this DNA serves to guide the synthesis of many identical RNA transcripts, which serve as working copies of the information stored in the archive. Many different sets of RNA molecules can be made by transcribing selected parts of a long DNA sequence, allowing each cell to use its information store differently.



(A) lysozyme

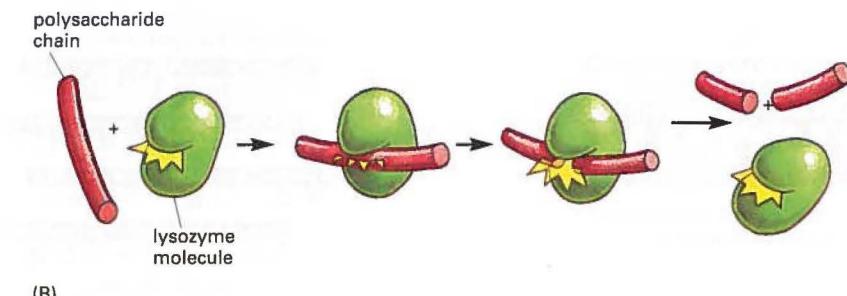


Figure 1-7 How a protein molecule acts as catalyst for a chemical reaction. (A) In a protein molecule the polymer chain folds up into a specific shape defined by its amino acid sequence. A groove in the surface of this particular folded molecule, the enzyme lysozyme, forms a catalytic site. (B) A polysaccharide molecule (red)—a polymer chain of sugar monomers—binds to the catalytic site of lysozyme and is broken apart, as a result of a covalent bond-breaking reaction catalyzed by the amino acids lining the groove.

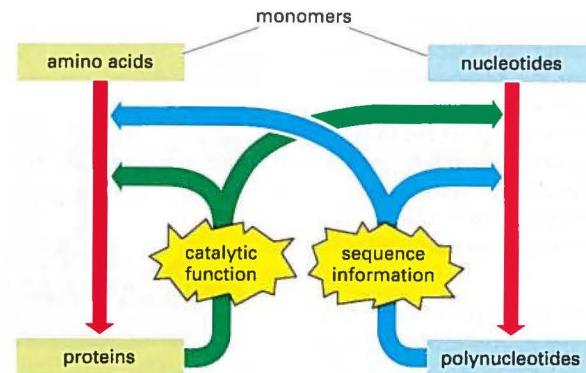
The monomers of protein, the **amino acids**, are quite different from those of DNA and RNA, and there are 20 types, instead of 4. Each amino acid is built around the same core structure through which it can be linked in a standard way to any other amino acid in the set; attached to this core is a side group that gives each amino acid a distinctive chemical character. Each of the protein molecules, or **polypeptides**, created by joining amino acids in a particular sequence folds into a precise three-dimensional form with reactive sites on its surface (Figure 1-7A). These amino acid polymers thereby bind with high specificity to other molecules and act as **enzymes** to catalyze reactions in which covalent bonds are made and broken. In this way they direct the vast majority of chemical processes in the cell (Figure 1-7B). Proteins have a host of other functions as well—maintaining structures, generating movements, sensing signals, and so on—each protein molecule performing a specific function according to its own genetically specified sequence of amino acids. Proteins, above all, are the molecules that put the cell's genetic information into action.

Thus, polynucleotides specify the amino acid sequences of proteins. Proteins, in turn, catalyze many chemical reactions, including those by which new DNA molecules are synthesized, and the genetic information in DNA is used to make both RNA and proteins. This feedback loop is the basis of the autocatalytic, self-reproducing behavior of living organisms (Figure 1-8).

All Cells Translate RNA into Protein in the Same Way

The translation of genetic information from the 4-letter alphabet of polynucleotides into the 20-letter alphabet of proteins is a complex process. The rules of this translation seem in some respects neat and rational, in other respects strangely arbitrary, given that they are (with minor exceptions) identical in all living things. These arbitrary features, it is thought, reflect frozen accidents in the early history of life—chance properties of the earliest organisms that were passed on by heredity and have become so deeply embedded in the constitution of all living cells that they cannot be changed without wrecking cell organization.

The information in the sequence of a messenger RNA molecule is read out in groups of three nucleotides at a time: each triplet of nucleotides, or **codon**, specifies (codes for) a single amino acid in a corresponding protein. Since there are 64 ($= 4 \times 4 \times 4$) possible codons, but only 20 amino acids, there are necessarily many cases in which several codons correspond to the same amino acid. The code is read out by a special class of small RNA molecules, the **transfer RNAs** (tRNAs). Each type of tRNA becomes attached at one end to a specific amino acid, and displays at its other end a specific sequence of three nucleotides—an **anticodon**—that enables it to recognize, through base-pairing, a particular codon or subset of codons in mRNA (Figure 1-9).



For synthesis of protein, a succession of tRNA molecules charged with their appropriate amino acids have to be brought together with an mRNA molecule and matched up by base-pairing through their anticodons with each of its successive codons. The amino acids then have to be linked together to extend the growing protein chain, and the tRNAs, relieved of their burdens, have to be released. This whole complex of processes is carried out by a giant multimolecular machine, the ribosome, formed of two main chains of RNA, called **ribosomal RNAs** (rRNAs), and more than 50 different proteins. This evolutionarily ancient molecular juggernaut latches onto the end of an mRNA molecule and then trundles along it, capturing loaded tRNA molecules and stitching together the amino acids they carry to form a new protein chain (Figure 1-10).

The Fragment of Genetic Information Corresponding to One Protein Is One Gene

DNA molecules as a rule are very large, containing the specifications for thousands of proteins. Segments of the entire DNA sequence are therefore transcribed into separate mRNA molecules, with each segment coding for a different protein. A **gene** is defined as the segment of DNA sequence corresponding to a single protein (or to a single catalytic or structural RNA molecule for those genes that produce RNA but not protein).

In all cells, the *expression* of individual genes is regulated: instead of manufacturing its full repertoire of possible proteins at full tilt all the time, the cell

Figure 1-8 Life as an autocatalytic process. Polynucleotides (nucleotide polymers) and proteins (amino acid polymers) provide the sequence information and the catalytic functions that serve—through a complex set of chemical reactions—to bring about the synthesis of more polynucleotides and proteins of the same types.

Figure 1-9 Transfer RNA. (A) A tRNA molecule specific for the amino acid tryptophan. One end of the tRNA molecule has tryptophan attached to it, while the other end displays the triplet nucleotide sequence CCA (its anticodon), which recognizes the tryptophan codon in messenger RNA molecules. (B) The three-dimensional structure of the tryptophan tRNA molecule. Note that the codon and the anticodon in (A) are in antiparallel orientations, like the two strands in a DNA double helix (see Figure 1-2), so that the sequence of the anticodon in the tRNA is read from right to left, while that of the codon in the mRNA is read from left to right.

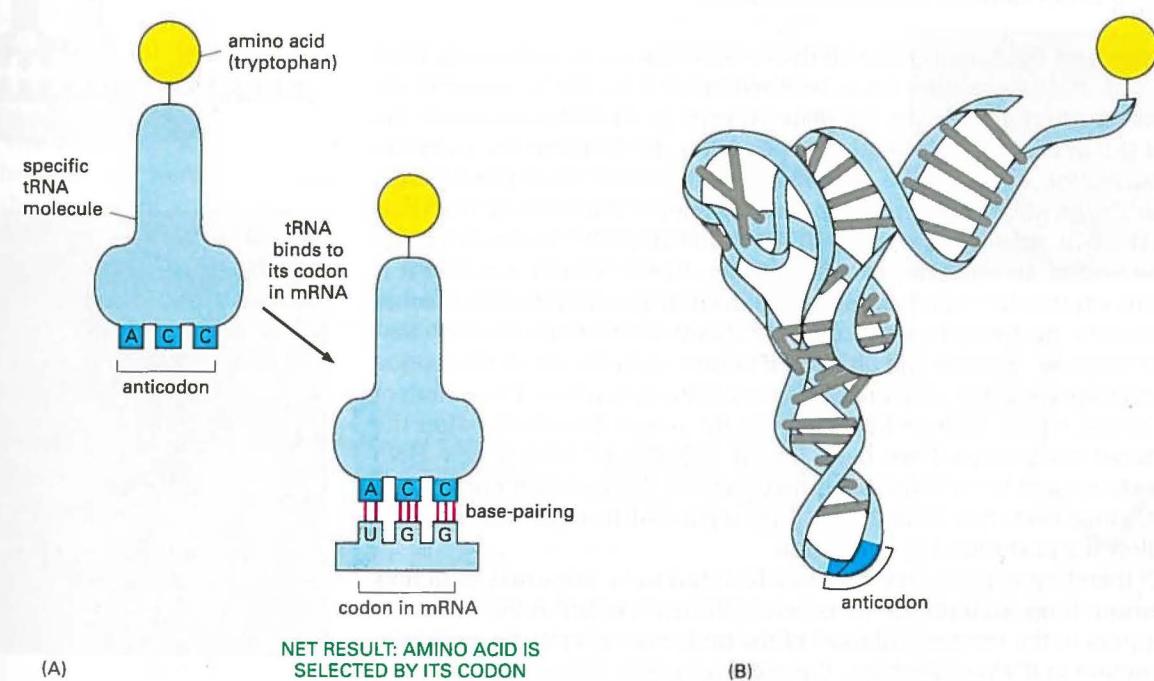


Figure 1–10 A ribosome at work. (A) The diagram shows how a ribosome moves along an mRNA molecule, capturing tRNA molecules that match the codons in the mRNA and using them to join amino acids into a protein chain. The mRNA specifies the sequence of amino acids. (B) The three-dimensional structure of a bacterial ribosome (pale green and blue), moving along an mRNA molecule (orange beads), with three tRNA molecules (yellow, green, and pink) at different stages in their process of capture and release. The ribosome is a giant assembly of more than 50 individual protein and RNA molecules. (B, courtesy of Joachim Frank, Yanhong Li, and Rajendra Agarwal.)

adjusts the rate of transcription and translation of different genes independently, according to need. Stretches of *regulatory DNA* are interspersed among the segments that code for protein, and these noncoding regions bind to special protein molecules that control the local rate of transcription (Figure 1–11). Other noncoding DNA is also present, some of it serving, for example, as punctuation, defining where the information for an individual protein begins and ends. The quantity and organization of the regulatory and other noncoding DNA vary widely from one class of organisms to another, but the basic strategy is universal. In this way, the **genome** of the cell—that is, the total of its genetic information as embodied in its complete DNA sequence—dictates not only the nature of the cell's proteins, but also when and where they are to be made.

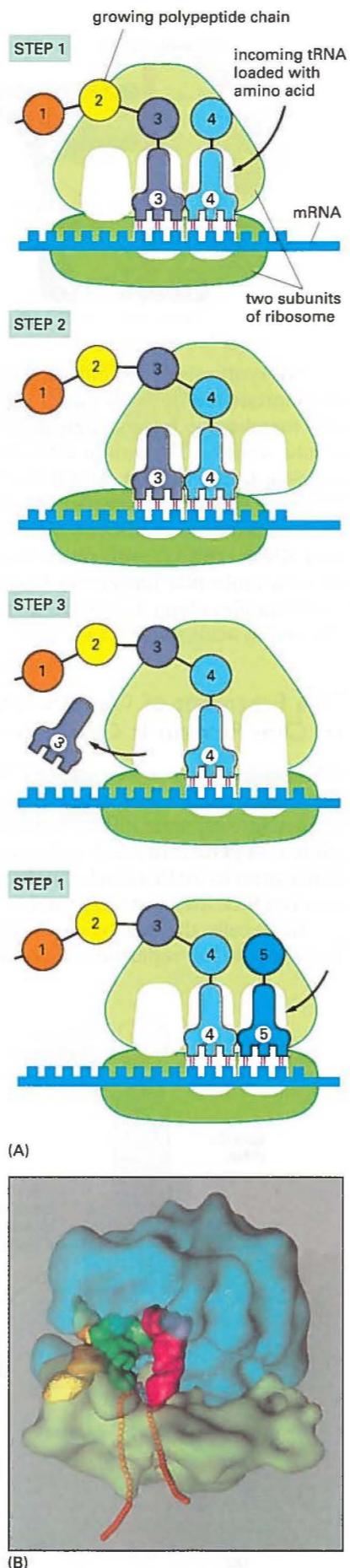
Life Requires Free Energy

A living cell is a system far from chemical equilibrium: it has a large internal free energy, meaning that if it is allowed to die and decay towards chemical equilibrium, a great deal of energy is released to the environment as heat. For the cell to make a new cell in its own image, it must take in free energy from the environment, as well as raw materials, to drive the necessary synthetic reactions. This consumption of free energy is fundamental to life. When it stops, a cell dies. Genetic information is also fundamental to life. Is there any connection?

The answer is yes: free energy is required for the propagation of information, and there is, in fact, a precise quantitative relationship between the two entities. To specify one bit of information—that is, one yes/no choice between two equally probable alternatives—costs a defined amount of free energy (measured in joules), depending on the temperature. The proof of this abstract general principle of statistical thermodynamics is quite arduous, and depends on the precise definition of the term “free energy” (discussed in Chapter 2). The basic idea, however, is not difficult to understand intuitively in the context of DNA synthesis.

To create a new DNA molecule with the same sequence as an existing DNA molecule, nucleotide monomers must be lined up in the correct sequence on the DNA strand that is used as the template. At each point in the sequence, the selection of the appropriate nucleotide depends on the fact that the correctly matched nucleotide binds to the template more strongly than mismatched nucleotides. The greater the difference in binding energy, the rarer are the occasions on which a mismatched nucleotide is accidentally inserted in the sequence instead of the correct nucleotide. A high-fidelity match, whether it is achieved through the direct and simple mechanism just outlined, or in a more complex way, with the help of a set of auxiliary chemical reactions, requires that a lot of free energy be released and dissipated as heat as each correct nucleotide is slotted into its place in the structure. This cannot happen unless the system of molecules carries a large store of free energy at the outset. Eventually, after the newly recruited nucleotides have been joined together to form a new DNA strand, a fresh input of free energy is required to force the matched nucleotides apart again, since each new strand has to be separated from its old template strand to allow the next round of replication.

The cell therefore requires free energy, which has to be imported somehow from its surroundings, to replicate its genetic information faithfully. The same principle applies to the synthesis of most of the molecules in cells. For example, in the production of RNAs or proteins, the existing genetic information dictates



the sequence of the new molecule through a process of molecular matching, and free energy is required to drive forward the many chemical reactions that construct the monomers from raw materials and link them together correctly.

All Cells Function as Biochemical Factories Dealing with the Same Basic Molecular Building Blocks

Because all cells make DNA, RNA, and protein, and these macromolecules are composed of the same set of subunits in every case, all cells have to contain and manipulate a similar collection of small molecules, including simple sugars, nucleotides, and amino acids, as well as other substances that are universally required for their synthesis. All cells, for example, require the phosphorylated nucleotide ATP (*adenosine triphosphate*) as a building block for the synthesis of DNA and RNA; and all cells also make and consume this molecule as a carrier of free energy and phosphate groups to drive many other chemical reactions.

Although all cells function as biochemical factories of a broadly similar type, many of the details of their small-molecule transactions differ, and it is not as easy as it is for the informational macromolecules to point out the features that are strictly universal. Some organisms, such as plants, require only the simplest of nutrients and harness the energy of sunlight to make from these almost all their own small organic molecules; other organisms, such as animals, feed on living things and obtain many of their organic molecules ready-made. We return to this point below.

All Cells Are Enclosed in a Plasma Membrane Across Which Nutrients and Waste Materials Must Pass

There is, however, at least one other feature of cells that is universal: each one is bounded by a membrane—the **plasma membrane**. This container acts as a selective barrier that enables the cell to concentrate nutrients gathered from its environment and retain the products it synthesizes for its own use, while excreting its waste products. Without a plasma membrane, the cell could not maintain its integrity as a coordinated chemical system.

This membrane is formed of a set of molecules that have the simple physicochemical property of being *amphipathic*—that is, consisting of one part that is hydrophobic (water-insoluble) and another part that is hydrophilic (water-soluble). When such molecules are placed in water, they aggregate spontaneously, arranging their hydrophobic portions to be as much in contact with one another as possible to hide them from the water, while keeping their hydrophilic portions exposed. Amphipathic molecules of appropriate shape, such as the phospholipid molecules that comprise most of the plasma membrane, spontaneously aggregate in water to form a *bilayer* that creates small closed vesicles (Figure 1–12). The phenomenon can be demonstrated in a test tube by simply mixing phospholipids and water together; under appropriate conditions, small vesicles form whose aqueous contents are isolated from the external medium.

Although the chemical details vary, the hydrophobic tails of the predominant membrane molecules in all cells are hydrocarbon polymers ($-\text{CH}_2-\text{CH}_2-\text{CH}_2-$), and their spontaneous assembly into a bilayered vesicle is but one of many examples of an important general principle: cells produce molecules whose chemical properties cause them to *self-assemble* into the structures that a cell needs.

Figure 1–11 (A) A diagram of a small portion of the genome of the bacterium *Escherichia coli*, containing genes (called *lacI*, *lacZ*, *lacY*, and *lacA*) coding for four different proteins. The protein-coding DNA segments (red) have regulatory and other noncoding DNA segments (yellow) between them. (B) An electron micrograph of DNA from this region, with a protein molecule (encoded by the *lacI* gene) bound to the regulatory segment; this protein controls the rate of transcription of the *lacZ*, *lacY*, and *lacA* genes. (C) A drawing of the structures shown in (B). (B, courtesy of Jack Griffith.)

