MOLECULAR BIOLOGY OF
THE CELL
THIRD EDITION

Bruce Alberts • Dennis Bray
Julian Lewis • Martin Raff • Keith Roberts
James D. Watson



Garland Publishing, Inc.
New York & London

Sarepta Exhibit 1040, page 1

GARLAND STAFF
Text Editor: Miranda Robertson
Managing Editor: Ruth Adams
Illustrator: Nigel Orme
Molecular Model Drawings: Kate Hesketh-Moore
Director of Electronic Publishing: John M-Roblin
Computer Specialist: Chuck Bartelt
Disk Preparation: Carol Winter
Copy Editor: Shirley M. Cobert
Production Editor: Douglas Goertzen
Production Coordinator: Perry Bessas
Indexer: Maija Hinkle

Bruce Alberts received his Ph.D. from Harvard University and is currently President of the National Academy of Sciences and Professor of Biochemistry and Biophysics at the University of California, San Francisco. *Dennis Bray* received his Ph.D. from the Massachusetts Institute of Technology and is currently a Medical Research Council Fellow in the Department of Zoology, University of Cambridge. *Julian Lewis* received his D.Phil. from the University of Oxford and is currently a Senior Scientist in the Imperial Cancer Research Fund Developmental Biology Unit, University of Oxford. *Martin Raff* received his M.D. from McGill University and is currently a Professor in the MRC Laboratory for Molecular Cell Biology and the Biology Department, University College London. *Keith Roberts* received his Ph.D. from the University of Cambridge and is currently Head of the Department of Cell Biology, the John Innes Institute, Norwich. *James D. Watson* received his Ph.D. from Indiana University and is currently Director of the Cold Spring Harbor Laboratory. He is the author of *Molecular Biology of the Gene* and, with Francis Crick and Maurice Wilkins, won the Nobel Prize in Medicine and Physiology in 1962.

© 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 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.].—3rd ed.
p. cm.
Includes bibliographical references and index.
ISBN 0-8153-1619-4 (hard cover).—ISBN 0-8153-1620-8 (pbk.)
1. Cytology. 2. Molecular biology. I. Alberts, Bruce.
[DNLM: 1. Cells. 2. Molecular Biology. QH 581.2 M718 1994]
QH581.2.M64 1994
574.87—dc20
DNLM/DLC
for Library of Congress
93-45907
CIP

Published by Garland Publishing, Inc.
717 Fifth Avenue, New York, NY 10022

Printed in the United States of America

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

Front cover: The photograph shows a rat nerve cell in culture. It is labeled (*yellow*) with a fluorescent antibody that stains its cell body and dendritic processes. Nerve terminals (*green*) from other neurons (not visible), which have made synapses on the cell, are labeled with a different antibody. (Courtesy of Olaf Mundigl and Pietro de Camilli.)

Dedication page: Gavin Borden, late president of Garland Publishing, weathered in during his mid-1980s climb near Mount McKinley with MBoC author Bruce Alberts and famous mountaineer guide Mugs Stump (1940–1992).

Back cover: The authors, in alphabetical order, crossing Abbey Road in London on their way to lunch. Much of this third edition was written in a house just around the corner. (Photograph by Richard Olivier.)

Small Molecules, Energy, and Biosynthesis

2

- The Chemical Components of a Cell
- Biological Order and Energy
- Food and the Derivation of Cellular Energy
- Biosynthesis and the Creation of Order
- The Coordination of Catabolism and Biosynthesis

"I must tell you that I can prepare urea without requiring a kidney or an animal, either man or dog." This sentence, written 165 years ago by the young German chemist Wöhler, signaled an end to the belief in a special *vital force* that exists in living organisms and gives rise to their distinctive properties and products. But what was a revelation in Wöhler's time is common knowledge today—living creatures are made of chemicals, obedient simply to the laws of chemistry and physics. This is not to say that no mysteries remain in biology: there are many areas of ignorance, as will become apparent in later chapters. But we should begin by emphasizing the enormous amount that is known.

We now have detailed information about the essential molecules of the cell—not just a small number of molecules, but thousands of them. In many cases we know their precise chemical structures and exactly how they are made and broken down. We know in general terms how chemical energy drives the biosynthetic reactions of the cell, how thermodynamic principles operate in cells to create molecular order, and how the myriad chemical changes occurring continuously within cells are controlled and coordinated.

In this and the next chapter we briefly survey the chemistry of the living cell. Here we deal with the processes involving small molecules: those mechanisms by which the cell synthesizes its fundamental chemical ingredients and by which it obtains its energy. Chapter 3 describes the giant molecules of the cell, which are polymers of a subset of the small molecules; these polymers are responsible both for the specificity of biological processes and for the transfer of biological information.

The Chemical Components of a Cell

Cell Chemistry Is Based on Carbon Compounds [1]

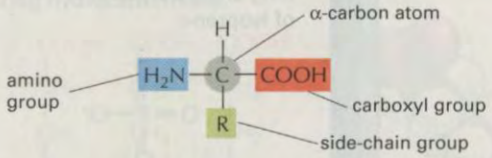
A living cell is composed of a restricted set of elements, four of which (C, H, N, and O) make up nearly 99% of its weight. This composition differs markedly from that of the earth's crust and is evidence of a distinctive type of chemistry (Figure 2–1). What is this special chemistry, and how did it evolve?

The most abundant substance of the living cell is water. It accounts for about 70% of a cell's weight, and most intracellular reactions occur in an aqueous environment. Life on this planet began in the ocean, and the conditions in that

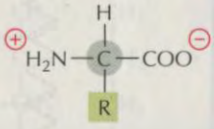
Sarepta Exhibit 1040, page 3

THE AMINO ACID

The general formula of an amino acid is

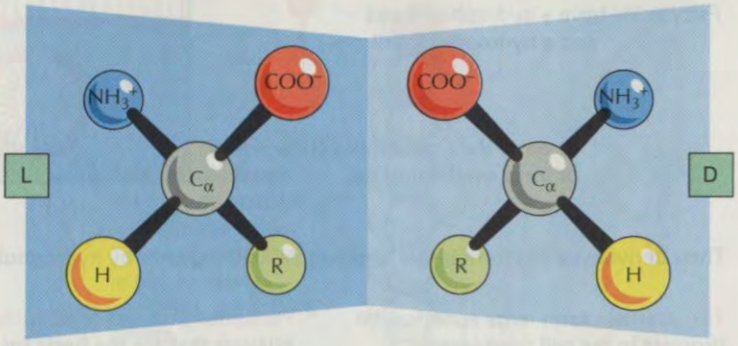


R is commonly one of 20 different side chains. At pH 7 both the amino and carboxyl groups are ionized.



OPTICAL ISOMERS

The α-carbon atom is asymmetric, which allows for two mirror image (or stereo-) isomers, D and L.

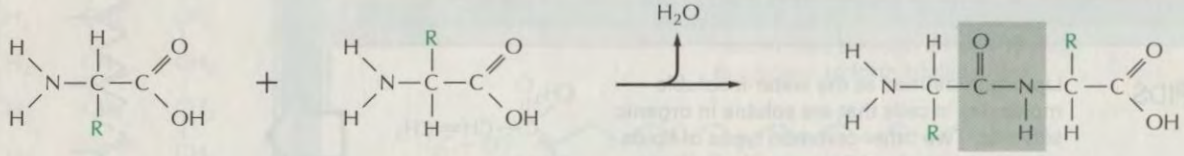


Proteins consist exclusively of L-amino acids.

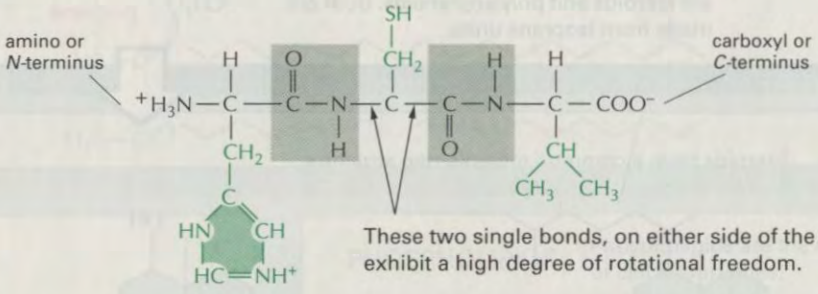
PEPTIDE BONDS

Amino acids are commonly joined together by an amide linkage, called a peptide bond.

peptide bond: The four atoms in each *gray box* form a rigid planar unit. There is no freedom of rotation about the C—N bond.



Proteins are long polymers of amino acids linked by peptide bonds, and they are always written with the *N*-terminus toward the left. The sequence of this tripeptide is His Cys Val.



These two single bonds, on either side of the rigid peptide unit, exhibit a high degree of rotational freedom.

FAMILIES OF AMINO ACIDS

The common amino acids are grouped according to whether their side chains are

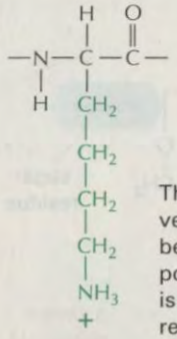
- acidic
- basic
- uncharged polar
- nonpolar

These 20 amino acids are given both three-letter and one-letter abbreviations.

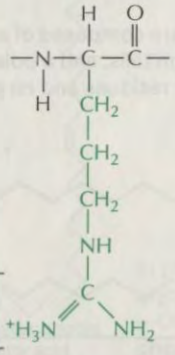
Thus: alanine = Ala = A

BASIC SIDE CHAINS

lysine (Lys, or K)

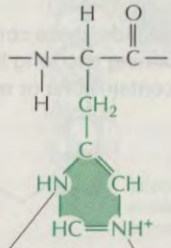


arginine (Arg, or R)



This group is very basic because its positive charge is stabilized by resonance.

histidine (His, or H)



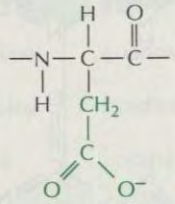
These nitrogens have a relatively weak affinity for an H^+ and are only partly positive at neutral pH.

56 **Panel 2–5 The 20 amino acids involved in the synthesis of proteins.**

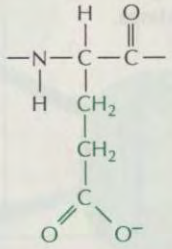
Sarepta Exhibit 1040, page 4

ACIDIC SIDE CHAINS

aspartic acid
(Asp, or D)



glutamic acid
(Glu, or E)



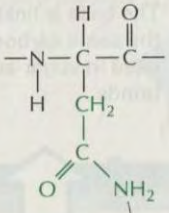
Amino acids with uncharged polar side chains are relatively hydrophilic and are usually on the outside of proteins, while the side chains on nonpolar amino acids tend to cluster together on the inside. Amino acids with basic or acidic side chains are very polar, and they are nearly always found on the outside of protein molecules.

The one letter code in alphabetical order:

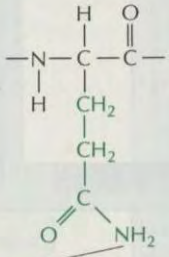
A = Ala	G = Gly	M = Met	S = Ser
C = Cys	H = His	N = Asn	T = Thr
D = Asp	I = Ileu	P = Pro	V = Val
E = Glu	K = Lys	Q = Gln	W = Trp
F = Phe	L = Leu	R = Arg	Y = Tyr

UNCHARGED POLAR SIDE CHAINS

asparagine
(Asn, or N)

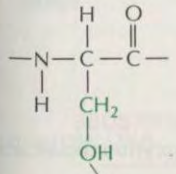


glutamine
(Gln, or Q)

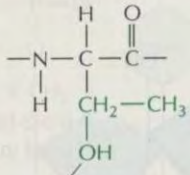


Although the amide N is not charged at neutral pH, it is polar.

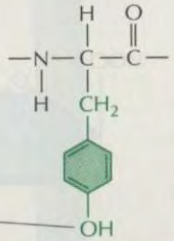
serine
(Ser, or S)



threonine
(Thr, or T)

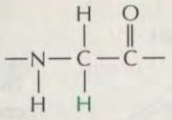


tyrosine
(Tyr, or Y)



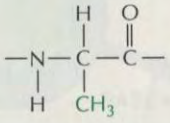
The —OH group is polar.

NONPOLAR SIDE CHAINS

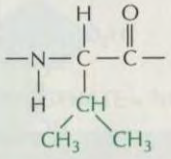


glycine
(Gly, or G)

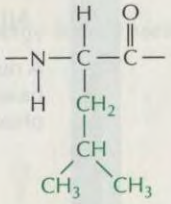
alanine
(Ala, or A)



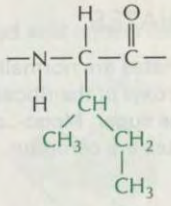
valine
(Val, or V)



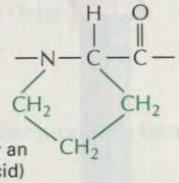
leucine
(Leu, or L)



isoleucine
(Ileu, or I)

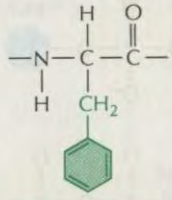


proline
(Pro, or P)

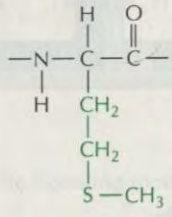


(actually an imino acid)

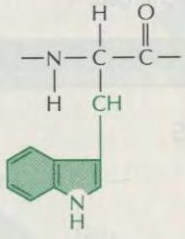
phenylalanine
(Phe, or F)

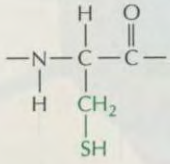


methionine
(Met, or M)



tryptophan
(Trp, or W)





cysteine
(Cys, or C)

Paired cysteines allow disulfide bonds to form in proteins.



Sarepta Exhibit 1040, page 5

Macromolecules: Structure, Shape, and Information

3

- Molecular Recognition Processes
- Nucleic Acids
- Protein Structure
- Proteins as Catalysts

In moving from the small molecules of the cell to the giant macromolecules, we encounter a transition of more than size alone. Even though proteins, nucleic acids, and polysaccharides are made from a limited repertoire of amino acids, nucleotides, and sugars, respectively, they can have unique and truly astounding properties that bear little resemblance to those of their simple chemical precursors. Biological macromolecules are composed of many thousands—sometimes millions—of atoms linked together in precisely defined spatial arrangements. Each of these macromolecules carries specific information. Incorporated in its structure is a series of biological messages that can be "read" in its interactions with other molecules, enabling it to perform a precise function.

In this chapter we examine the structures of macromolecules, emphasizing proteins and nucleic acids, and explain how they have adapted in the course of evolution to perform specific functions. We consider the principles by which these molecules catalyze chemical transformations, build complex multimolecular structures, generate movement, and—most fundamental of all—store and transmit hereditary information.

Molecular Recognition Processes [1]

Macromolecules typically have molecular weights between about 10,000 and 1 million and are intermediate in size between the organic molecules of the cell discussed in Chapter 2 and the large macromolecular assemblies and organelles that will be discussed in subsequent chapters (Figure 3–1). One small molecule, water, constitutes 70% of the total mass of a cell; nearly all of the remaining cell mass is due to macromolecules (Table 3–1).

As described in Chapter 2, a macromolecule is assembled from low-molecular-weight subunits that are repeatedly added to one end to form a long, chainlike polymer. Usually only one family of subunits is used to construct each chain: amino acids are linked to other amino acids to form proteins, nucleotides are linked to other nucleotides to form nucleic acids, and sugars are linked to other sugars to form polysaccharides. Because the precise sequence of subunits is crucial to the function of a macromolecule, its biosynthesis requires mechanisms to ensure that the correct subunit goes into the polymer at each position in the chain.

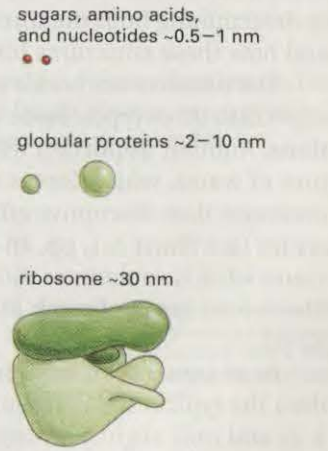


Figure 3–1 The size of protein molecules compared to some other cell components. The ribosome is an important macromolecular assembly composed of about 60 protein and RNA molecules.

Sarepta Exhibit 1040, page 6

Table 3–1 Approximate Chemical Compositions of a Typical Bacterium and a Typical Mammalian Cell

Component	Percent of Total Cell Weight	
	E. Coli *Bacterium*	*Mammalian Cell*
H_2O	70	70
Inorganic ions (Na^+, K^+, Mg^{2+}, Ca^{2+}, Cl^-, etc.)	1	1
Miscellaneous small metabolites	3	3
Proteins	15	18
RNA	6	1.1
DNA	1	0.25
Phospholipids	2	3
Other lipids	—	2
Polysaccharides	2	2
Total cell volume:	2×10^{-12} cm^3	4×10^{-9} cm^3
Relative cell volume:	1	2000

Proteins, polysaccharides, DNA, and RNA are macromolecules. Lipids are not generally classed as macromolecules even though they share some of their features; for example, most are synthesized as linear polymers of a smaller molecule (the acetyl group on acetyl CoA), and they self-assemble into larger structures (membranes). Note that water and protein comprise most of the mass of both mammalian and bacterial cells.

The Specific Interactions of a Macromolecule Depend on Weak, Noncovalent Bonds [2]

A macromolecular chain is held together by *covalent* bonds, which are strong enough to preserve the sequence of subunits for long periods of time. Although the sequence of subunits determines the information content of a macromolecule, utilizing that information depends largely on much weaker, *noncovalent* bonds. These weak bonds form between different parts of the same macromolecule and between different macromolecules. They therefore play a major part in determining both the three-dimensional structure of macromolecular chains and how these structures interact with one another.

The noncovalent bonds encountered in biological molecules are usually classified into three types: **ionic bonds, hydrogen bonds,** and **van der Waals attractions.** Another important weak force is created by the three-dimensional structure of water, which forces exposed hydrophobic groups together in order to minimize their disruptive effect on the hydrogen-bonded network of water molecules (see Panel 2–1, pp. 48–49). This expulsion from the aqueous solution generates what is sometimes thought of as a fourth kind of weak, noncovalent bond. These four types of weak attractive forces are the subject of Panel 3–1, pages 92–93.

In an aqueous environment each noncovalent bond is 30 to 300 times weaker than the typical covalent bonds that hold biological molecules together (Table 3–2) and only slightly stronger than the average energy of thermal collisions at

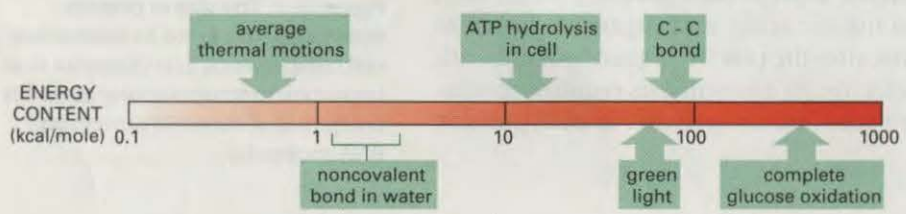


Figure 3–2 Comparative energies of some important molecular events in cells. Note that energy is displayed on a logarithmic scale.

Sarepta Exhibit 1040, page 7

Table 3–2 Covalent and Noncovalent Chemical Bonds

Bond Type	Length (nm)	Strength (kcal/mole)*	
		In Vacuum	*In Water*
Covalent	0.15	90	90
Ionic	0.25	80	3
Hydrogen	0.30	4	1
van der Waals attraction (per atom)	0.35	0.1	0.1

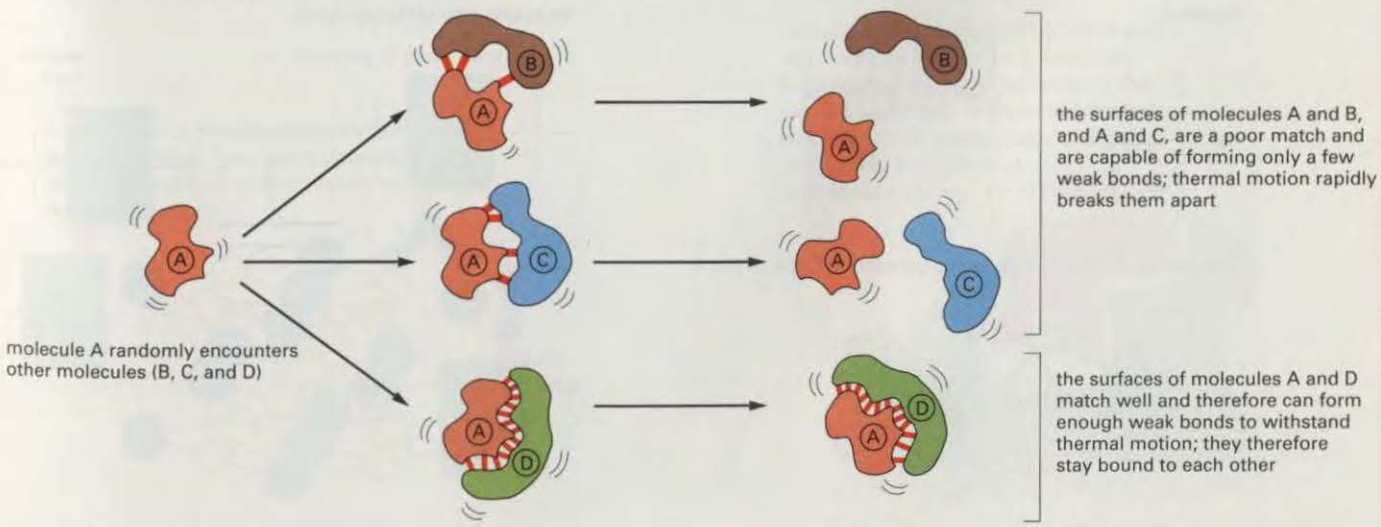
*The strength of a bond can be measured by the energy required to break it, here given in kilocalories per mole (kcal/mole). (*One kilocalorie* is the quantity of energy needed to raise the temperature of 1000 g of water by 1°C. An alternative unit in wide use is the kilojoule, kJ, equal to 0.24 kcal.) Individual bonds vary a great deal in strength, depending on the atoms involved and their precise environment, so that the above values are only a rough guide. Note that the aqueous environment in a cell will greatly weaken both the ionic and the hydrogen bonds between nonwater molecules (Panel 3–1, pp. 92–93). The bond length is the center-to-center distance between the two interacting atoms; the length given here for a hydrogen bond is that between its two nonhydrogen atoms.

37°C (Figure 3–2). A single noncovalent bond—unlike a single covalent bond—is therefore too weak to withstand the thermal motions that tend to pull molecules apart. Large numbers of noncovalent bonds are needed to hold two molecular surfaces together, and these can form between two surfaces only when large numbers of atoms on the surfaces are precisely matched to each other (Figure 3–3). The exacting requirements for matching account for the specificity of biological recognition, such as occurs between an enzyme and its substrate.

As explained at the top of Panel 3–1, atoms behave almost as if they were hard spheres with a definite radius (their van der Waals radius). The requirement that no two atoms overlap limits the possible bond angles in a polypeptide chain (Figure 3–4). These and other steric interactions severely constrain the number of three-dimensional arrangements of atoms (or **conformations**) that are possible. Nevertheless, a long flexible chain such as a protein can still fold in an enormous number of ways. Each conformation will have a different set of weak intrachain interactions, and it is the total strength of these interactions that determines which conformations will form.

Most proteins in a cell fold stably in only one way: during the course of evolution the sequence of amino acid subunits in each protein has been selected so that one conformation is able to form many more favorable intrachain interactions than any other.

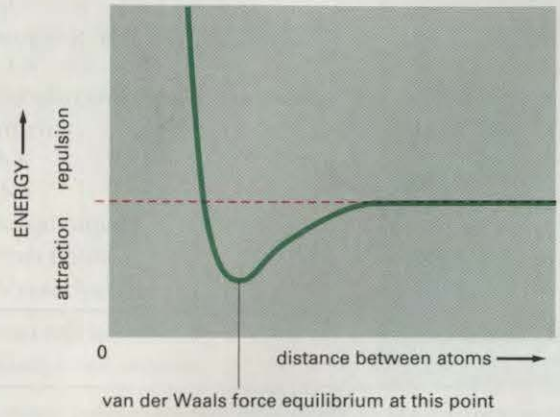
Figure 3–3 Noncovalent bonds. How weak bonds mediate recognition between macromolecules.



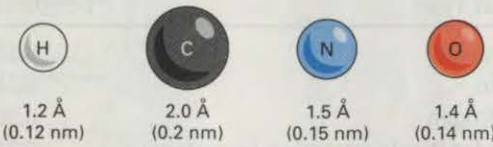
Sarepta Exhibit 1040, page 8

VAN DER WAALS FORCES

At very short distances any two atoms show a weak bonding interaction due to their fluctuating electrical charges. This force is known as van der Waals attraction. However, two atoms will very strongly repel each other if they are brought too close together. This van der Waals repulsion plays a major part in limiting the possible conformations of a molecule.



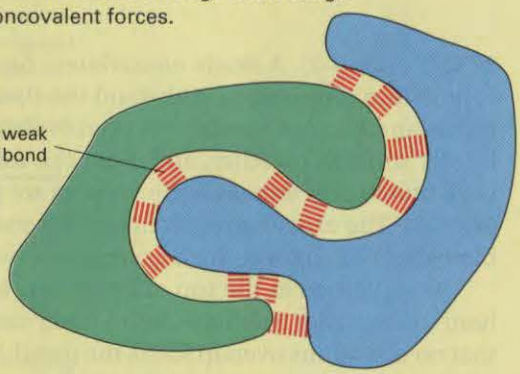
Each type of atom has a radius, known as its van der Waals radius, at which van der Waals forces are in equilibrium.



Two atoms will be attracted to each other by van der Waals forces until the distance between them equals the sum of their van der Waals radii. Although they are individually very weak, these van der Waals attractions can become important when two macromolecular surfaces fit very close together.

WEAK CHEMICAL BONDS

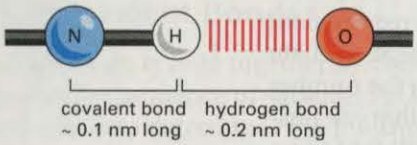
Organic molecules can interact with other molecules through short-range noncovalent forces.



Weak chemical bonds have less than 1/20 the strength of a strong covalent bond. They are strong enough to provide tight binding only when many of them are formed simultaneously.

HYDROGEN BONDS

A hydrogen atom is shared between two other atoms (both electronegative, such as O and N) to give a hydrogen bond.

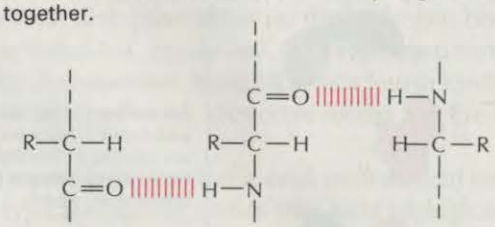


Hydrogen bonds are strongest when the three atoms are in a straight line:

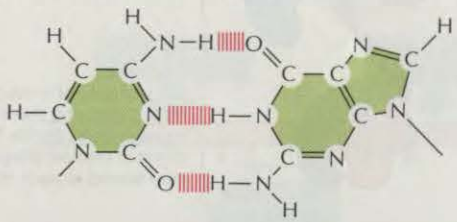


Examples in macromolecules:

Amino acids in polypeptide chains hydrogen-bonded together.

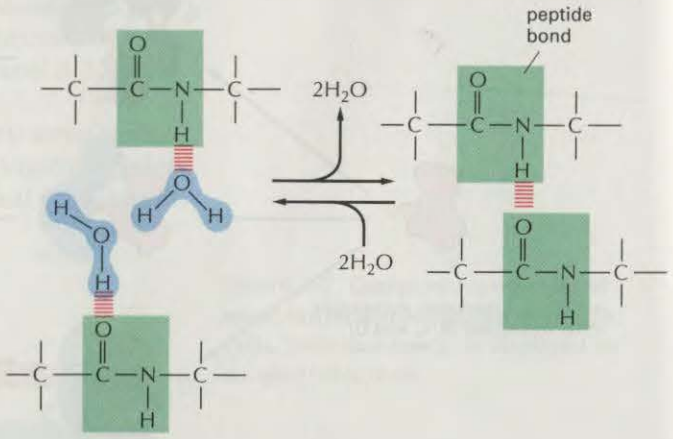


Two bases, G and C, hydrogen-bonded in DNA or RNA.



HYDROGEN BONDS IN WATER

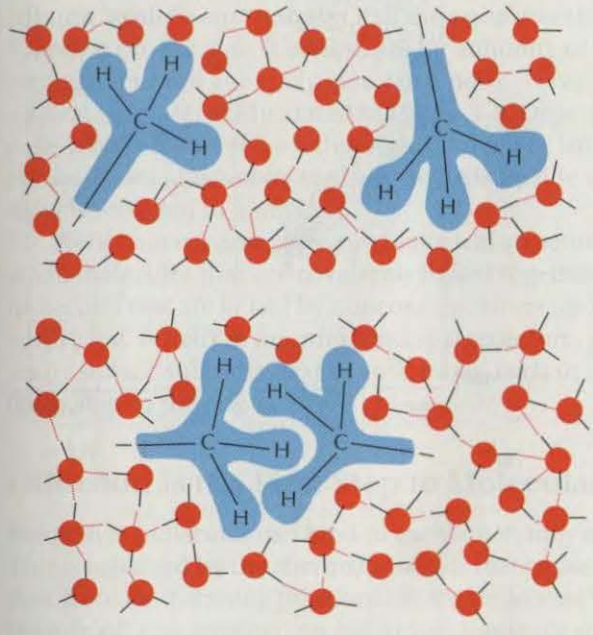
Any molecules that can form hydrogen bonds to each other can alternatively form hydrogen bonds to water molecules. Because of this competition with water molecules, the hydrogen bonds formed between two molecules dissolved in water are relatively weak.



92 **Panel 3–1 The principal types of weak noncovalent forces that hold macromolecules together.**

Sarepta Exhibit 1040, page 9

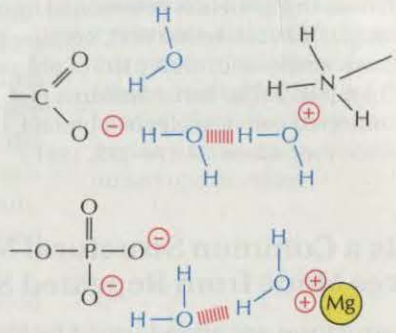
HYDROPHOBIC FORCES



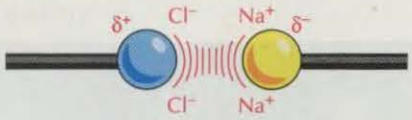
Water forces hydrophobic groups together in order to minimize their disruptive effects on the hydrogen-bonded water network. Hydrophobic groups held together in this way are sometimes said to be held together by "hydrophobic bonds," even though the attraction is actually caused by a repulsion from the water.

IONIC BONDS IN AQUEOUS SOLUTIONS

Charged groups are shielded by their interactions with water molecules. Ionic bonds are therefore quite weak in aqueous solution.

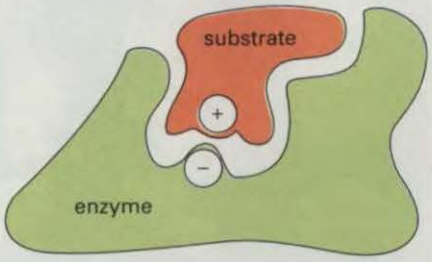


Ionic bonds are further weakened by the presence of salts, whose atoms form the counterions that cluster around ions of opposite charge.



Measurement of the extent of destabilization of an interaction by salt provides a quantitative estimate of the total number of ionic bonds involved.

Despite being weakened by water and salt, ionic bonds are very important in biological systems; an enzyme that binds a positively charged substrate will often have a negatively charged amino acid side chain at the appropriate place.



IONIC BONDS

Ionic interactions occur either between fully charged groups (ionic bond) or between partially charged groups.



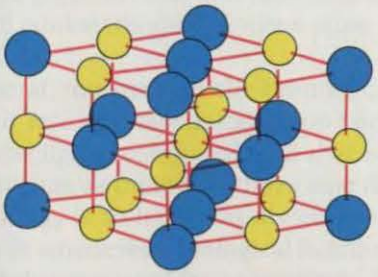
The force of attraction between the two charges δ^+ and δ^- is

$$\text{force} = \frac{\delta^+ \delta^-}{r^2 D} \quad \text{(Coulomb's law)}$$

where D = dielectric constant
(1 for vacuum; 80 for water)

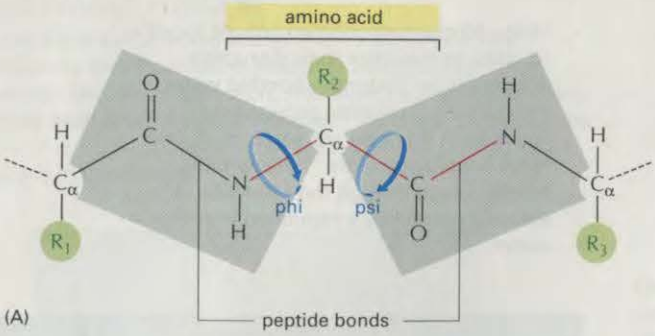
r = distance of separation

In the absence of water, ionic forces are very strong. They are responsible for the strength of such minerals as marble and agate.



NaCl crystal

Sarepta Exhibit 1040, page 10



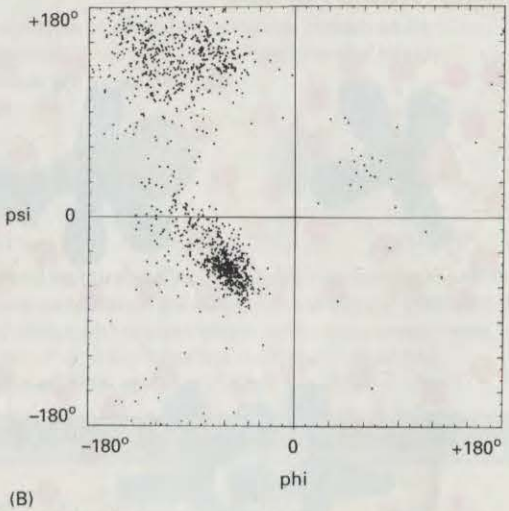


Figure 3–4 Steric limitations on the bond angles in a polypeptide chain. (A) Each amino acid contributes three bonds (colored *red*) to its polypeptide chain. The peptide bond is planar (*gray* shading) and does not permit rotation. By contrast, rotation can occur about the C_α—C bond, whose angle of rotation is called psi (ψ), and about the N—C_α bond, whose angle of rotation is called phi (φ). The R group denotes an amino acid side chain. (B) The conformation of the main-chain atoms in a protein is determined by one pair of phi and psi angles for each amino acid; because of steric collisions within each amino acid, most pairs of phi and psi angles do not occur. In this so-called Ramachandran plot, each dot represents an observed pair of angles in a protein. (B, from J. Richardson, *Adv. Prot. Chem.* 34:174–175, 1981.)

A Helix Is a Common Structural Motif in Biological Structures Made from Repeated Subunits [3]

Biological structures are often formed by linking subunits that are very similar to each other—such as amino acids or nucleotides—into a long, repetitive chain. If all the subunits are identical, neighboring subunits in the chain will often fit together in only one way, adjusting their relative positions so as to minimize the

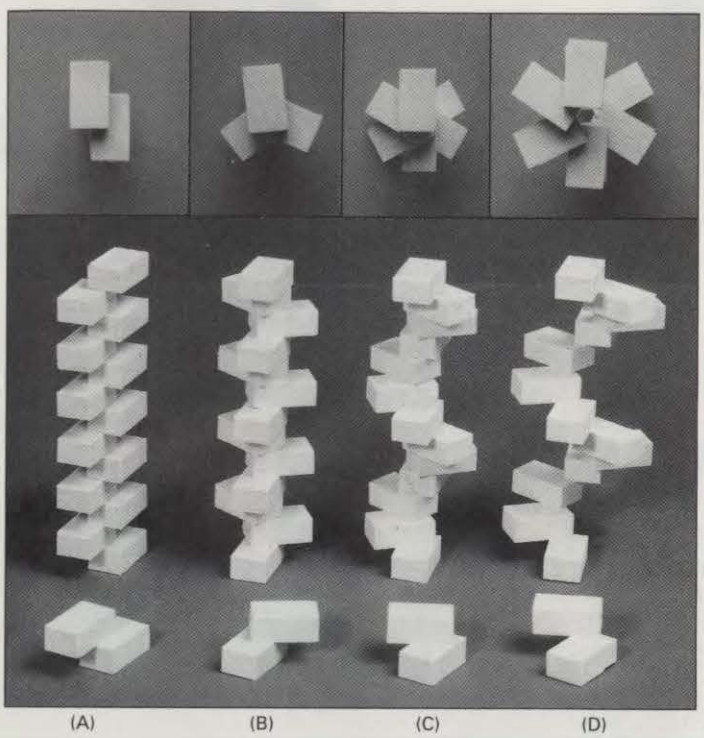


Figure 3–5 A helix will form when a series of subunits bind to each other in a regular way. In the foreground the interaction between two subunits is shown; behind it are the helices that result. These helices have two (A), three (B), and six (C and D) subunits per turn. At the top, the arrangement of subunits has been photographed from directly above the helix. Note that the helix in (D) has a wider path than that in (C).

Sarepta Exhibit 1040, page 11

free energy of the contact between them. In this case, each subunit will be positioned in exactly the same way in relation to its neighboring subunits, so that subunit 3 will fit onto subunit 2 in the same way that subunit 2 fits onto subunit 1, and so on. Because it is very rare for subunits to join up in a straight line, this arrangement will generally result in a **helix**—a regular structure that resembles a spiral staircase, as illustrated in Figure 3–5. Depending on the twist of the staircase, a helix is said to be either right-handed or left-handed (Figure 3–6). Handedness is not affected by turning the helix upside down, but it is reversed if the helix is reflected in a mirror.

Helices occur commonly in biological structures, whether the subunits are small molecules that are covalently linked together (as in DNA) or large protein molecules that are linked by noncovalent forces (as in actin filaments). This is not surprising. A helix is an unexceptional structure, generated simply by placing many similar subunits next to each other, each in the same strictly repeated relationship to the one before.

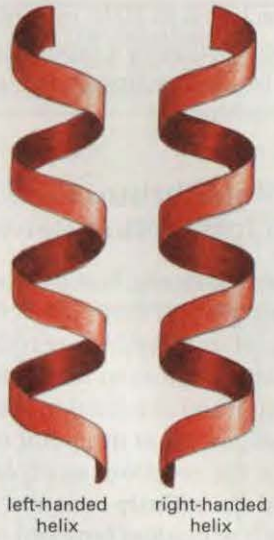


Figure 3–6 Comparison of a left-handed and a right-handed helix. As a reference, it is useful to remember that standard screws, which insert when turned clockwise, are right-handed. Note that a helix preserves the same handedness when it is turned upside down.

Diffusion Is the First Step to Molecular Recognition [4]

Before two molecules can bind to each other, they must come into close contact. This is achieved by the thermal motions that cause molecules to wander, or *diffuse*, from their starting positions. As the molecules in a liquid rapidly collide and bounce off one another, an individual molecule moves first one way and then another, its path constituting a "random walk" (Figure 3–7). The average distance that each type of molecule travels from its starting point is proportional to the square root of the time involved: that is, if it takes a particular molecule 1 second on average to go 1 μm, it will go 2 μm in 4 seconds, 10 μm in 100 seconds, and so on. Diffusion is therefore an efficient way for molecules to move limited distances but an inefficient way for molecules to move long distances.

Experiments performed by injecting fluorescent dyes and other labeled molecules into cells show that the diffusion of small molecules through the cytoplasm is nearly as rapid as it is in water. A molecule the size of ATP, for example, requires only about 0.2 second to diffuse an average distance of 10 μm—the diameter of a small animal cell. Large macromolecules, however, move much more slowly. Not only is their diffusion rate intrinsically slower, but their movement is retarded by frequent collisions with many other macromolecules that are held in place by molecular associations in the cytoplasm (Figure 3–8).

Thermal Motions Bring Molecules Together and Then Pull Them Apart [4]

Encounters between two macromolecules or between a macromolecule and a small molecule occur randomly through simple diffusion. An encounter may lead immediately to the formation of a complex between the two molecules, in which case the rate of complex formation is said to be *diffusion-limited.* Alternatively, the rate of complex formation may be slower, requiring some adjustment of the structure of one or both molecules before the interacting surfaces can fit together, so that most often the two colliding molecules will bounce off each other without sticking. In either case once the two interacting surfaces have come sufficiently close together, they will form multiple weak bonds with each other that persist until random thermal motions cause the molecules to dissociate again (see Figure 3–3).

In general, the stronger the binding of the molecules in the complex, the slower their rate of dissociation. At one extreme the total energy of the bonds formed is negligible compared with that of thermal motion, and the two molecules dissociate as rapidly as they came together. At the other extreme the total bond energy is so high that dissociation rarely occurs. Strong interactions occur in cells whenever a biological function requires that two macromolecules remain tightly associated for a long time—for example, when a gene regulatory



Figure 3–7 A random walk. Molecules in solution move in a random fashion due to the continual buffeting they receive in collisions with other molecules. This movement allows small molecules to diffuse from one part of the cell to another in a surprisingly short time: such molecules will generally diffuse across a typical animal cell in less than a second.

Sarepta Exhibit 1040, page 12

protein binds to DNA to turn off a gene. Weaker interactions occur when the function demands a rapid change in the structure of a complex—for example, when two interacting proteins change partners during the movements of a protein machine.

The Equilibrium Constant Is a Measure of the Strength of an Interaction Between Two Molecules [5]

The precise strength of the bonding between two molecules is a useful index of the specificity of their interaction. To illustrate how the binding strength is measured, let us consider a reaction in which molecule A binds to molecule B. The reaction will proceed until it reaches an *equilibrium point*, at which the rates of formation and dissociation are equal (Figure 3–9). The concentrations of A, B, and the complex AB at this point can be used to determine an **equilibrium constant** (***K***) for the reaction, as explained in Figure 3–9. This constant is sometimes termed the **affinity constant** and is commonly employed as a measure of the strength of binding between two molecules: the *stronger* the binding, the *larger* is the value of the affinity constant.

The equilibrium constant of a reaction in which two molecules bind to each other is related directly to the standard free-energy change for the binding (ΔG°) by the equation described in Table 3–3. The table also lists the ΔG° values corresponding to a range of K values. Affinity constants for simple binding interactions in biological systems often range between 10^3 and 10^{12} liters/mole; this corresponds to binding energies in the range 4–17 kcal/mole, which could arise from 4 to 17 average hydrogen bonds.



Figure 3–8 Macromolecules in the cell cytoplasm. The drawing is approximately to scale and emphasizes the crowding in the cytoplasm. Only the macromolecules are shown: RNAs are shown in *blue*, ribosomes in *green*, and proteins in *red*. Macromolecules diffuse relatively slowly in the cytoplasm because they interact with many other macromolecules; small molecules, by contrast, diffuse nearly as rapidly as they do in water. (Adapted from D.S. Goodsell, *Trends in Biochem. Sci.* 16:203–206, 1991.)

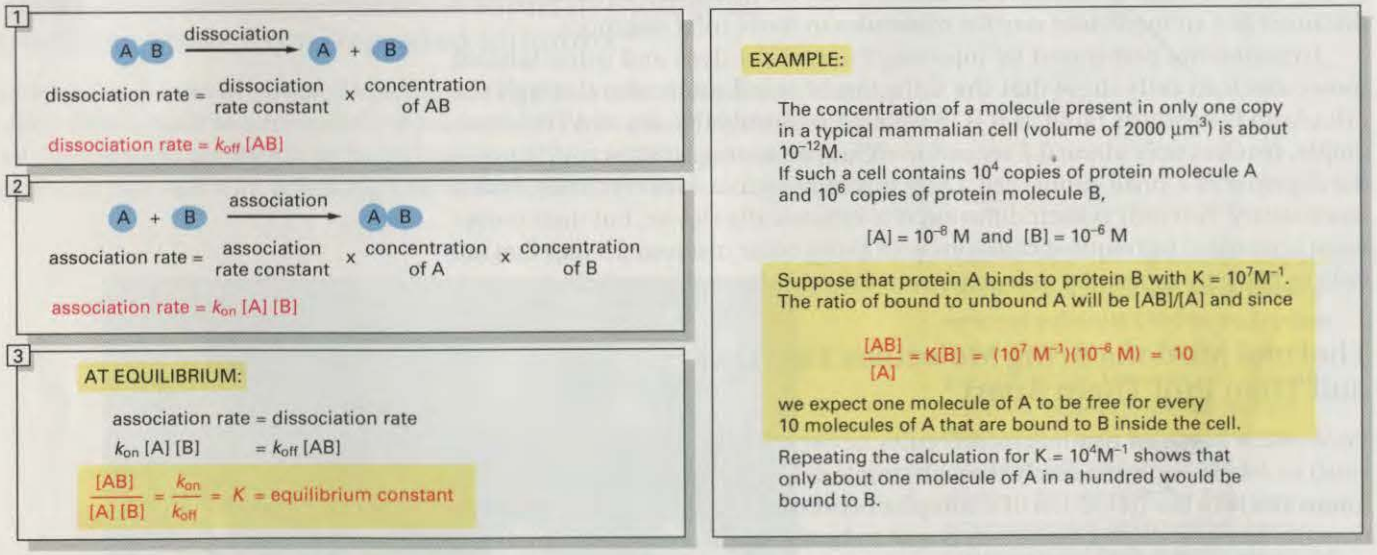


Figure 3–9 The principle of equilibrium. The equilibrium between molecules A and B and the complex AB is maintained by a balance between the two opposing reactions shown in (1) and (2). As shown in (3), the ratio of the rate constants for the association and the dissociation reactions is equal to the *equilibrium constant (K)* for the reaction. Molecules A and B must collide in order to react, and the rate in reaction (2) is therefore proportional to the product of their individual concentrations. As a result, the product [A] × [B] appears in the final expression for *K*, where [ ] indicates concentration.

As traditionally defined, the concentrations of products appear in the numerator and the concentrations of reactants appear in the denominator of the equation for an equilibrium constant. Thus the equilibrium constant in (3) is that for the *association* reaction A + B → AB. For simple binding interactions this constant is called the *affinity constant* or *association constant* (in units of liters per mole); the larger the value of the association constant (K_a), the stronger is the binding between A and B. The reciprocal of K_a is the *dissociation constant* (in units of moles per liter); the smaller the value of the dissociation constant (K_d), the stronger is the binding between A and B.

Sarepta Exhibit 1040, page 13

Atoms and Molecules Move Very Rapidly [6]

The chemical reactions in a cell occur at amazingly fast rates. A typical enzyme molecule, for example, will catalyze on the order of 1000 reactions per second, and rates of more than 10^6 reactions per second are achieved by some enzymes. Since each reaction requires a separate encounter between an enzyme and a substrate molecule, such rates are possible only because the molecules are moving so rapidly. Molecular motions can be classified broadly into three kinds: (1) the movement of a molecule from one place to another (translational motion), (2) the rapid back-and-forth movement of covalently linked atoms with respect to one another (vibrations), and (3) rotations. All of these motions are important in bringing the surfaces of interacting molecules together.

The rates of molecular motions can be measured by a variety of spectroscopic techniques. These indicate that a large globular protein is constantly tumbling, rotating about its axis about a million times per second. The rates of diffusional encounters due to translational movements are proportional to the concentration of the diffusing molecule. If ATP is present at its typical intracellular concentration of about 1 mM, for instance, each site on a protein molecule will be bombarded by about 10^6 random collisions with ATP molecules per second; for an ATP concentration tenfold lower, the number of collisions would drop to 10^5 per second and so on.

Once two molecules have collided and are in the correct relative orientation, a chemical reaction can occur between them extremely rapidly. When one appreciates how quickly molecules move and react, the observed rates of enzymatic catalysis do not seem so amazing.

Molecular Recognition Processes Can Never Be Perfect [7]

All molecules possess energy—the kinetic energy of their translational movements, vibrations, and rotations and the potential energy stored in their electron distributions. Through molecular collisions this energy is randomly distributed to all of the atoms present, so that most atoms will have energy levels close to the average, with only a small proportion possessing very high energy. Although the favored conformations or states for a molecule will be those of lowest free energy (see p. 75), states of higher energy occur through unusually violent collisions. Given the temperature, it is possible to calculate the probability that an atom or a molecule will be in a particular energy state (see Table 3–3). The probability of a high-energy state becomes smaller relative to a low-energy state as the difference in free energy between the two increases. It reaches zero, however, only when this energy difference becomes infinite.

Because of the random factor in molecular interactions, minor "side reactions" are bound to occur occasionally. As a consequence, a cell continually makes errors. Even reactions that are very energetically unfavorable will take place occasionally. Two atoms joined to each other by a covalent bond, for example, will eventually be subjected to an especially energetic collision and fall apart. Similarly, the specificity of an enzyme for its substrate cannot be absolute because the recognition of one molecule as distinct from another can never be perfect. Mistakes could be avoided completely only if the cell could evolve mechanisms with infinite energy differences between alternatives. Since this is not possible, cells are forced to tolerate a certain level of failure and have instead evolved a variety of repair reactions to correct those errors that are the most damaging.

On the other hand, errors are essential to life as we know it. If it were not for occasional mistakes in the maintenance of DNA sequences, evolution could not occur.

Table 3–3 The Relationship Between Free-Energy Differences and Equilibrium Constants

Equilibrium Constant $\frac{[AB]}{[A][B]} = K$ (liters/mole)	Free Energy of AB minus Free Energy of A + B (kcal/mole)
10^5	−7.1
10^4	−5.7
10^3	−4.3
10^2	−2.8
10	−1.4
1	0
10^{-1}	1.4
10^{-2}	2.8
10^{-3}	4.3
10^{-4}	5.7
10^{-5}	7.1

If the reaction A + B⇌AB is allowed to come to equilibrium, the relative amounts of A, B, and AB will depend on the free-energy difference, $\Delta G°$, between them. The above values are given for 37°C (310°K) and are calculated from the equation

$$\Delta G° = -RT \ln \frac{[AB]}{[A][B]}$$

or

$$\frac{[AB]}{[A][B]} = e^{-\Delta G°/RT} = e^{-\Delta G°/(1.623)}$$

Here $\Delta G°$ is in kilocalories per mole and represents the free-energy difference under standard conditions (where all components are present at a concentration of 1.0 mole/liter); T is the temperature in kelvins (°K).

Similar principles apply to the even simpler case of a reaction A⇌A*, where a molecule is interconvertible between two states A and A* differing in free energy by an amount $\Delta G°$. The quantities of molecules in the two states at equilibrium will be in the ratio

$$\frac{[A^*]}{[A]} = e^{-\Delta G°/RT}$$

Thus, we see from this table that, if there is a favorable free-energy change of 4.3 kcal/mole for the transition A → A*, there will be 1000 times more molecules in state A* than state A.

Sarepta Exhibit 1040, page 14

Summary

The sequence of subunits in a macromolecule contains information that determines the three-dimensional contours of its surface. These contours in turn govern the recognition between one molecule and another, or between different parts of the same molecule, by means of weak, noncovalent bonds. The attractive forces are of four types: ionic bonds, van der Waals attractions, hydrogen bonds, and an interaction between nonpolar groups caused by their hydrophobic expulsion from water. Two molecules will recognize each other by a process in which they meet by random diffusion, stick together for a while, and then dissociate. The strength of this interaction is generally expressed in terms of an equilibrium constant. Since the only way to make recognition infallible is to make the energy of binding infinitely large, living cells constantly make errors; those that are intolerable are corrected by specific repair processes.

Nucleic Acids [8]

Genes Are Made of DNA [9]

It has been obvious for as long as humans have sown crops or raised animals that each seed or fertilized egg must contain a hidden plan, or design, for the development of the organism. In modern times the science of genetics grew up around the premise of invisible information-containing elements, called **genes,** that are distributed to each daughter cell when a cell divides. Therefore, before dividing, a cell has to make a copy of its genes in order to give a complete set to each daughter cell. The genes in the sperm and egg cells carry the hereditary information from one generation to the next.

The inheritance of biological characteristics must involve patterns of atoms that follow the laws of physics and chemistry: in other words, genes must be formed from molecules. At first the nature of these molecules was hard to imagine. What kind of molecule could be stored in a cell and direct the activities of a developing organism and also be capable of accurate and almost unlimited replication?

By the end of the nineteenth century biologists had recognized that the carriers of inherited information were the *chromosomes* that become visible in the nucleus as a cell begins to divide. But the evidence that the deoxyribonucleic acid (DNA) in these chromosomes is the substance of which genes are made came only much later, from studies on bacteria. In 1944 it was shown that adding purified DNA from one strain of bacteria to a second, slightly different bacterial strain conferred heritable properties characteristic of the first strain upon the second. Because it had been commonly believed that only proteins have enough conformational complexity to carry genetic information, this discovery came as a surprise, and it was not generally accepted until the early 1950s. Today the idea that DNA carries genetic information in its long chain of nucleotides is so fundamental to biological thought that it is sometimes difficult to realize the enormous intellectual gap that it filled.

DNA Molecules Consist of Two Long Chains Held Together by Complementary Base Pairs [10]

The difficulty that geneticists had in accepting DNA as the substance of genes is understandable, considering the simplicity of its chemistry. A DNA chain is a long, unbranched polymer composed of only four types of subunits. These are the deoxyribonucleotides containing the bases adenine (A), cytosine (C), guanine (G), and thymine (T). The nucleotides are linked together by covalent phosphodiester bonds that join the 5′ carbon of one deoxyribose group to the 3′ carbon of the next (see Panel 2–6, pp. 58–59). The four kinds of bases are attached to this

Sarepta Exhibit 1040, page 15

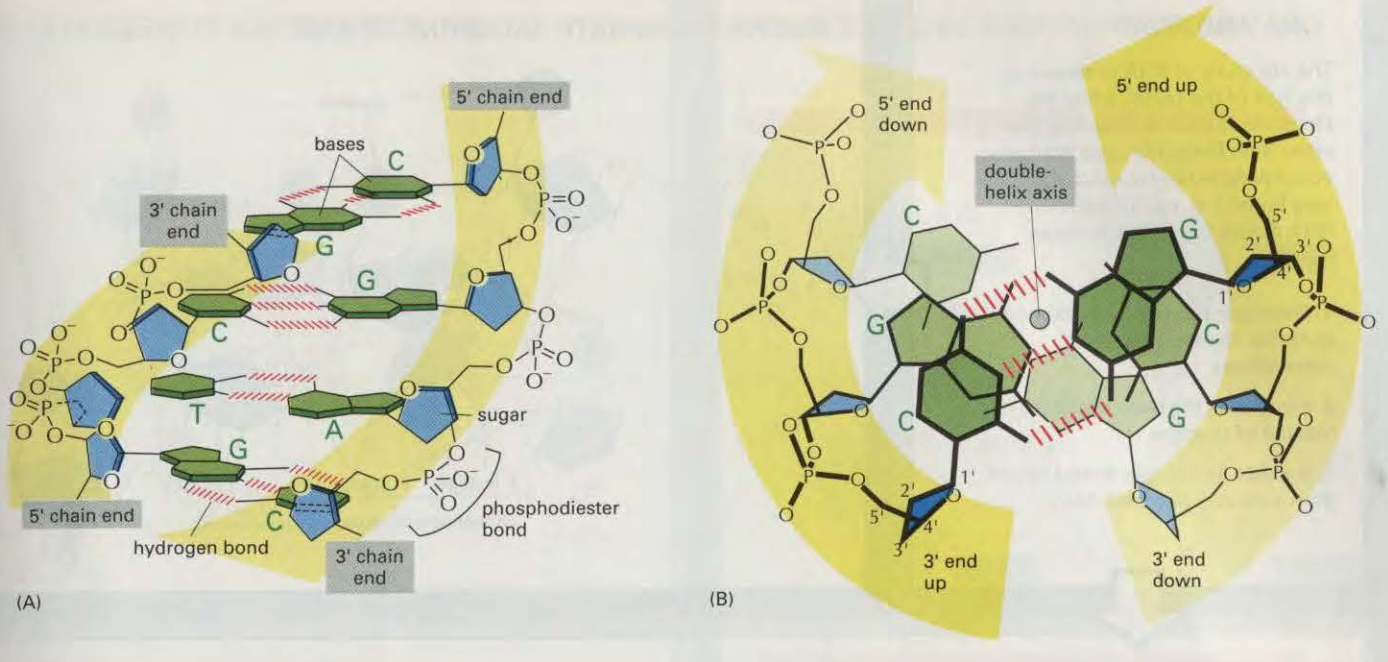


Figure 3–10 The DNA double helix. (A) A short section of the helix viewed from its side. Four complementary base pairs are shown. The bases are shown in *green*, while the deoxyribose sugars are *blue*. (B) The helix viewed from an end. Note that the two DNA strands run in opposite directions and that each base pair is held together by either two or three hydrogen bonds (see also Panel 3–2, pp. 100–101).

repetitive sugar-phosphate chain almost like four kinds of beads strung on a necklace.

How can a long chain of nucleotides encode the instructions for an organism or even a cell? And how can these messages be copied from one generation of cells to the next? The answers lie in the structure of the DNA molecule.

Early in the 1950s x-ray diffraction analyses of specimens of DNA pulled into fibers suggested that the DNA molecule is a helical polymer composed of two strands. The helical structure of DNA was not surprising since, as we have seen, a helix will often form if each of the neighboring subunits in a polymer is regularly oriented. But the finding that DNA is two-stranded was of crucial significance. It provided the clue that led, in 1953, to the construction of a model that fitted the observed x-ray diffraction pattern and thereby solved the puzzle of DNA structure and function.

An essential feature of the model was that all of the bases of the DNA molecule are on the *inside* of the double helix, with the sugar phosphates on the outside. This demands that the bases on one strand be extremely close to those on the other, and the fit proposed required specific *base-pairing* between a large purine base (A or G, each of which has a double ring) on one chain and a smaller pyrimidine base (T or C, each of which has a single ring) on the other chain (Figure 3–10).

Both evidence from earlier biochemical experiments and conclusions derived from model building suggested that **complementary base pairs** (also called *Watson-Crick base pairs*) form between A and T and between G and C. Biochemical analyses of DNA preparations from different species had shown that, although the nucleotide composition of DNA varies a great deal (for example, from 13% A residues to 36% A residues in the DNA of different types of bacteria), there is a general rule that quantitatively [G] = [C] and [A] = [T]. Model building revealed that the numbers of effective hydrogen bonds that could be formed between G and C or between A and T were greater than for any other combinations (see Panel 3–2, pp. 100–101). The double-helical model for DNA thus neatly explained the quantitative biochemistry.

The Structure of DNA Provides an Explanation for Heredity [11]

A gene carries biological information in a form that must be precisely copied and transmitted from each cell to all of its progeny. The implications of the discov-

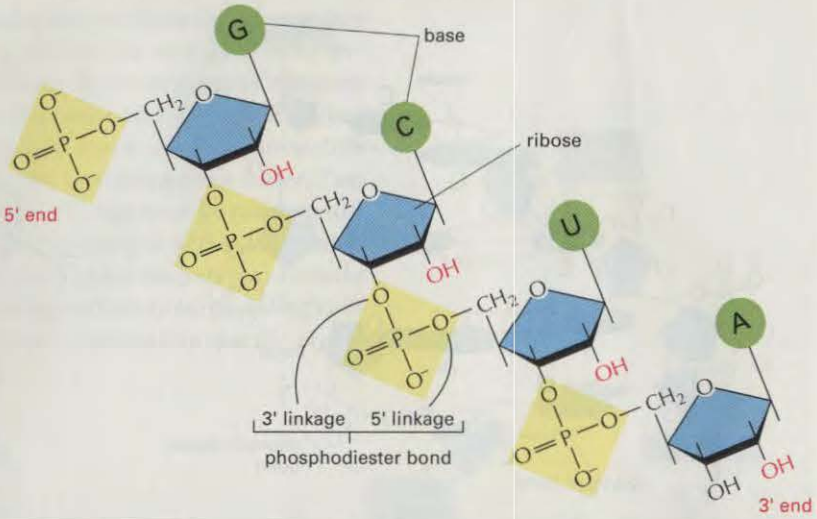
Sarepta Exhibit 1040, page 16

DNA AND RNA

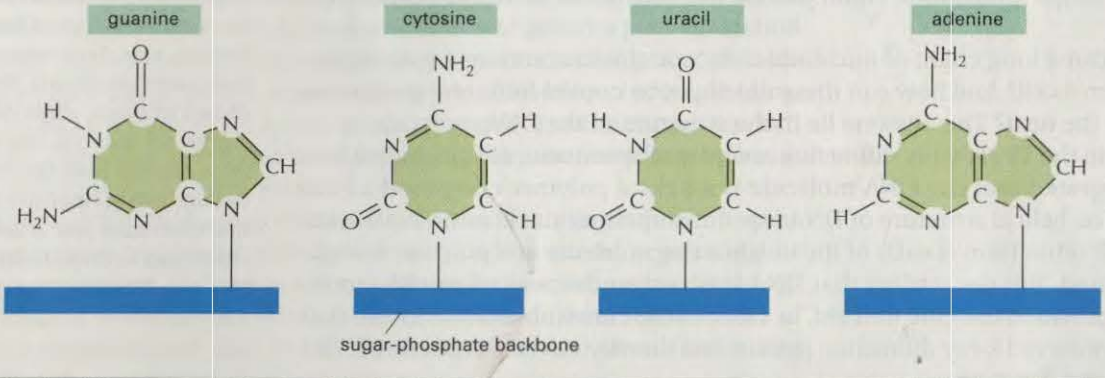
The structure of RNA is shown in this half of the panel, while the structure of DNA is shown in the other half. Both DNA and RNA are linear polymers of nucleotides (see Panel 2–6, pp. 58–59). RNA differs from DNA in three ways:

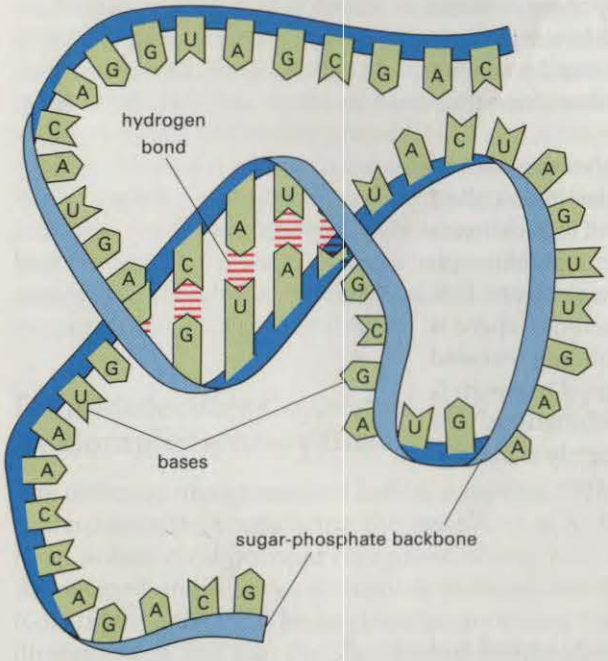
1. the sugar phosphate backbone contains ribose rather than deoxyribose
2. it contains the base uracil (U) instead of thymine (T)
3. it exists as a single strand rather than a double-stranded helix

SUGAR-PHOSPHATE BACKBONE OF RNA



FOUR BASES OF RNA

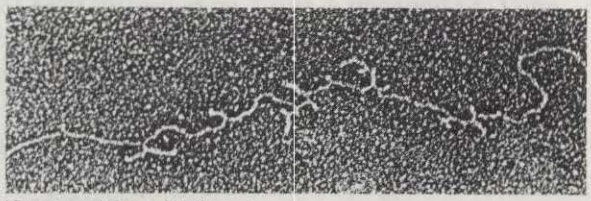




RNA SINGLE STRAND

RNA is single-stranded, but it contains local regions of short complementary base-pairing that can form from a random matching process. Regions of base-pairing can be seen in the electron micrograph as branches off the stretched-out chain.

ELECTRON MICROGRAPH OF RNA

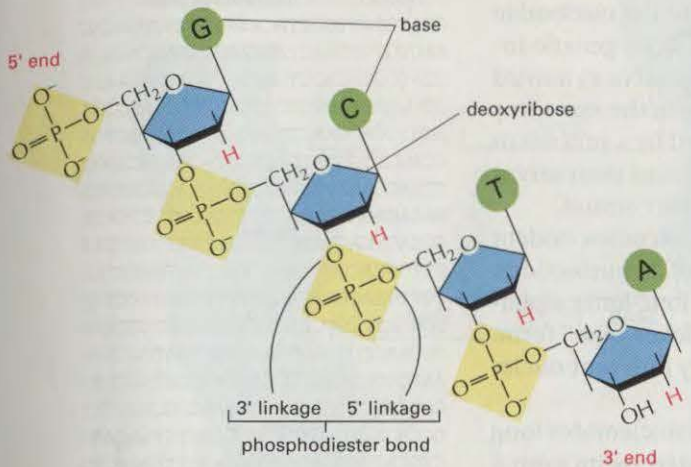


(Courtesy of Peter Wellauer.)

100 **Panel 3–2 DNA and RNA structures.**

Sarepta Exhibit 1040, page 17

SUGAR-PHOSPHATE BACKBONE OF DNA

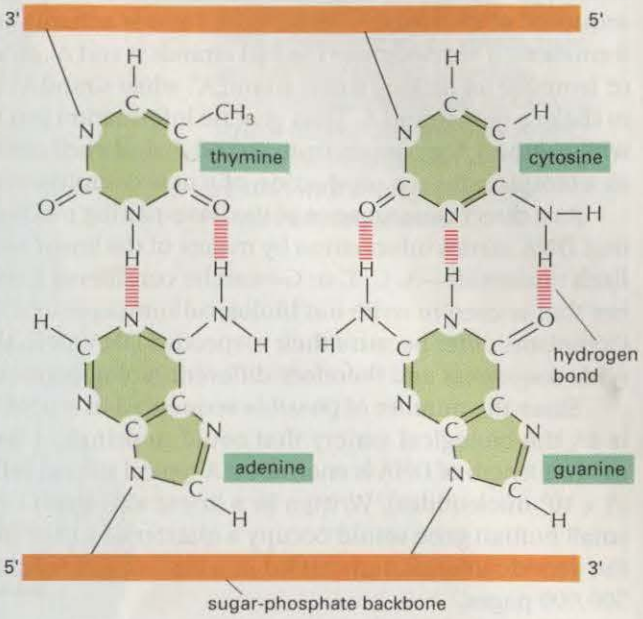


ELECTRON MICROGRAPH OF DNA



(Courtesy of Mei Lie Wong.)

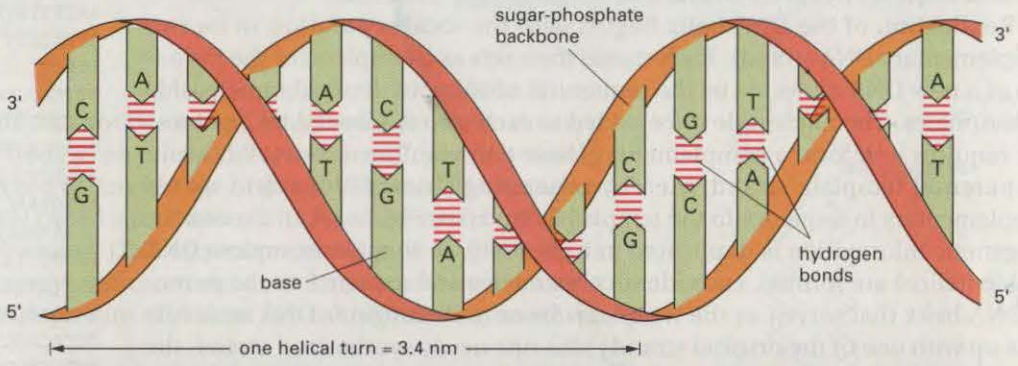
FOUR BASES AS BASE PAIRS OF DNA

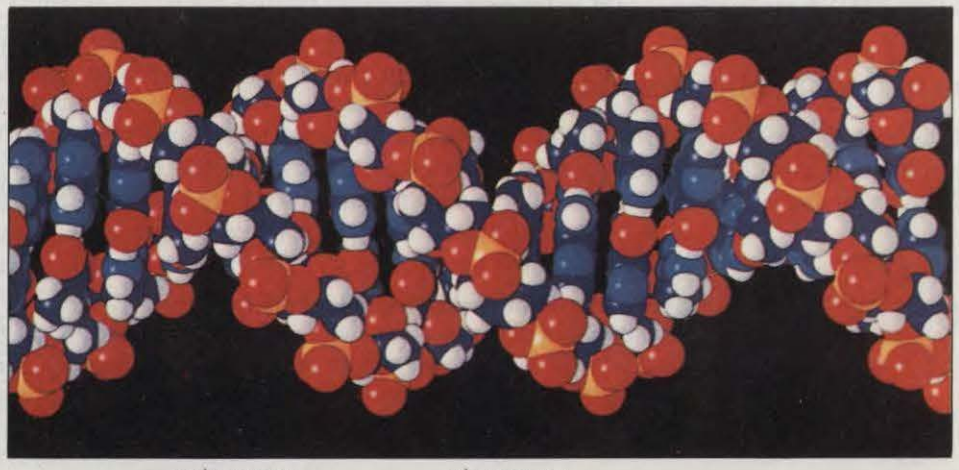


Specific hydrogen bonding between G and C and between A and T (A and U in RNA) generates complementary base-pairing.

DNA DOUBLE HELIX

In a DNA molecule two antiparallel strands that are complementary in their nucleotide sequence are paired in a right-handed double helix with about 10 nucleotide pairs per helical turn. A schematic representation (*top*) and a space-filling model (*bottom*) are illustrated here.





major groove minor groove

Sarepta Exhibit 1040, page 18

ery of the DNA double helix were profound because the structure immediately suggested how information transfer could be accomplished. Since each strand contains a nucleotide sequence that is exactly complementary to the nucleotide sequence of its partner strand, both strands actually carry the same genetic information. If we designate the two strands A and A′, strand A can serve as a mold or *template* for making a new strand A′, while strand A′ can serve in the same way to make a new strand A. Thus genetic information can be copied by a process in which strand A separates from strand A′ and each separated strand then serves as a template for the production of a new complementary partner strand.

As a direct consequence of the base-pairing mechanism, it becomes evident that DNA carries information by means of the linear sequence of its nucleotides. Each nucleotide—A, C, T, or G—can be considered a letter in a four-letter alphabet that is used to write out biological messages in a linear "ticker-tape" form. Organisms differ because their respective DNA molecules carry different nucleotide sequences and therefore different biological messages.

Since the number of possible sequences in a DNA chain n nucleotides long is 4^n, the biological variety that could in principle be generated using even a modest length of DNA is enormous. A typical animal cell contains a meter of DNA (3×10^9 nucleotides). Written in a linear alphabet of four letters, an unusually small human gene would occupy a quarter of a page of text (Figure 3–11), while the genetic information carried in a human cell would fill a book of more than 500,000 pages.

Although the principle underlying gene replication is both elegant and simple, the actual machinery by which this copying is carried out in the cell is complicated and involves a complex of proteins that form a "replication machine." The fundamental reaction is that shown in Figure 3–12, in which the enzyme *DNA polymerase* catalyzes the addition of a deoxyribonucleotide to the 3′ end of a DNA chain. Each nucleotide added to the chain is a *deoxyribonucleoside triphosphate;* the release of pyrophosphate from this activated nucleotide and its subsequent hydrolysis provide the energy for the **DNA replication** reaction and make it effectively irreversible.

Replication of the DNA helix begins with the local separation of its two complementary DNA strands. Each strand then acts as a template for the formation of a new DNA molecule by the sequential addition of deoxyribonucleoside triphosphates. The nucleotide to be added at each step is selected by a process that requires it to form a complementary base pair with the next nucleotide in the parental template strand, thereby generating a new DNA strand that is complementary in sequence to the template strand (see Figure 3–12). Eventually, the genetic information is duplicated in its entirety, so that two complete DNA double helices are formed, each identical in nucleotide sequence to the parental DNA helix that served as the template. Since each daughter DNA molecule ends up with one of the original strands plus one newly synthesized strand, the mechanism of DNA replication is said to be *semiconservative* (Figure 3–13).

Errors in DNA Replication Cause Mutations [12]

One of the most impressive features of DNA replication is its accuracy. Several proofreading mechanisms are used to eliminate incorrectly positioned nucleotides; as a result, the sequence of nucleotides in a DNA molecule is copied with

CCCTGTGGAGCCACACCCTAGGGTTGGCCA
ATCTACTCCCAGGAGCAGGGAGGGCAGGAG
CCAGGGCTGGGCATAAAAGTCAGGGCAGAG
CCATCTATTGCTTACATTTGCTTCTGACAC
AACTGTGTTCACTAGCAACTCAAACAGACA
CCATGGTGCACCTGACTCCTGAGGAGAAGT
CTGCCGTTACTGCCCTGTGGGGCAAGGTGA
ACGTGGATGAAGTTGGTGGTGAGGCCCTGG
GCAGGTTGGTATCAAGGTTACAAGACAGGT
TTAAGGAGACCAATAGAAACTGGGCATGTG
GAGACAGAGAAGACTCTTGGGTTTCTGATA
GGCACTGACTCTCTCTGCCTATTGGTCTAT
TTTCCCACCCTTAGGCTGCTGGTGGTCTAC
CCTTGGACCCAGAGGTTCTTTGAGTCCTTT
GGGGATCTGTCCACTCCTGATGCTGTTATG
GGCAACCCTAAGGTGAAGGCTCATGGCAAG
AAAGTGCTCGGTGCCTTTAGTGATGGCCTG
GCTCACCTGGACAACCTCAAGGGCACCTTT
GCCACACTGAGTGAGCTGCACTGTGACAAG
CTGCACGTGGATCCTGAGAACTTCAGGGTG
AGTCTATGGGACCCTTGATGTTTTCTTTCC
CCTTCTTTTCTATGGTTAAGTTCATGTCAT
AGGAAGGGGAGAAGTAACAGGGTACAGTTT
AGAATGGGAAACAGACGAATGATTGCATCA
GTGTGGAAGTCTCAGGATCGTTTTAGTTTC
TTTTATTTGCTGTTCATAACAATTGTTTTC
TTTTGTTTAATTCTTGCTTTCTTTTTTTTT
CTTCTCCGCAATTTTTACTATTATACTTAA
TGCCTTAACATTGTGTATAACAAAAGGAAA
TATCTCTGAGATACATTAAGTAACTTAAAA
AAAAACTTTACACAGTCTGCCTAGTACATT
ACTATTTGGAATATATGTGTGCTTATTTGC
ATATTCATAATCTCCCTACTTTATTTTCTT
TTATTTTTAATTGATACATAATCATTATAC
ATATTTATGGGTTAAAGTGTAATGTTTTAA
TATGTGTACACATATTGACCAAATCAGGGT
AATTTTGCATTTGTAATTTTAAAAAATGCT
TTCTTCTTTTAATATACTTTTTTGTTTATC
TTATTTCTAATACTTTCCCTAATCTCTTTC
TTTCAGGGCAATAATGATACAATGTATCAT
GCCTCTTTGCACCATTCTAAAGAATAACAG
TGATAATTTCTGGGTTAAGGCAATAGCAAT
ATTTCTGCATATAAATATTTCTGCATATAA
ATTGTAACTGATGTAAGAGGTTTCATATTG
CTAATAGCAGCTACAATCCAGCTACCATTC
TGCTTTTATTTTATGGTTGGGATAAGGCTG
GATTATTCTGAGTCCAAGCTAGGCCCTTTT
GCTAATCATGTTCATACCTCTTATCTTCCT
CCCACAGCTCCTGGGCAACGTGCTGGTCTG
TGTGCTGGCCCATCACTTTGGCAAAGAATT
CACCCCACCAGTGCAGGCTGCCTATCAGAA
AGTGGTGGCTGGTGTGGCTAATGCCCTGGC
CCACAAGTATCACTAAGCTCGCTTTCTTGC
TGTCCAATTTCTATTAAAGGTTCCTTTGTT
CCCTAAGTCCAACTACTAAACTGGGGGATA
TTATGAAGGGCCTTGAGCATCTGGATTCTG
CCTAATAAAAAACATTTATTTTCATTGCAA
TGATGTATTTAAATTATTTCTGAATATTTT
ACTAAAAAGGGAATGTGGGAGGTCAGTGCA
TTTAAAACATAAAGAAATGATGAGCTGTTC
AAACCTTGGGAAAATACACTATATCTTAAA
CTCCATGAAAGAAGGTGAGGCTGCAACCAG
CTAATGCACATTGGCAACAGCCCCTGATGC
CTATGCCTTATTCATCCCTCAGAAAAGGAT
TCTTGTAGAGGCTTGATTTGCAGGTTAAAG
TTTTGCTATGCTGTATTTTACATTACTTAT
TGTTTTAGCTGTCCTCATGAATGTCTTTTC

Figure 3–11 The DNA sequence of the human β-globin gene. The gene encodes one of the two subunits of the hemoglobin molecule, which carries oxygen in the blood. Only one of the two DNA strands is shown (the "coding strand"), since the other strand has a precisely complementary sequence. The sequence should be read from left to right in successive lines down the page, as if it were normal English text.

Sarepta Exhibit 1040, page 19

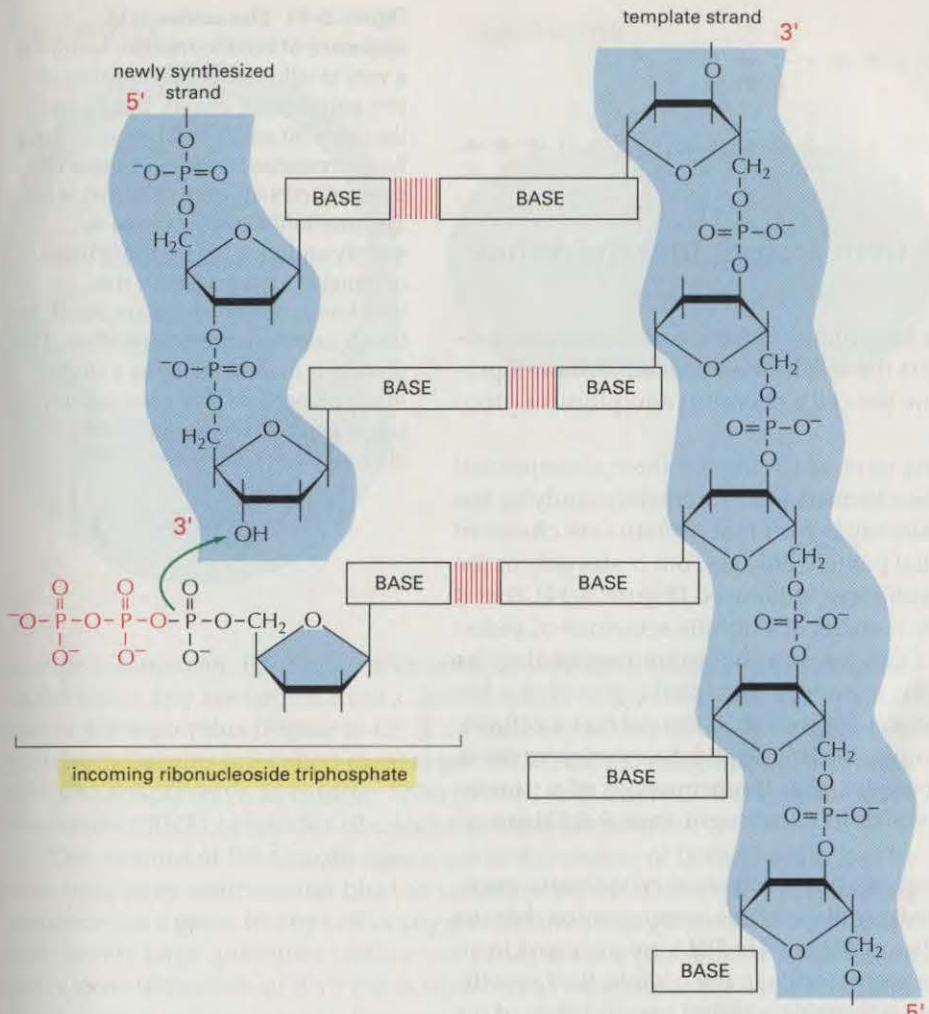


Figure 3–12 DNA synthesis. The addition of a deoxyribonucleotide to the 3′ end of a polynucleotide chain is the fundamental reaction by which DNA is synthesized. As shown, base-pairing between this incoming deoxyribonucleotide and an existing strand of DNA (the *template* strand) guides the formation of a new strand of DNA with a complementary nucleotide sequence.

fewer than one mistake in 10^9 nucleotides added. Very rarely, however, the replication machinery skips or adds a few nucleotides, or puts a T where it should have put a C, or an A instead of a G. Any change of this kind in the DNA sequence constitutes a genetic mistake, called a **mutation,** which will be copied in all future cell generations since "wrong" DNA sequences are copied as faithfully as "correct" ones. The consequence of such an error can be great, for even a single nucleotide change can have important effects on the cell, depending on where the mutation has occurred.

Geneticists demonstrated conclusively in the early 1940s that genes specify the structure of individual proteins. Thus a mutation in a gene, caused by an alteration in its DNA sequence, may lead to the inactivation of a crucial protein and result in cell death, in which case the mutation will be lost. On the other hand, a mutation may be *silent* and not affect the function of the protein. Very rarely, a mutation will create a gene with an improved or novel useful function. In this case organisms carrying the mutation will have an advantage, and the mutated gene may eventually replace the original gene in the population through natural selection.

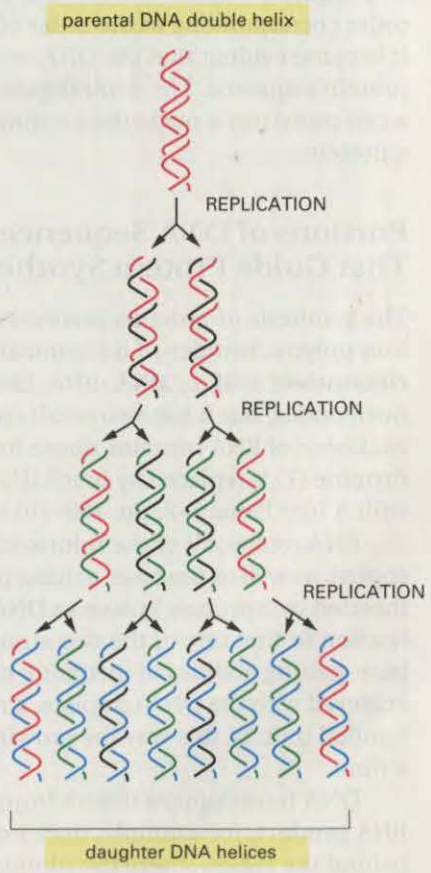


Figure 3–13 The semiconservative replication of DNA. In each round of replication each of the two strands of DNA is used as a template for the formation of a complementary DNA strand. The original strands therefore remain intact through many cell generations.

Sarepta Exhibit 1040, page 20

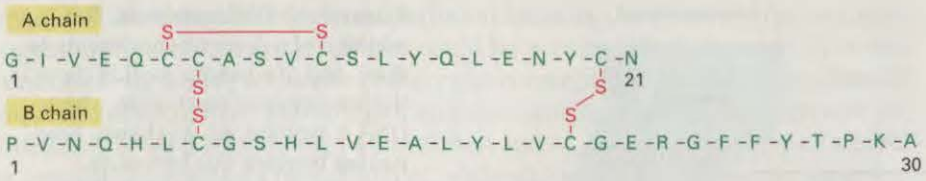


Figure 3–14 The amino acid sequence of bovine insulin. Insulin is a very small protein that consists of two polypeptide chains, one 21 and the other 30 amino acid residues long. Each chain has a unique, genetically determined sequence of amino acids. The one-letter symbols used to specify amino acids are those listed in Panel 2–5, pages 56–57; the S—S bonds shown in *red* are disulfide bonds between cysteine residues. The protein is made initially as a single long polypeptide chain (encoded by a single gene) that is subsequently cleaved to give the two chains.

The Nucleotide Sequence of a Gene Determines the Amino Acid Sequence of a Protein [13]

DNA is relatively inert chemically. The information it contains is expressed indirectly via other molecules: DNA directs the synthesis of specific RNA and protein molecules, which in turn determine the cell's chemical and physical properties.

At about the time that biophysicists were analyzing the three-dimensional structure of DNA by x-ray diffraction, biochemists were intensively studying the chemical structure of proteins. It was already known that proteins are chains of amino acids joined together by sequential peptide linkages; but it was only in the early 1950s, when the small protein *insulin* was sequenced (Figure 3–14), that it was discovered that each type of protein consists of a unique sequence of amino acids. Just as solving the structure of DNA was seminal in understanding the molecular basis of genetics and heredity, so sequencing insulin provided a key to understanding the structure and function of proteins. If insulin had a definite, genetically determined sequence, then presumably so did every other protein. It seemed reasonable to suppose, moreover, that the properties of a protein would depend on the precise order in which its constituent amino acids are arranged.

Both DNA and protein are composed of a linear sequence of subunits; eventually, the analysis of the proteins made by mutant genes demonstrated that the two sequences are *co-linear*—that is, the nucleotides in DNA are arranged in an order corresponding to the order of the amino acids in the protein they specify. It became evident that the DNA sequence contains a coded specification of the protein sequence. The central question in molecular biology then became how a cell translates a nucleotide sequence in DNA into an amino acid sequence in a protein.

Portions of DNA Sequence Are Copied into RNA Molecules That Guide Protein Synthesis [14]

The synthesis of proteins involves copying specific regions of DNA (the *genes*) into polynucleotides of a chemically and functionally different type known as **ribonucleic acid**, or **RNA**. RNA, like DNA, is composed of a linear sequence of nucleotides, but it has two small chemical differences: (1) the sugar-phosphate backbone of RNA contains ribose instead of a deoxyribose sugar and (2) the base thymine (T) is replaced by uracil (U), a very closely related base that likewise pairs with A (see Panel 3–2, pp. 100–101).

RNA retains all of the information of the DNA sequence from which it was copied, as well as the base-pairing properties of DNA. Molecules of RNA are synthesized by a process known as **DNA transcription,** which is similar to DNA replication in that one of the two strands of DNA acts as a template on which the base-pairing abilities of incoming nucleotides are tested. When a good match is achieved with the DNA template, a ribonucleotide is incorporated as a covalently bonded unit. In this way the growing RNA chain is elongated one nucleotide at a time.

DNA transcription differs from DNA replication in a number of ways. The RNA product, for example, does not remain as a strand annealed to DNA. Just behind the region where the ribonucleotides are being added, the original DNA helix re-forms and releases the RNA chain. Thus RNA molecules are single-

Sarepta Exhibit 1040, page 21

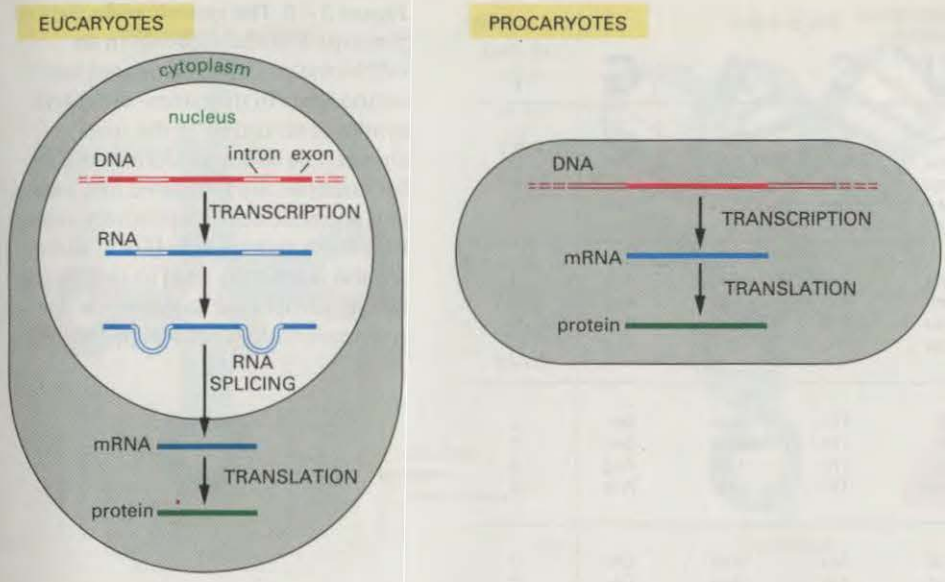


Figure 3–15 The transfer of information from DNA to protein. The transfer proceeds by means of an RNA intermediate called messenger RNA (mRNA). In procaryotic cells the process is simpler than in eucaryotic cells. In eucaryotes the coding regions of the DNA (in the *exons*, shown in color) are separated by noncoding regions (the *introns*). As indicated, these introns must be removed by an enzymatically catalyzed RNA-splicing reaction to form the mRNA.

stranded. Moreover, RNA molecules are relatively short compared to DNA molecules since they are copied from a limited region of the DNA—enough to make one or a few proteins (Figure 3–15). RNA transcripts that direct the synthesis of protein molecules are called **messenger RNA (mRNA)** molecules, while other RNA transcripts serve as *transfer RNAs (tRNAs)* or form the RNA components of ribosomes (rRNA) or smaller ribonucleoprotein particles.

The amount of RNA made from a particular region of DNA is controlled by *gene regulatory proteins* that bind to specific sites on DNA close to the coding sequences of a gene. In any cell at any given time, some genes are used to make RNA in very large quantities while other genes are not transcribed at all. For an active gene thousands of RNA transcripts can be made from the same DNA segment in each cell generation. Because each mRNA molecule can be translated into many thousands of copies of a polypeptide chain, the information contained in a small region of DNA can direct the synthesis of millions of copies of a specific protein. The protein *fibroin*, for example, is the major component of silk. In each silk gland cell a single fibroin gene makes 10^4 copies of mRNA, each of which directs the synthesis of 10^5 molecules of fibroin—producing a total of 10^9 molecules of fibroin in just 4 days.

Eucaryotic RNA Molecules Are Spliced to Remove Intron Sequences [15]

In bacterial cells most proteins are encoded by a single uninterrupted stretch of DNA sequence that is copied without alteration to produce an mRNA molecule. In 1977 molecular biologists were astonished by the discovery that most eucaryotic genes have their coding sequences (called *exons*) interrupted by noncoding sequences (called *introns*). To produce a protein, the entire length of the gene, including both its introns and its exons, is first transcribed into a very large RNA molecule—the *primary transcript*. Before this RNA molecule leaves the nucleus, a complex of RNA-processing enzymes removes all of the intron sequences, thereby producing a much shorter RNA molecule. After this RNA-processing step, called **RNA splicing,** has been completed, the RNA molecule moves to the cytoplasm as an mRNA molecule that directs the synthesis of a particular protein (see Figure 3–15).

This seemingly wasteful mode of information transfer in eucaryotes is presumed to have evolved because it makes protein synthesis much more versatile. The primary RNA transcripts of some genes, for example, can be spliced in various ways to produce different mRNAs, depending on the cell type or stage of

Sarepta Exhibit 1040, page 22

1st position (5′ end)	2nd position U	2nd position C	2nd position A	2nd position G	3rd position (3′ end)
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	STOP	STOP	A
	Leu	Ser	STOP	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

Figure 3–16 The genetic code. Sets of three nucleotides (*codons*) in an mRNA molecule are translated into amino acids in the course of protein synthesis according to the rules shown. The codons GUG and GAG, for example, are translated into valine and glutamic acid, respectively. Note that those codons with U or C as the second nucleotide tend to specify the more hydrophobic amino acids (compare with Panel 2–5, pp. 56–57).

development. This allows different proteins to be produced from the same gene. Moreover, because the presence of numerous introns facilitates genetic recombination events between exons, this type of gene arrangement is likely to have been profoundly important in the early evolutionary history of genes, speeding up the process whereby organisms evolve new proteins from parts of preexisting ones instead of evolving totally new amino acid sequences.

Sequences of Nucleotides in mRNA Are "Read" in Sets of Three and Translated into Amino Acids [16]

The rules by which the nucleotide sequence of a gene is translated into the amino acid sequence of a protein, the so-called **genetic code,** were deciphered in the early 1960s. The sequence of nucleotides in the mRNA molecule that acts as an intermediate was found to be read in serial order in groups of three. Each triplet of nucleotides, called a **codon,** specifies one amino acid. Since RNA is a linear polymer of four different nucleotides, there are $4^3 = 64$ possible codon triplets (remember that it is the *sequence* of nucleotides in the triplet that is important). However, only 20 different amino acids are commonly found in proteins, so that most amino acids are specified by several codons; that is, the genetic code is *degenerate.* The code (shown in Figure 3–16) has been highly conserved during evolution: with a few minor exceptions, it is the same in organisms as diverse as bacteria, plants, and humans.

In principle, each RNA sequence can be translated in any one of three different *reading frames* depending on where the decoding process begins (Figure 3–17). In almost every case only one of these reading frames will produce a functional protein. Since there are no punctuation signals except at the beginning and end of the RNA message, the reading frame is set at the initiation of the translation process and is maintained thereafter.

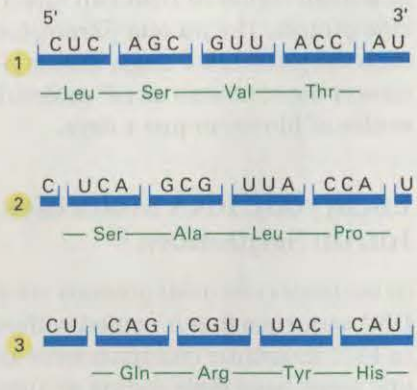
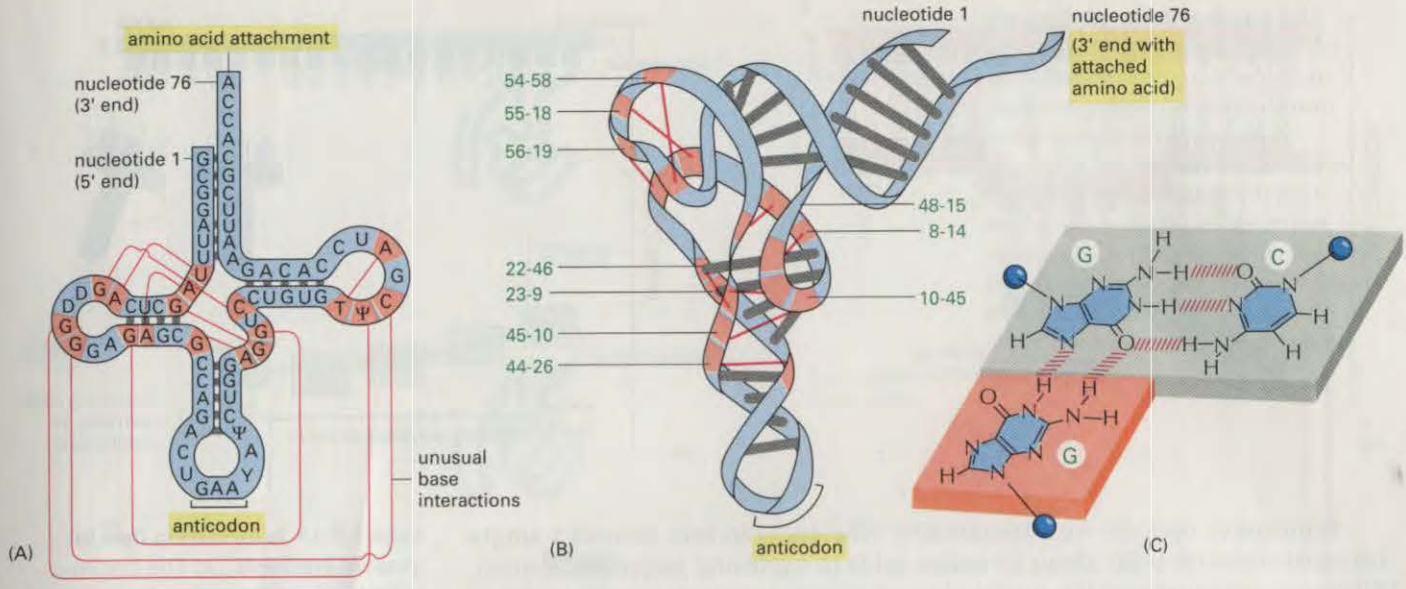


Figure 3–17 The three possible reading frames in protein synthesis. In the process of translating a nucleotide sequence (*blue*) into an amino acid sequence (*green*), the sequence of nucleotides in an mRNA molecule is read from the 5′ to the 3′ end in sequential sets of three nucleotides. In principle, therefore, the same RNA sequence can specify three completely different amino acid sequences, depending on the "reading frame."

tRNA Molecules Match Amino Acids to Groups of Nucleotides [17]

The codons in an mRNA molecule do not directly recognize the amino acids they specify in the way that an enzyme recognizes a substrate. The **translation** of mRNA into protein depends on "adaptor" molecules that recognize both an

Sarepta Exhibit 1040, page 23



amino acid and a group of three nucleotides. These adaptors consist of a set of small RNA molecules known as **transfer RNAs (tRNAs)**, each about 80 nucleotides in length.

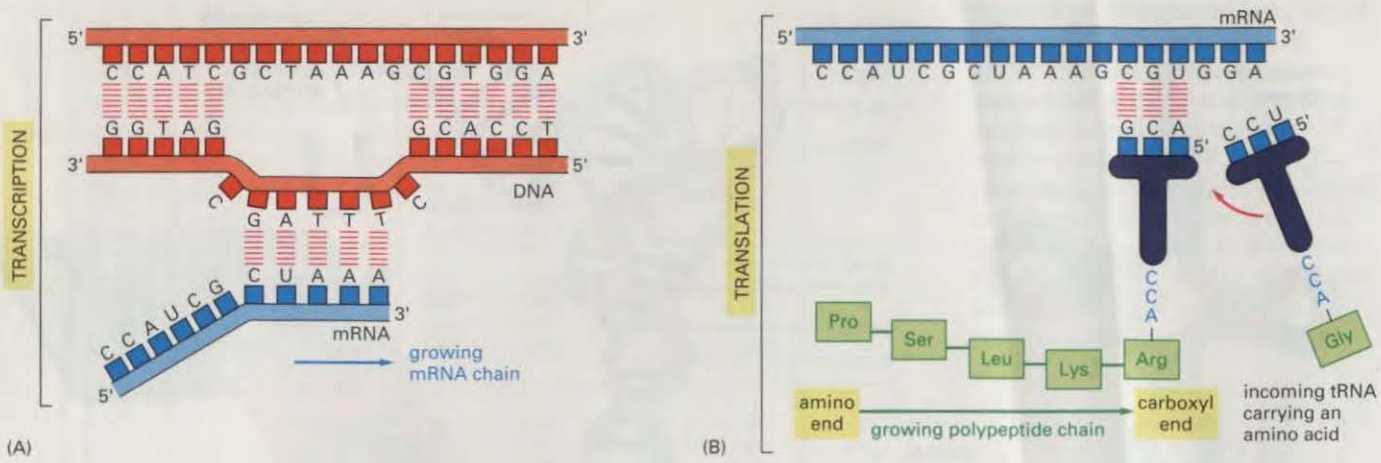
A tRNA molecule has a folded three-dimensional conformation that is held together in part by noncovalent base-pairing interactions like those that hold together the two strands of the DNA helix. In the single-stranded tRNA molecule, however, the complementary base pairs form between nucleotide residues in the *same* chain, which causes the tRNA molecule to fold up in a unique way that is important for its function as an adaptor. Four short segments of the molecule contain a double-helical structure, producing a molecule that looks like a "cloverleaf" in two dimensions. This cloverleaf is in turn further compacted into a highly folded, L-shaped conformation that is held together by more complex hydrogen-bonding interactions (Figure 3–18). Two sets of unpaired nucleotide residues at either end of the "L" are especially important for the function of the tRNA molecule in protein synthesis: one forms the *anticodon* that base-pairs to a complementary triplet in an mRNA molecule (the codon), while the *CCA sequence* at the 3′ end of the molecule is attached covalently to a specific amino acid (see Figure 3–18A).

Figure 3–18 Phenylalanine tRNA of yeast. (A) The molecule is drawn with a cloverleaf shape to show the complementary base-pairing (*short gray bars*) that occurs in the helical regions of the molecule. (B) The actual shape of the molecule, based on x-ray diffraction analysis, is shown schematically. Complementary base pairs are indicated as *long gray bars*. In addition, the nucleotides involved in unusual base-pair interactions that hold different parts of the molecule together are colored *red* and are connected by a *red line* in both (A) and (B). The pairs are numbered in (B). (C) One of the unusual base-pair interactions. Here one base forms hydrogen-bond interactions with two others; several such "base triples" help fold up this tRNA molecule.

The RNA Message Is Read from One End to the Other by a Ribosome [18]

The codon recognition process by which genetic information is transferred from mRNA via tRNA to protein depends on the same type of base-pair interactions that mediate the transfer of genetic information from DNA to DNA and from DNA to RNA (Figure 3–19). But the mechanics of ordering the tRNA molecules on the mRNA are complicated and require a **ribosome,** a complex of more than 50 different proteins associated with several structural RNA molecules (rRNAs). Each ribosome is a large protein-synthesizing machine on which tRNA molecules position themselves so as to read the genetic message encoded in an mRNA molecule. The ribosome first finds a specific start site on the mRNA that sets the reading frame and determines the amino-terminal end of the protein. Then, as the ribosome moves along the mRNA molecule, it translates the nucleotide sequence into an amino acid sequence one codon at a time, using tRNA molecules to add amino acids to the growing end of the polypeptide chain (Figure 3–20). When a ribosome reaches the end of the message, both it and the freshly made carboxyl end of the protein are released from the 3′ end of the mRNA molecule into the cytoplasm.

Sarepta Exhibit 1040, page 24



Ribosomes operate with remarkable efficiency: in one second a single bacterial ribosome adds about 20 amino acids to a growing polypeptide chain. Ribosome structure and the mechanism of protein synthesis are discussed in Chapter 6.

Figure 3–19 Information flow in protein synthesis. (A) The nucleotides in an mRNA molecule are joined together to form a complementary copy of a segment of one strand of DNA. (B) They are then matched three at a time to complementary sets of three nucleotides in the anticodon regions of tRNA molecules. At the other end of each type of tRNA molecule, a specific amino acid is held in a high-energy linkage, and when matching occurs, this amino acid is added to the end of the growing polypeptide chain. Thus translation of the mRNA nucleotide sequence into an amino acid sequence depends on complementary base-pairing between codons in the mRNA and corresponding tRNA anticodons. The molecular basis of information transfer in translation is therefore very similar to that in DNA replication and transcription. Note that the mRNA is both synthesized and translated starting from its 5′ end.

Some RNA Molecules Function as Catalysts [19]

RNA molecules have commonly been viewed as strings of nucleotides with a relatively uninteresting chemistry. In 1981 this view was shattered by the discovery of a catalytic RNA molecule with the type of sophisticated chemical reactivity that biochemists had previously associated only with proteins. The ribosomal RNA molecules of the ciliated protozoan *Tetrahymena* are initially synthesized as a large precursor from which one of the rRNAs is produced by an RNA-splicing reaction. The surprise came with the discovery that this splicing can occur *in vitro* in the absence of protein. It was subsequently shown that the intron sequence itself has an enzymelike catalytic activity that carries out the two-step reaction illustrated in Figure 3–21. The 400-nucleotide-long intron sequence was then synthesized in a test tube and shown to fold up to form a complex surface that can function like an enzyme in reactions with other RNA molecules. For example, it can bind two specific substrates tightly—a guanine nucleotide and an RNA chain—and catalyze their covalent attachment so as to sever the RNA chain at a specific site (Figure 3–22).

In this model reaction, which mimics the first step in Figure 3–21, the same intron sequence acts repeatedly to cut many RNA chains. Although RNA splicing is most commonly achieved by means that are not autocatalytic (discussed

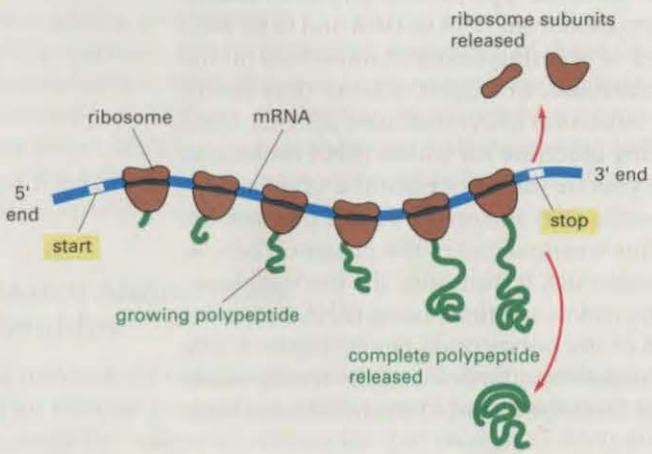


Figure 3–20 Synthesis of a protein by ribosomes attached to an mRNA molecule. Ribosomes become attached to a start signal near the 5′ end of the mRNA molecule and then move toward the 3′ end, synthesizing protein as they go. A single mRNA will usually have a number of ribosomes traveling along it at the same time, each making a separate but identical polypeptide chain; the entire structure is known as a *polyribosome*.

Sarepta Exhibit 1040, page 25

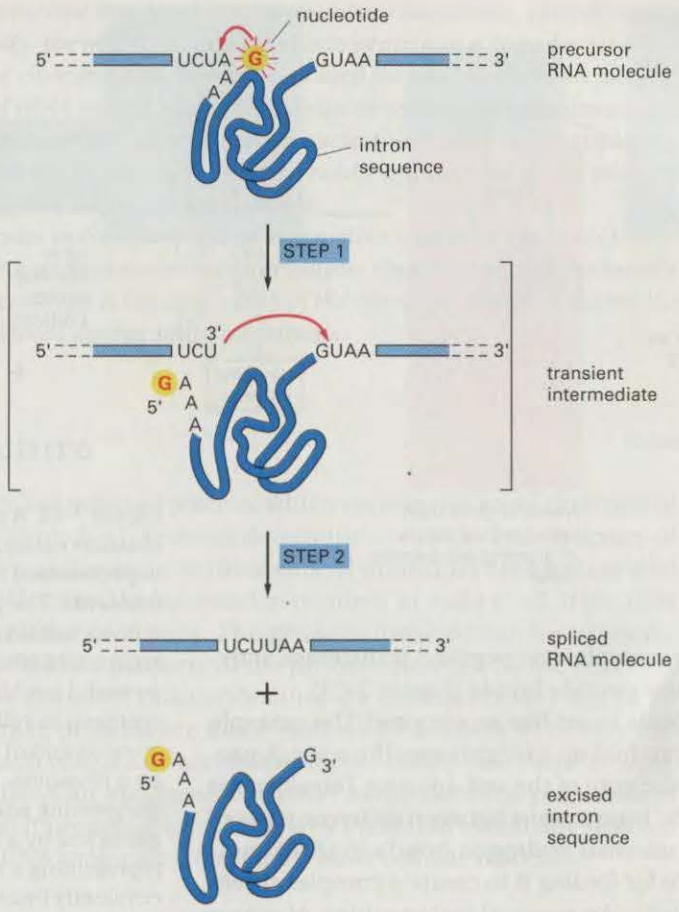


Figure 3–21 A self-splicing RNA molecule. The diagram shows the self-splicing reaction in which an intron sequence catalyzes its own excision from a *Tetrahymena* ribosomal RNA molecule. As shown, the reaction is initiated when a G nucleotide is added to the intron sequence, cleaving the RNA chain in the process; the newly created 3′ end of the RNA chain then attacks the other side of the intron to complete the reaction.

in Chapter 8), self-splicing RNAs with intron sequences related to that in *Tetrahymena* have been discovered in other types of cells, including fungi and bacteria. This suggests that these RNA sequences may have arisen before the eucaryotic and procaryotic lineages diverged about 1.5 billion years ago.

Several other families of catalytic RNAs have recently been discovered. Most tRNAs, for example, are initially synthesized as a larger precursor RNA, and an RNA molecule has been shown to play the major catalytic role in an RNA-protein complex that recognizes these precursors and cleaves them at specific sites. A catalytic RNA sequence also plays an important part in the life cycle of many plant viroids. Most remarkably, ribosomes are now suspected to function largely by RNA-based catalysis, with the ribosomal proteins playing a supporting role to the ribosomal RNAs (rRNAs), which make up more than half the mass of the

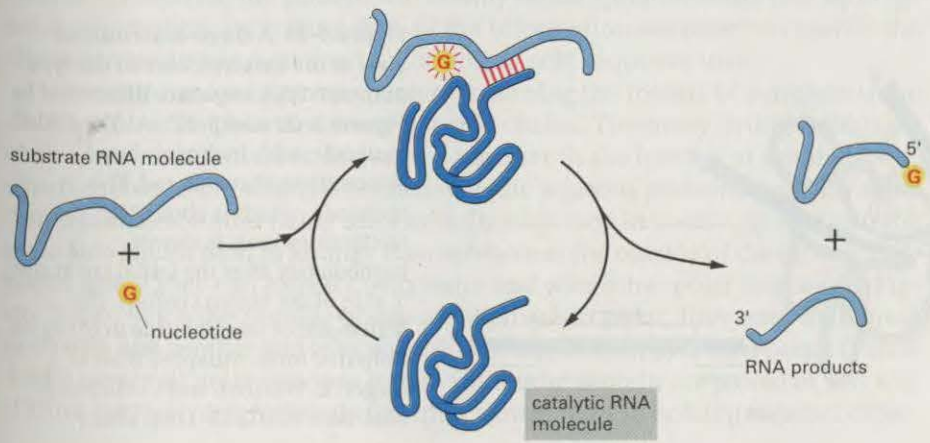


Figure 3–22 An enzymelike reaction catalyzed by the purified *Tetrahymena* intron sequence. In this reaction, which corresponds to the first step in Figure 3–21, both a specific substrate RNA molecule and a G nucleotide become tightly bound to the surface of the catalytic RNA molecule. The nucleotide is then covalently attached to the substrate RNA molecule, cleaving it at a specific site. The release of the resulting two RNA chains frees the intron sequence for further cycles of reaction.

Sarepta Exhibit 1040, page 26

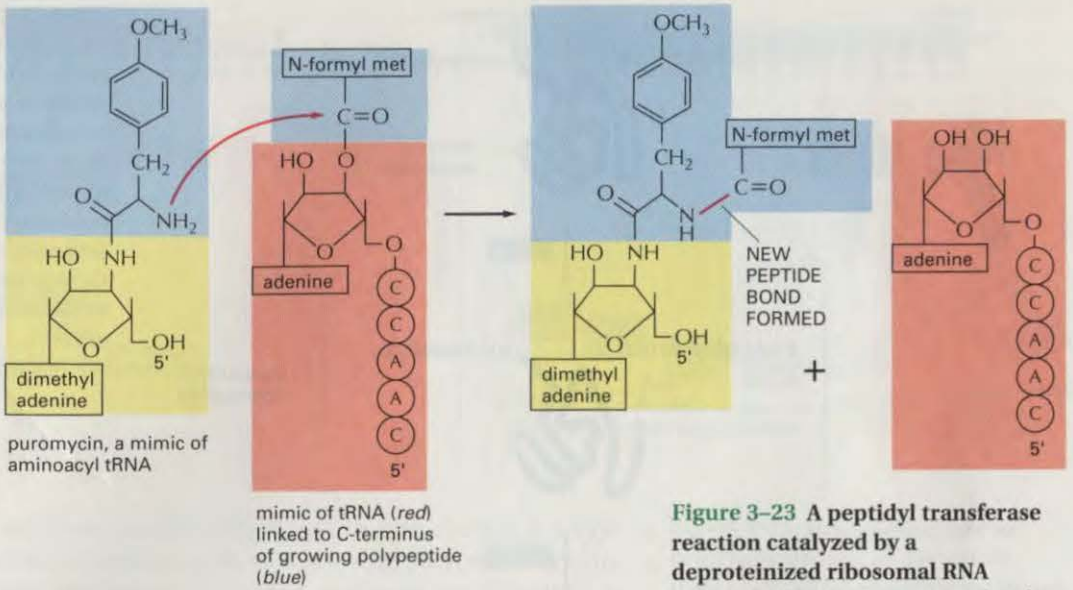


Figure 3–23 A peptidyl transferase reaction catalyzed by a deproteinized ribosomal RNA molecule. The puromycin molecule mimics a tRNA charged with the amino acid tyrosine, and it acts as a powerful inhibitor of protein synthesis in cells by adding to the growing end of a polypeptide chain on a ribosome. In this model reaction the growing polypeptide chain end is mimicked by a hexanucleotide (*red*, representing a tRNA) that is covalently linked to N-formyl methionine (representing the polypeptide). A highly purified large rRNA molecule catalyzes the addition of the puromycin to the N-formyl methionine, forming a new peptide bond and releasing the hexanucleotide.

ribosome. The large rRNA by itself, for example, has peptidyl transferase activity and will catalyze the formation of new peptide bonds (Figure 3–23).

How is it possible for an RNA molecule to act like an enzyme? The example of tRNA indicates that RNA molecules can fold up in highly specific ways. A proposed three-dimensional structure for the core of the self-splicing *Tetrahymena* intron sequence is shown in Figure 3–24. Interactions between different parts of this RNA molecule (analogous to the unusual hydrogen bonds in tRNA molecules—see Figure 3–18) are responsible for folding it to create a complex three-dimensional surface with catalytic activity. An unusual juxtaposition of atoms presumably strains covalent bonds and thereby makes selected atoms in the folded RNA chain unusually reactive.

As explained in Chapter 1, the discovery of catalytic RNA molecules has profoundly changed our views of how the first living cells arose.

Summary

Genetic information is carried in the linear sequence of nucleotides in DNA. Each molecule of DNA is a double helix formed from two complementary strands of nucleotides held together by hydrogen bonds between G-C and A-T base pairs. Duplication of the genetic information occurs by the polymerization of a new complementary strand onto each of the old strands of the double helix during DNA replication.

The expression of the genetic information stored in DNA involves the translation of a linear sequence of nucleotides into a co-linear sequence of amino acids in proteins. A limited segment of DNA is first copied into a complementary strand of RNA.

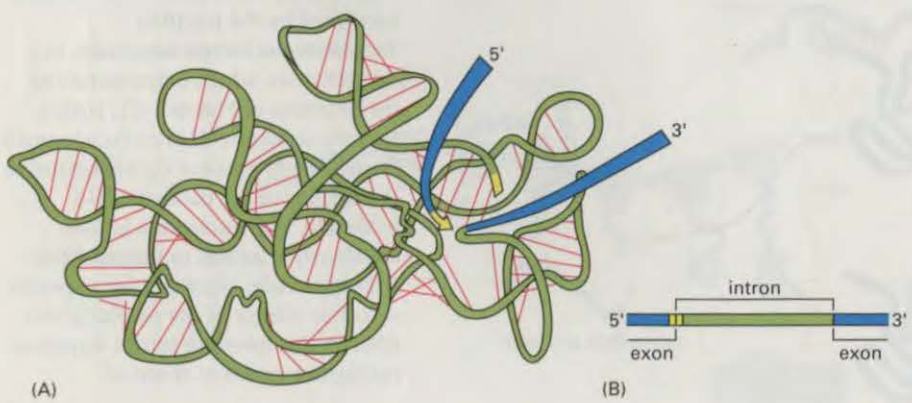


Figure 3–24 A three-dimensional view of the catalytic core of the type of intron RNA sequence illustrated in Figures 3–21 and 3–22. (A) The folded molecule, with hydrogen-bond interactions shown in *red*. This molecule, which is about 240 nucleotides long, is shown immediately after the initial cut at the 5′ side of the intron (*yellow*). (B) Schematic of the molecule in (A) in its unfolded form. (Adapted from L. Jaeger, E. Westhof, and F. Michel, *J. Mol. Biol.* 221:1153–1164, 1991.)

Sarepta Exhibit 1040, page 27

*This primary RNA transcript is spliced to remove intron sequences, producing an mRNA molecule. Finally, the mRNA is translated into protein in a complex set of reactions that occur on a ribosome. The amino acids used for protein synthesis are first attached to a family of tRNA molecules, each of which recognizes, by complementary base-pairing interactions, particular sets of three nucleotides in the mRNA (*codons*). The sequence of nucleotides in the mRNA is then read from one end to the other in sets of three, according to a universal genetic code.*

Other RNA molecules in cells function as enzymelike catalysts. These RNA molecules fold up to create a surface containing nucleotides that have become unusually reactive. One of these catalysts is the large rRNA of the ribosome, which catalyzes the formation of peptide bonds during protein synthesis.

Protein Structure [20]

To a large extent, cells are made of protein, which constitutes more than half of their dry weight (see Table 3–1). Proteins determine the shape and structure of the cell and also serve as the main instruments of molecular recognition and catalysis. Although DNA stores the information required to make a cell, it has little direct influence on cellular processes. The gene for hemoglobin, for example, cannot carry oxygen; that is a property of the protein specified by the gene.

DNA and RNA are chains of nucleotides that are chemically very similar to one another. In contrast, proteins are made from an assortment of 20 very different amino acids, each with a distinct chemical personality (see Panel 2–5, pp. 56–57). This variety allows for enormous versatility in the chemical properties of different proteins, and it presumably explains why evolution eventually selected proteins rather than RNA molecules to catalyze most cellular reactions.

The Shape of a Protein Molecule Is Determined by Its Amino Acid Sequence [21]

Many of the bonds in a long polypeptide chain allow free rotation of the atoms they join, giving the protein backbone great flexibility. In principle, then, any protein molecule could adopt an almost unlimited number of shapes (*conformations*). Most polypeptide chains, however, fold into only one particular conformation determined by their amino acid sequence. This is because the backbones and side chains of the amino acids associate with one another and with water to form various weak noncovalent bonds (see Panel 3–1, pp. 92–93). Provided that the appropriate side chains are present at crucial positions in the chain, large forces are developed that make one particular conformation especially stable.

Most proteins can fold spontaneously into their correct shape. By treatment with certain solvents, a protein can be unfolded, or *denatured*, to give a flexible polypeptide chain that has lost its native conformation. When the denaturing solvent is removed, the protein will usually refold spontaneously into its original conformation, indicating that all the information necessary to specify the shape of a protein is contained in the amino acid sequence itself.

One of the most important factors governing the folding of a protein is the distribution of its polar and nonpolar side chains. The many hydrophobic side chains in a protein tend to be pushed together in the interior of the molecule, which enables them to avoid contact with the aqueous environment (just as oil droplets coalesce after being mechanically dispersed in water). By contrast, the polar side chains tend to arrange themselves near the outside of the protein molecule, where they can interact with water and with other polar molecules (Figure 3–25). Since the peptide bonds are themselves polar, they tend to interact both with one another and with polar side chains to form hydrogen bonds (Figure 3–26); nearly all polar residues buried within the protein are paired in this way (Figure 3–27). Hydrogen bonds thus play a major part in holding together differ-

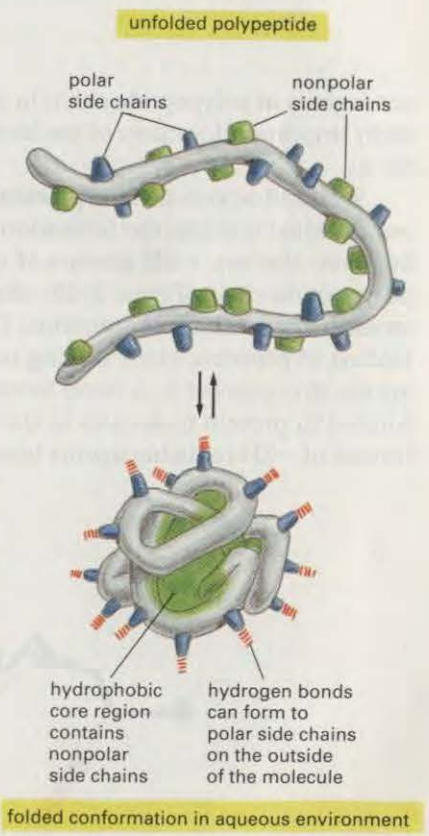


Figure 3–25 How a protein folds into a globular conformation. The polar amino acid side chains tend to gather on the outside of the protein, where they can interact with water. The nonpolar amino acid side chains are buried on the inside to form a hydrophobic core that is "hidden" from water.

Sarepta Exhibit 1040, page 28

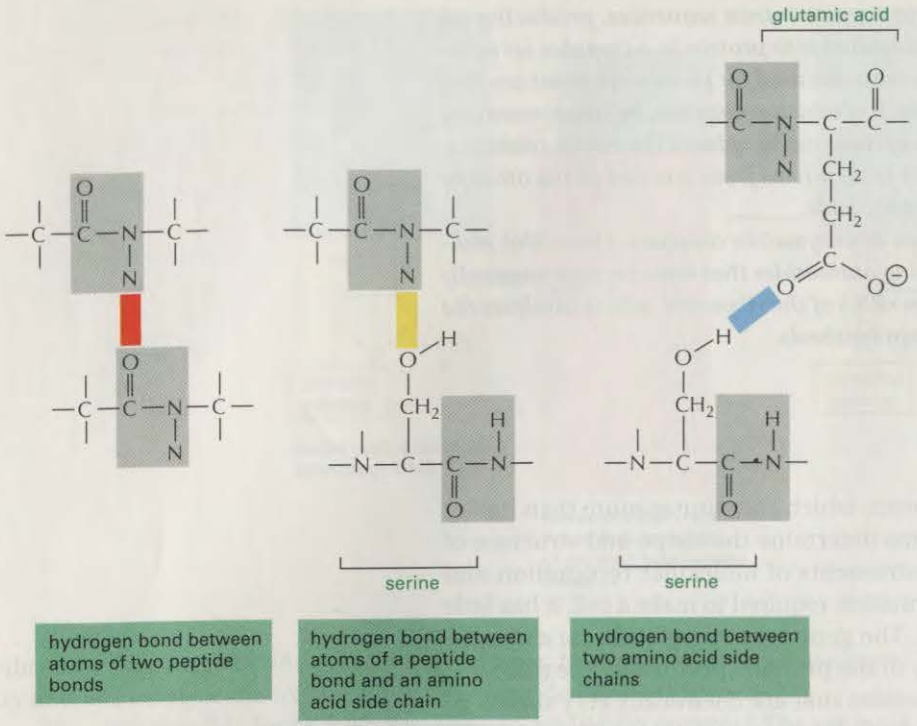


Figure 3–26 Hydrogen bonding. Some of the hydrogen bonds (shown in color) that can form between the amino acids in a protein. The peptide bonds are shaded in *gray*.

ent regions of polypeptide chain in a folded protein molecule. They are also crucially important for many of the binding interactions that occur on protein surfaces.

Secreted or cell-surface proteins often form additional *covalent* intrachain bonds. Most notably, the formation of **disulfide bonds** (also called S—S bonds) between the two —SH groups of neighboring cysteine residues in a folded polypeptide chain (Figure 3–28) often serves to stabilize the three-dimensional structure of extracellular proteins. These bonds are not required for the specific folding of proteins, since folding occurs normally in the presence of reducing agents that prevent S—S bond formation. In fact, S—S bonds are rarely, if ever, formed in protein molecules in the cytosol because the high cytosolic concentration of —SH reducing agents breaks such bonds.

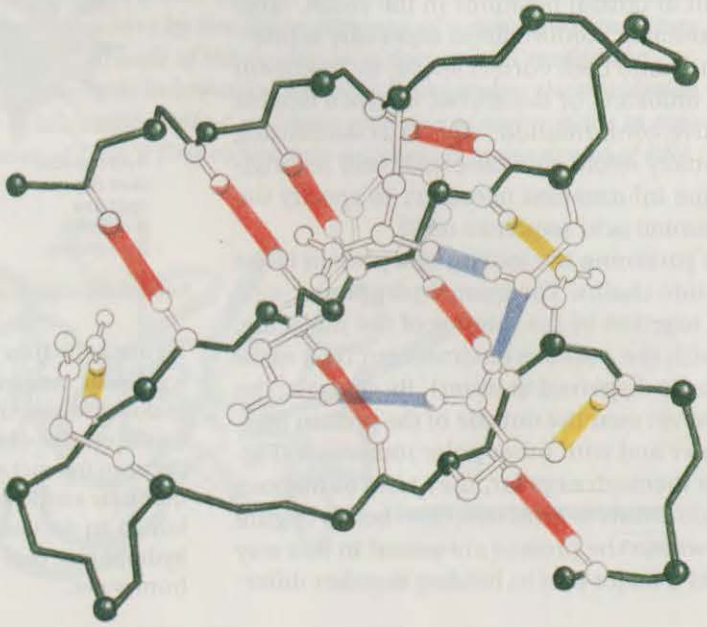


Figure 3–27 Details of intramolecular hydrogen bonds in a protein. In this region of the enzyme lysozyme, hydrogen bonds form between two side chains (*blue*), between a side chain and an atom in a peptide bond (*yellow*), or between atoms in two peptide bonds (*red*). For reference, see Figure 3–26. (After C.K. Mathews and K.E. van Holde, Biochemistry. Redwood City, CA: Benjamin/Cummings, 1990.)

Sarepta Exhibit 1040, page 29

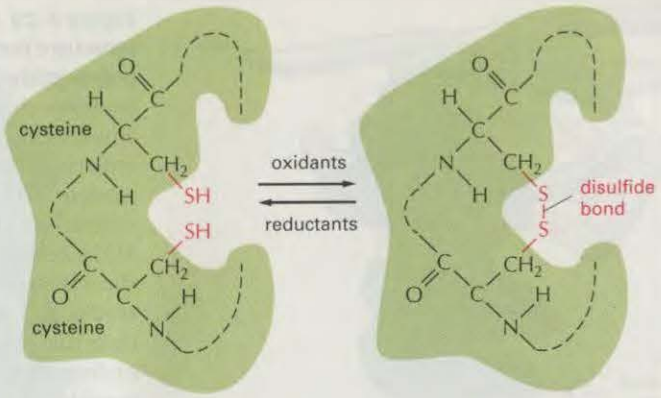


Figure 3–28 Disulfide-bond formation. The drawing illustrates the formation of a covalent disulfide bond between the side chains of neighboring cysteine residues in a protein.

The net result of all the individual amino acid interactions is that most protein molecules fold up spontaneously into precisely defined conformations. Those that are compact and globular have an inner core composed of clustered hydrophobic side chains—packed into a tight, nearly crystalline arrangement—while a very complex and irregular exterior surface is formed by the more polar side chains. The positioning and chemistry of the different atoms on this intricate surface make each protein unique and enable it to bind specifically to other macromolecular surfaces and to certain small molecules (discussed below). From both a chemical and a structural standpoint, proteins are the most sophisticated molecules known.

Common Folding Patterns Recur in Different Protein Chains [22]

Although all the information required for the folding of a protein chain is contained in its amino acid sequence, we have not yet learned how to "read" this information so as to predict the detailed three-dimensional structure of a protein whose sequence is known. Consequently, the folded conformation can be determined only by an elaborate *x-ray diffraction analysis* performed on crystals of the protein or, if the protein is very small, by nuclear magnetic resonance techniques (see Chapter 4). So far, more than 100 types of protein folds have been discovered by this technique. Each protein has a specific conformation so intricate and irregular that it would require a chapter to describe it in full three-dimensional detail.

When the three-dimensional structures of different protein molecules are compared, it becomes clear that, although the overall conformation of each protein is unique, several structural patterns recur repeatedly in parts of these macromolecules. Two patterns are particularly common because they result from regular hydrogen-bonding interactions between the peptide bonds themselves rather than between the side chains of particular amino acids. Both patterns were correctly predicted in 1951 from model-building studies based on the different x-ray diffraction patterns of silk and hair. The two regular patterns discovered are now known as the *β sheet*, which occurs in the protein fibroin, found in silk, and the *α helix*, which occurs in the protein α-keratin, found in skin and its appendages, such as hair, nails, and feathers.

The core of most (but not all) globular proteins contains extensive regions of **β sheet.** In the example illustrated in Figure 3–29, which shows part of an antibody molecule, an *antiparallel β sheet* is formed when an extended polypeptide chain folds back and forth upon itself, with each section of the chain running in the direction opposite to that of its immediate neighbors. This gives a very rigid structure held together by hydrogen bonds that connect the peptide bonds in neighboring chains. The antiparallel β sheet and the closely related *parallel β sheet* (which is formed by regions of polypeptide chain that run in the same direction) frequently serve as the framework around which globular proteins are constructed.

Sarepta Exhibit 1040, page 30

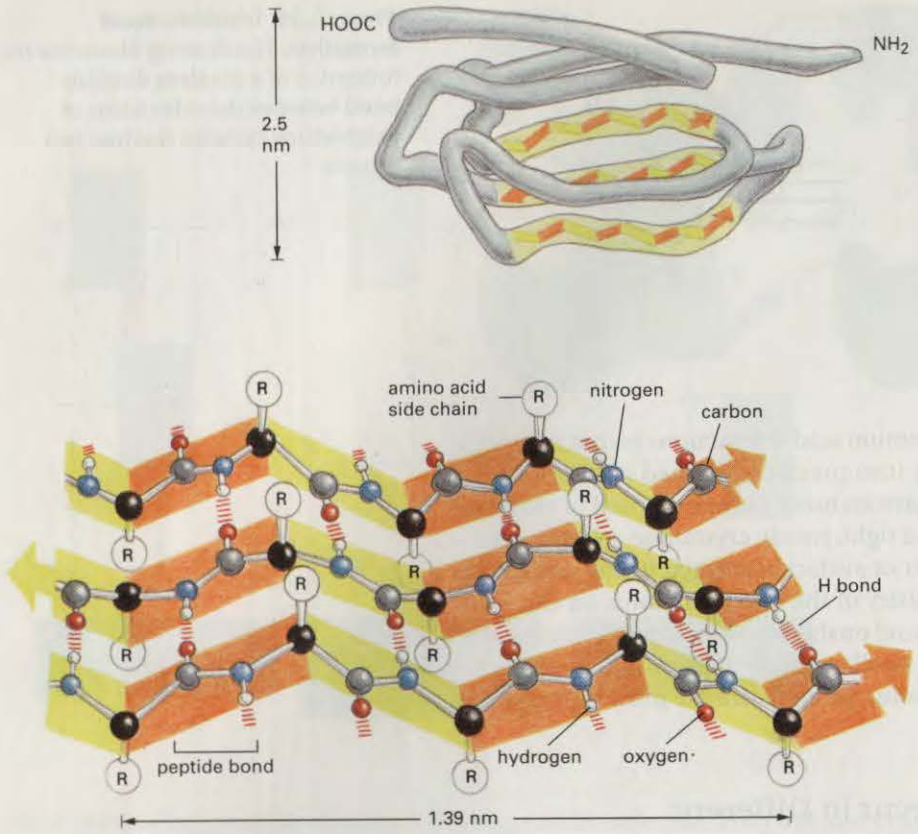
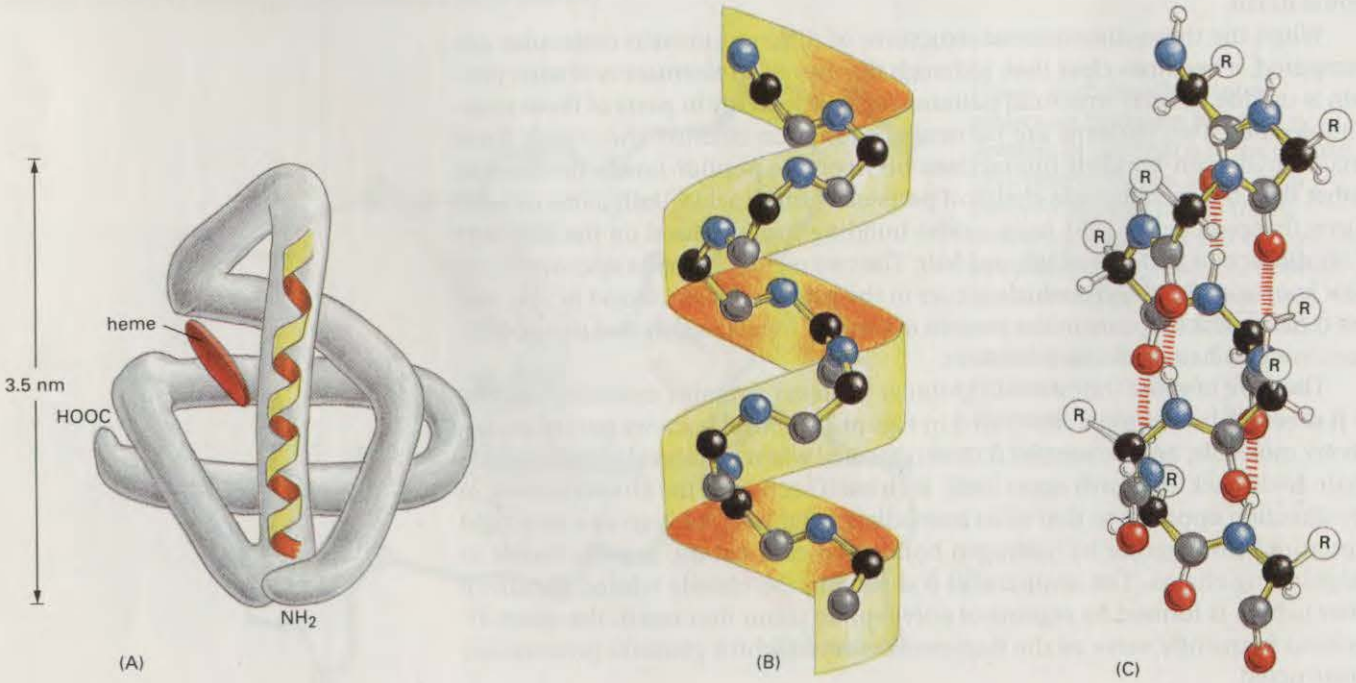


Figure 3–29 A β sheet is a common structure formed by parts of the polypeptide chain in globular proteins. At the top, a domain of 115 amino acids from an immunoglobulin molecule is shown; it consists of a sandwichlike structure of two β sheets, one of which is drawn in color. At the bottom, a perfect antiparallel β sheet is shown in detail, with the amino acid side chains denoted R. Note that every peptide bond is hydrogen-bonded to a neighboring peptide bond. The actual sheet structures in globular proteins are usually less regular than the β sheet shown here, and most sheets are slightly twisted (see Figure 3–31).

An **α helix** is generated when a single polypeptide chain turns regularly about itself to make a rigid cylinder in which each peptide bond is regularly hydrogen-bonded to other peptide bonds nearby in the chain. Many globular proteins contain short regions of such α helices (Figure 3–30), and those portions of a transmembrane protein that cross the lipid bilayer are usually α helices because of the constraints imposed by the hydrophobic lipid environment (discussed in Chapter 10).

Figure 3–30 An α helix is another common structure formed by parts of the polypeptide chain in proteins. (A) The oxygen-carrying molecule myoglobin (153 amino acids long) is shown, with one region of α helix outlined in color. (B) A perfect α helix is shown in outline. (C) As in the β sheet, every peptide bond in an α helix is hydrogen-bonded to a neighboring peptide bond. Note that for clarity in (B) both the side chains [which protrude radially along the outside of the helix and are denoted by R in (C)] and the hydrogen atom are omitted on the α-carbon atom of each amino acid (see also Figure 3–31).



Sarepta Exhibit 1040, page 31

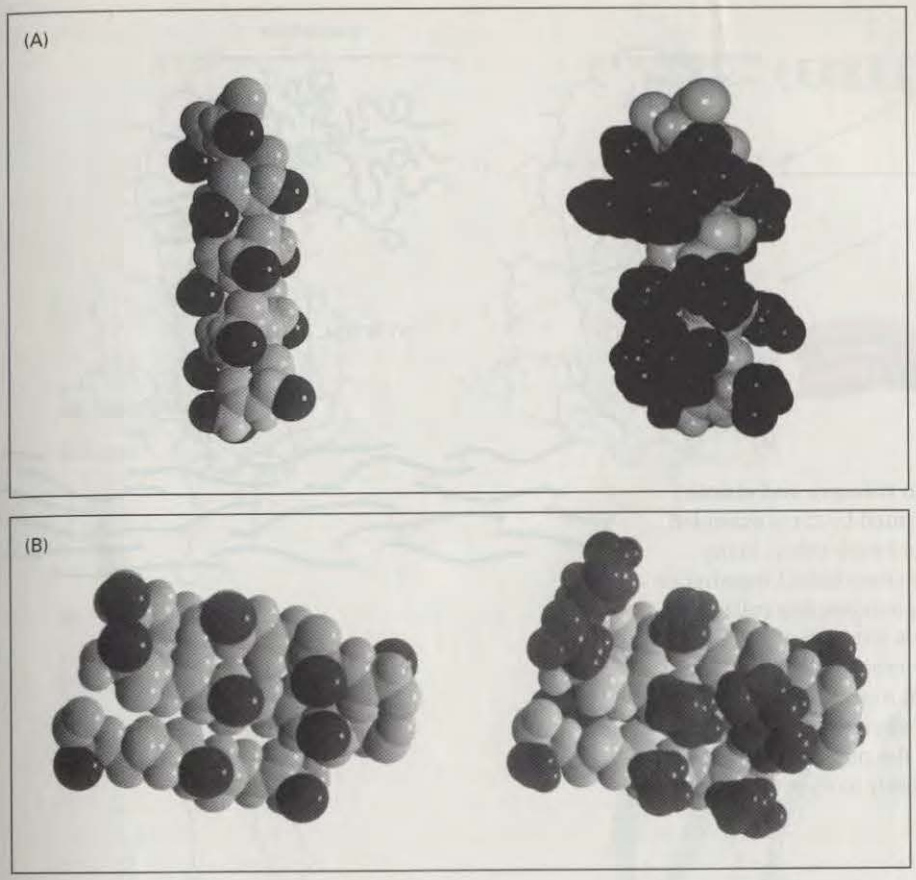


Figure 3–31 Space-filling models of an α helix and a β sheet with (*right*) and without (*left*) their amino acid side chains. (A) An α helix (part of the structure of myoglobin). (B) A region of β sheet (part of the structure of an immunoglobulin domain). In the photographs on the left, each side chain is represented by a single darkly shaded atom (the R groups in Figures 3–29 and 3–30), while the entire side chain is shown on the right. (Courtesy of Richard J. Feldmann.)

In aqueous environments an isolated α helix is usually not stable on its own. Two identical α helices that have a repeating arrangement of nonpolar side chains, however, will twist around each other gradually to form a particularly stable structure known as a *coiled-coil* (see p. 125). Long rodlike coiled-coils are found in many fibrous proteins, such as the intracellular α-keratin fibers that reinforce skin and its appendages.

Space-filling representations of an α helix and a β sheet from actual proteins are shown with and without their side chains in Figure 3–31.

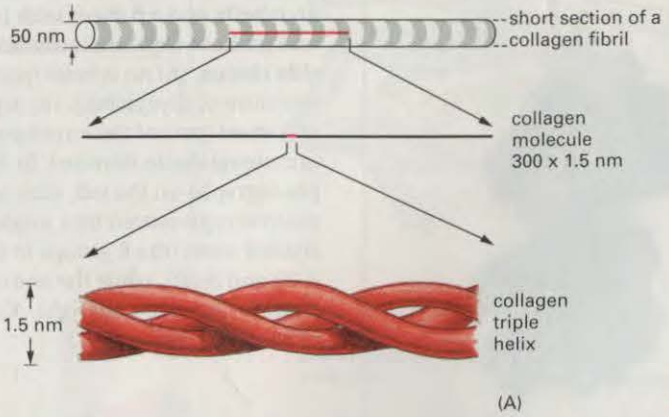
Proteins Are Amazingly Versatile Molecules [23]

Because of the variety of their amino acid side chains, proteins are remarkably versatile with respect to the types of structures they can form. Contrast, for example, two abundant proteins secreted by cells in connective tissue—collagen and elastin—both present in the extracellular matrix. In **collagen** molecules three separate polypeptide chains, each rich in the amino acid proline and containing the amino acid glycine at every third residue, are wound around one another to generate a regular triple helix. These collagen molecules are packed together into fibrils in which adjacent molecules are tied together by covalent cross-links between neighboring lysine residues, giving the fibril enormous tensile strength (Figure 3–32).

Elastin is at the opposite extreme. Its relatively loose and unstructured polypeptide chains are cross-linked covalently to generate a rubberlike elastic meshwork that enables tissues such as arteries and lungs to deform and stretch without damage. As illustrated in Figure 3–32, the elasticity is due to the ability of individual protein molecules to uncoil reversibly whenever a stretching force is applied.

It is remarkable that the same basic chemical structure—a chain of amino acids—can form so many different structures: a rubberlike elastic meshwork

Sarepta Exhibit 1040, page 32



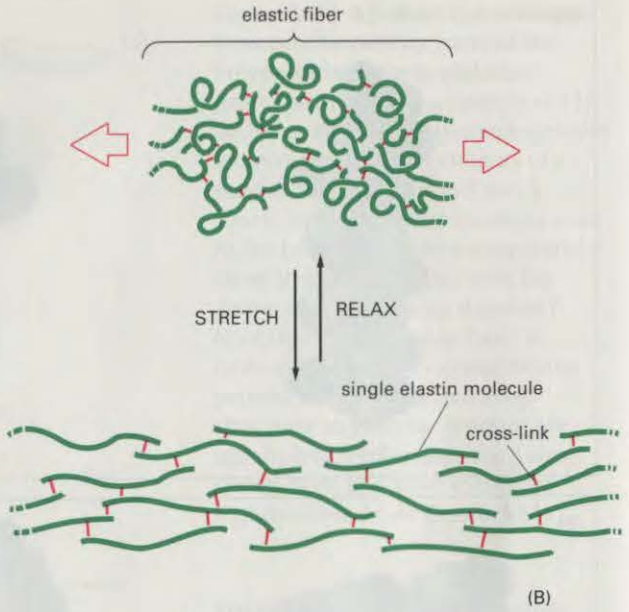


Figure 3–32 Contrast between collagen and elastin. (A) *Collagen* is a triple helix formed by three extended protein chains that wrap around each other. Many rodlike collagen molecules are cross-linked together in the extracellular space to form inextensible collagen fibrils (*top*) that have the tensile strength of steel. (B) *Elastin* polypeptide chains are cross-linked together to form elastic fibers. Each elastin molecule uncoils into a more extended conformation when the fiber is stretched. The striking contrast between the physical properties of elastin and collagen is due entirely to their very different amino acid sequences.

(elastin), an inextensible cable with the tensile strength of steel (collagen), or any of the wide variety of catalytic surfaces on the globular proteins that function as enzymes. Figure 3–33 illustrates and compares the range of shapes that could, in theory, be adopted by a polypeptide chain 300 amino acids long. As we have already emphasized, the conformation actually adopted depends on the amino acid sequence.

Proteins Have Different Levels of Structural Organization [24]

In describing the structure of a protein, it is helpful to distinguish various levels of organization. The amino acid sequence is called the **primary structure** of the protein. Regular hydrogen-bond interactions within contiguous stretches of polypeptide chain give rise to α helices and β sheets, which constitute the protein's **secondary structure.** Certain combinations of α helices and β sheets pack together to form compactly folded globular units, each of which is called a protein **domain.** Domains are usually constructed from a section of polypeptide chain that contains between 50 and 350 amino acids, and they seem to be the modular units from which proteins are constructed (see below). While small proteins may contain only a single domain, larger proteins contain a number of domains, which are often connected by relatively open lengths of polypeptide chain. Finally, individual polypeptides often serve as subunits for the formation of larger molecules, sometimes called *protein assemblies* or *protein complexes,* in which the subunits are bound to one another by a large number of weak, noncovalent interactions; in extracellular proteins these interactions are often stabilized by disulfide bonds.

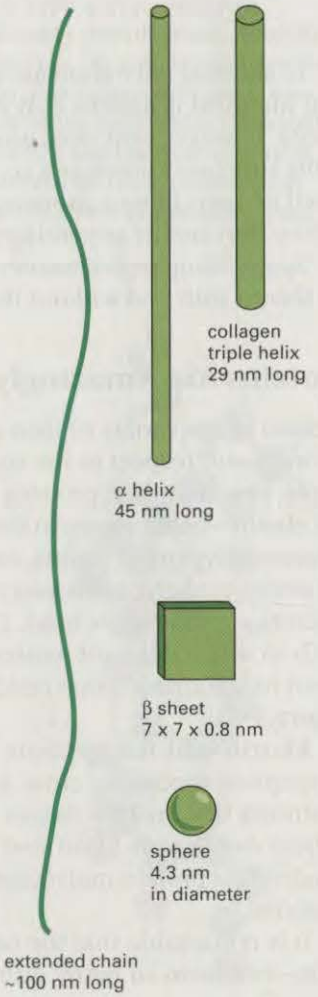


Figure 3–33 Some possible sizes and shapes of a protein molecule 300 amino acid residues long. The structure formed is determined by the amino acid sequence. (Adapted from D.E. Metzler, Biochemistry. New York: Academic Press, 1977.)

Sarepta Exhibit 1040, page 33

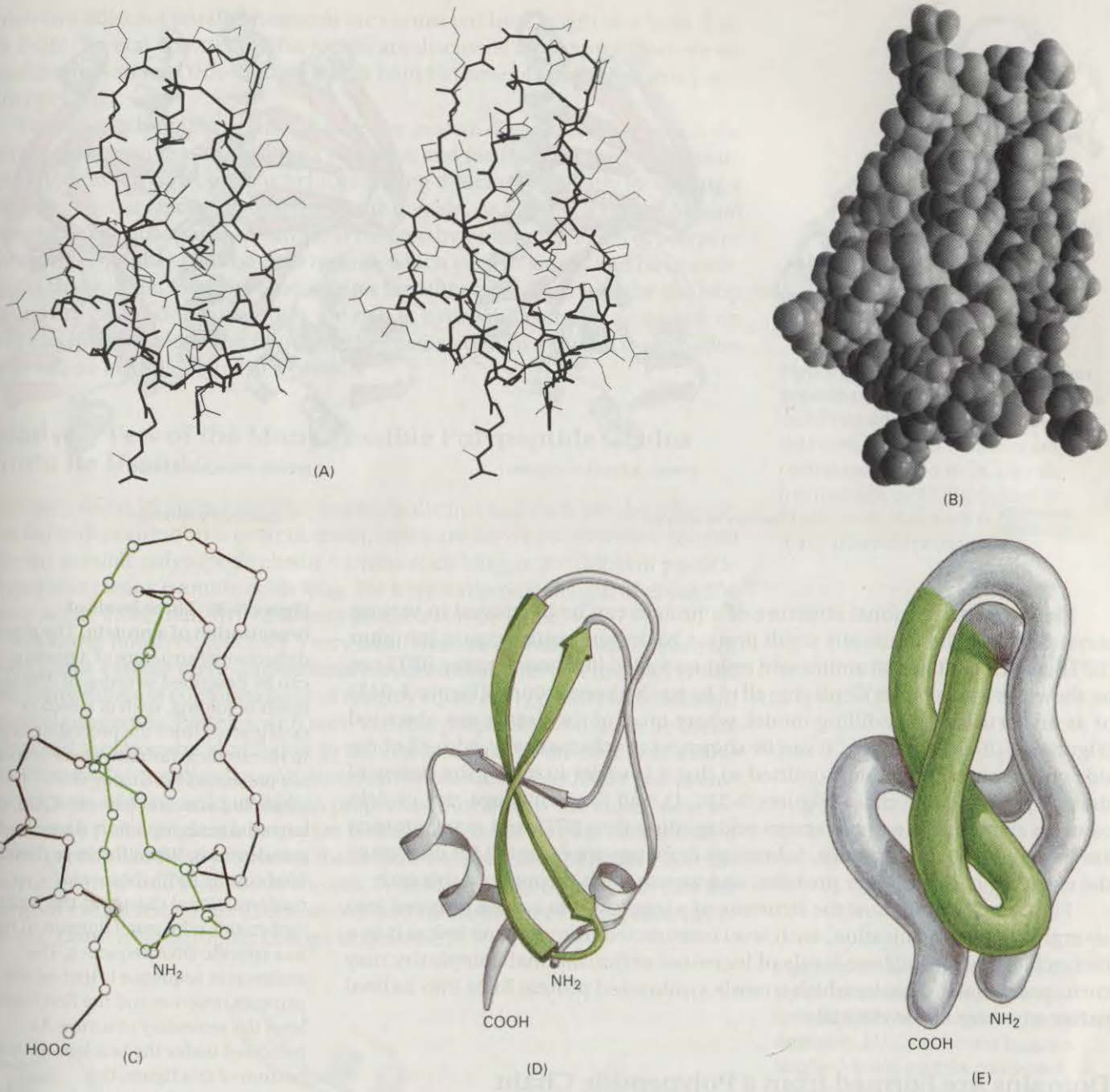
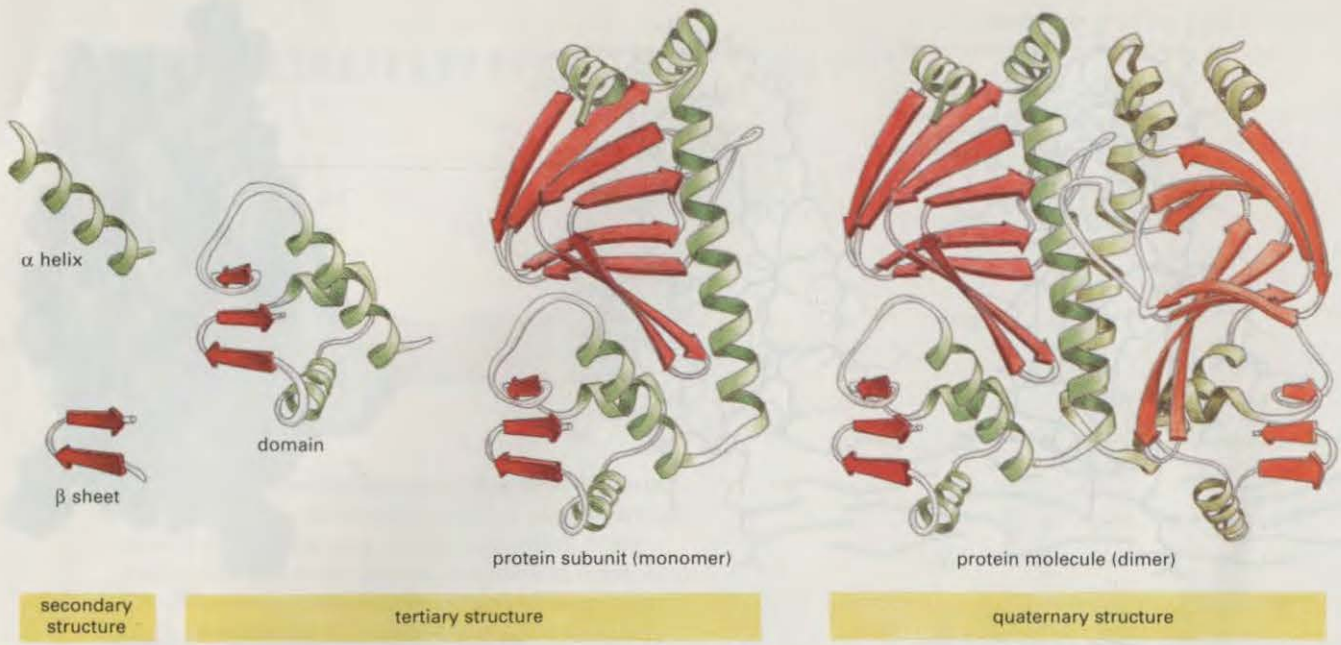


Figure 3–34 Basic pancreatic trypsin inhibitor (BPTI). The three-dimensional conformation of this small protein is shown in five commonly used representations. (A) A stereo pair illustrating the positions of all nonhydrogen atoms. The main chain is shown with heavy lines and the side chains with thin lines. (B) Space-filling model showing the van der Waals radii of all atoms (see Panel 3–1, pp. 92–93). (C) Backbone wire model composed of lines that connect each α carbon along the polypeptide backbone. (D) "Ribbon model," which represents all regions of regular hydrogen-bonded interactions as either helices (α helices) or sets of arrows (β sheets) pointing toward the carboxyl-terminal end of the chain. (E) "Sausage model," which shows the course of the polypeptide chain but omits all detail. In the bottom three panels the *hairpin beta motif* is colored *green*; this motif is also found in many other proteins (see text). Note that the core of all globular proteins is densely packed with atoms. Thus the impression of an open structure produced by models (C), (D), and (E) is misleading. (B and C, courtesy of Richard J. Feldmann; A and D, courtesy of Jane Richardson.)

Sarepta Exhibit 1040, page 34



The three-dimensional structure of a protein can be illustrated in various ways. Consider the unusually small protein basic pancreatic trypsin inhibitor (BPTI), which contains 58 amino acid residues folded into one domain. BPTI can be shown as a stereo pair displaying all of its nonhydrogen atoms (Figure 3–34A) or as an accurate space-filling model, where most of the details are obscured (Figure 3–34B). Alternatively, it can be shown more schematically, with all of the side chains and actual atoms omitted so that it is easier to follow the course of the main polypeptide chain (Figures 3–34C, D, and E). An average-size protein contains about six times more amino acid residues than BPTI, and many proteins are more than 20 times its size. Schematic drawings are essential for displaying the structure of these larger proteins, and we use them throughout this text.

Figure 3–35 shows how the structure of a large protein can be resolved into several levels of organization, each level constructed from the one below it in a hierarchical fashion. These levels of increased organizational complexity may correspond to the steps by which a newly synthesized protein folds into its final native structure inside the cell.

Domains Are Formed from a Polypeptide Chain That Winds Back and Forth, Making Sharp Turns at the Protein Surface [24]

A protein domain can be viewed as the basic structural unit of a protein structure. The core of each domain is largely composed of a set of interconnected β sheets or α helices or both. These regular secondary structures are favored because they permit an extensive hydrogen bonding between the backbone atoms, which is essential for stabilizing the interior of the domain, where water is not available to form hydrogen bonds with the polar carbonyl oxygen or amide hydrogen of the peptide bond.

Because there are only a limited number of ways of combining α helices and β sheets to make a globular structure, certain combinations of these elements, called **motifs,** occur repeatedly in the core of many unrelated proteins. One example is the *hairpin beta motif* found in BPTI (colored *green* in Figure 3–34D), which consists of two antiparallel β strands joined by a sharp turn formed by a loop of polypeptide chain. Another example is the *beta-alpha-beta motif,* in

Figure 3–35 Three levels of organization of a protein. The three-dimensional structure of a protein can be described in terms of different levels of folding, each of which is constructed from the preceding one in hierarchical fashion. These levels are illustrated here using the catabolite activator protein (CAP), a bacterial gene regulatory protein with two domains. When the large domain binds cyclic AMP, it causes a conformational change in the protein that enables the small domain to bind to a specific DNA sequence. The amino acid sequence is termed the *primary structure* and the first folding level the *secondary structure.* As indicated under the brackets at the bottom of this figure, the combination of the second and third folding levels shown here is commonly termed the *tertiary structure,* and the fourth level (the assembly of subunits) the *quaternary structure* of a protein. (Modified from a drawing by Jane Richardson.)

Sarepta Exhibit 1040, page 35

which two adjacent parallel β strands are connected by a length of α helix (Figure 3–36). Several other common motifs are discussed in Chapter 9, where we consider the various DNA-binding motifs found in several families of gene regulatory proteins.

Various combinations of motifs form the protein domain itself, in which the polypeptide chain tends to wind its way back and forth across the entire structure, either as a β sheet or an α helix, reversing direction suddenly by making a tight turn when it reaches the surface of the domain. As a result, a typical domain is a compact structure whose surface is covered by protruding loops of polypeptide chain (Figure 3–37). The *loop regions*, which vary in length and have an irregular shape, often form the binding sites for other molecules. Because the loop regions are exposed to water, they are rich in hydrophilic amino acids, and on this basis their positions can frequently be predicted from a careful examination of the amino acid sequence of a protein.

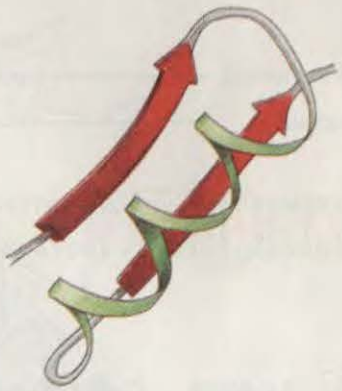


Figure 3–36 Example of a common protein motif. In the beta-alpha-beta motif two adjacent parallel strands that form a β sheet structure are connected by an α helix. Like the hairpin beta motif highlighted in Figure 3–34, this motif is found in many different proteins.

Relatively Few of the Many Possible Polypeptide Chains Would Be Useful

Since each of the 20 amino acids is chemically distinct and each can, in principle, occur at any position in a protein chain, there are $20 \times 20 \times 20 \times 20 = 160{,}000$ different possible polypeptide chains 4 amino acids long, or 20^n different possible polypeptide chains n amino acids long. For a typical protein length of about 300 amino acids, more than 10^{390} different proteins can be made.

We know, however, that only a very small fraction of these possible proteins would adopt a stable three-dimensional conformation. The vast majority would have many different conformations of roughly equal energy, each with different chemical properties. Proteins with such variable properties would not be useful and would therefore be eliminated by natural selection in the course of evolution. Present-day proteins have an amazingly sophisticated structure and chemistry because of their unique folding properties. Not only is the amino acid sequence such that a single conformation is extremely stable, but this conformation has the precise chemical properties that enable the protein to perform a specific catalytic or structural function in the cell. Proteins are so precisely built that the change of even a few atoms in one amino acid can sometimes disrupt the structure and cause a catastrophic change in function.

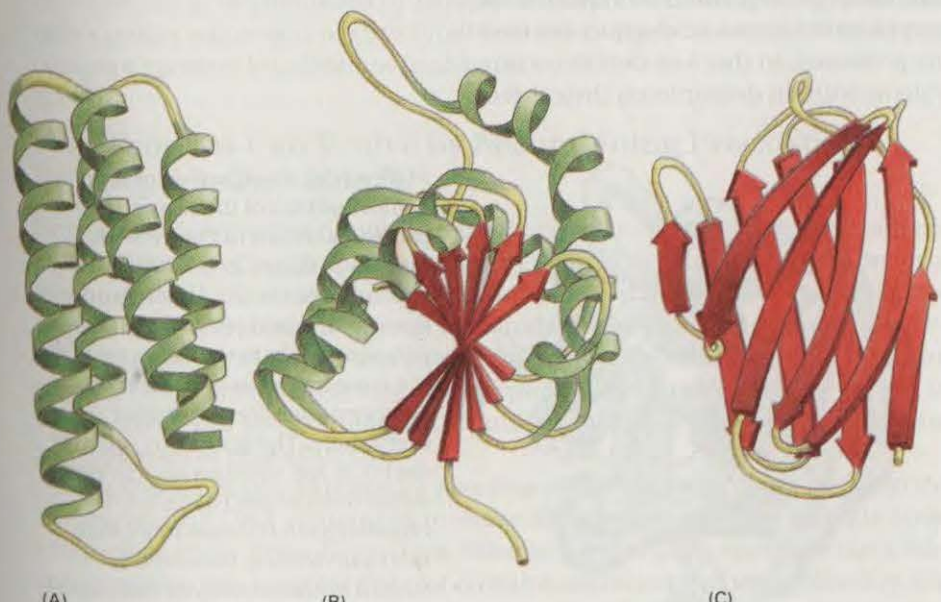


Figure 3–37 Ribbon models of the three-dimensional structure of several differently organized protein domains. (A) Cytochrome b_{562}, a single-domain protein composed almost entirely of α helices. (B) The NAD-binding domain of lactic dehydrogenase, composed of a mixture of α helices and β sheets. (C) The variable domain of an immunoglobin light chain, composed of a sandwich of two β sheets. In these examples the α helices are shown in *green*, while strands organized as β sheets are denoted by *red arrows*. Note that the polypeptide chain generally traverses back and forth across the entire domain, making sharp turns only at the protein surface. The protruding *loop regions* (*yellow*) often form the binding sites for other molecules. (Drawings courtesy of Jane Richardson.)

Sarepta Exhibit 1040, page 36

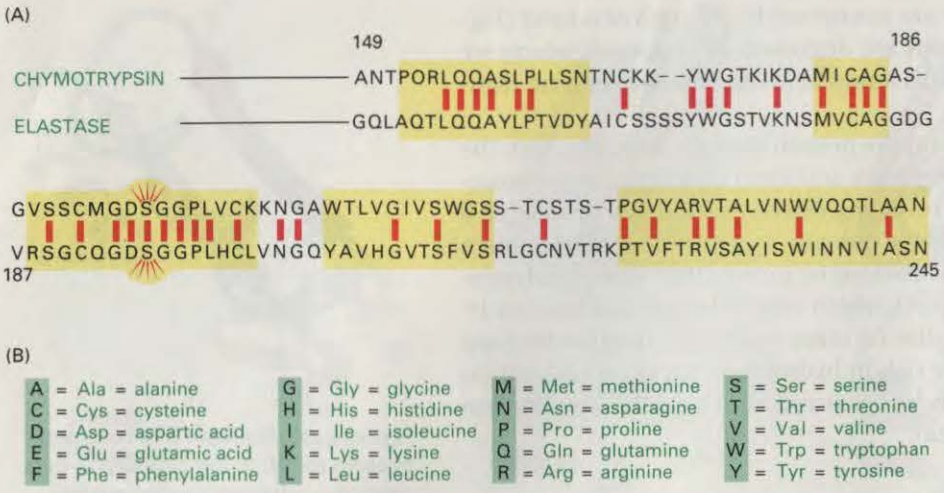


Figure 3–38 (A) Comparison of the amino acid sequences of two members of the serine protease family of enzymes. The carboxyl-terminal portions of the two proteins are shown (amino acids 149 to 245). Identical amino acids are connected by colored bars, and the serine residue in the active site at position 195 is highlighted. In the *yellow* boxed sections of the polypeptide chains, each amino acid occupies a closely equivalent position in the three-dimensional structures of the two enzymes (see Figure 3–39). (B) The standard one-letter and three-letter codes for amino acids. (Modified from J. Greer, *Proc. Natl. Acad. Sci. USA* 77:3393–3397, 1980.)

New Proteins Usually Evolve by Alterations of Old Ones [25]

Cells have genetic mechanisms that allow genes to be duplicated, modified, and recombined in the course of evolution. Consequently, once a protein with useful surface properties has evolved, its basic structure can be incorporated in many other proteins. Proteins of different but related function in present-day organisms often have similar amino acid sequences. Such families of proteins are believed to have evolved from a single ancestral gene that duplicated in the course of evolution to give rise to other genes in which mutations gradually accumulated to produce related proteins with new functions.

Consider the **serine proteases,** a family of protein-cleaving (proteolytic) enzymes that includes the digestive enzymes chymotrypsin, trypsin, and elastase and some of the proteases in the blood-clotting and complement enzymatic cascades. When two of these enzymes are compared, about 40% of the positions in their amino acid sequences are found to be occupied by the same amino acid (Figure 3–38). The similarity of their three-dimensional conformations as determined by x-ray crystallography is even more striking: most of the detailed twists and turns in their polypeptide chains, which are several hundred amino acids long, are identical (Figure 3–39).

The story that we have told for the serine proteases could be repeated for hundreds of other protein families. In many cases the amino acid sequences have diverged much further than for the serine proteases, so that one cannot be sure of a family relationship between two proteins without determining their three-

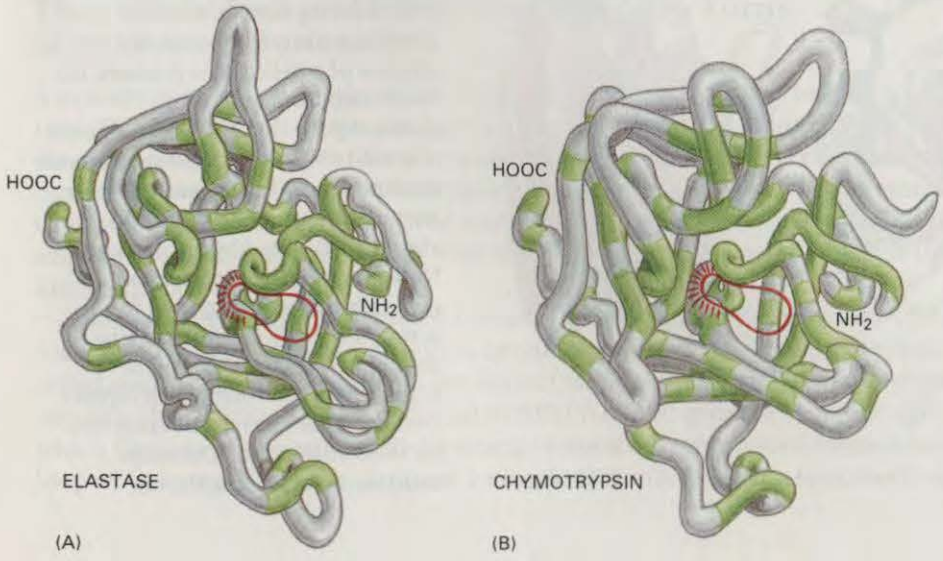
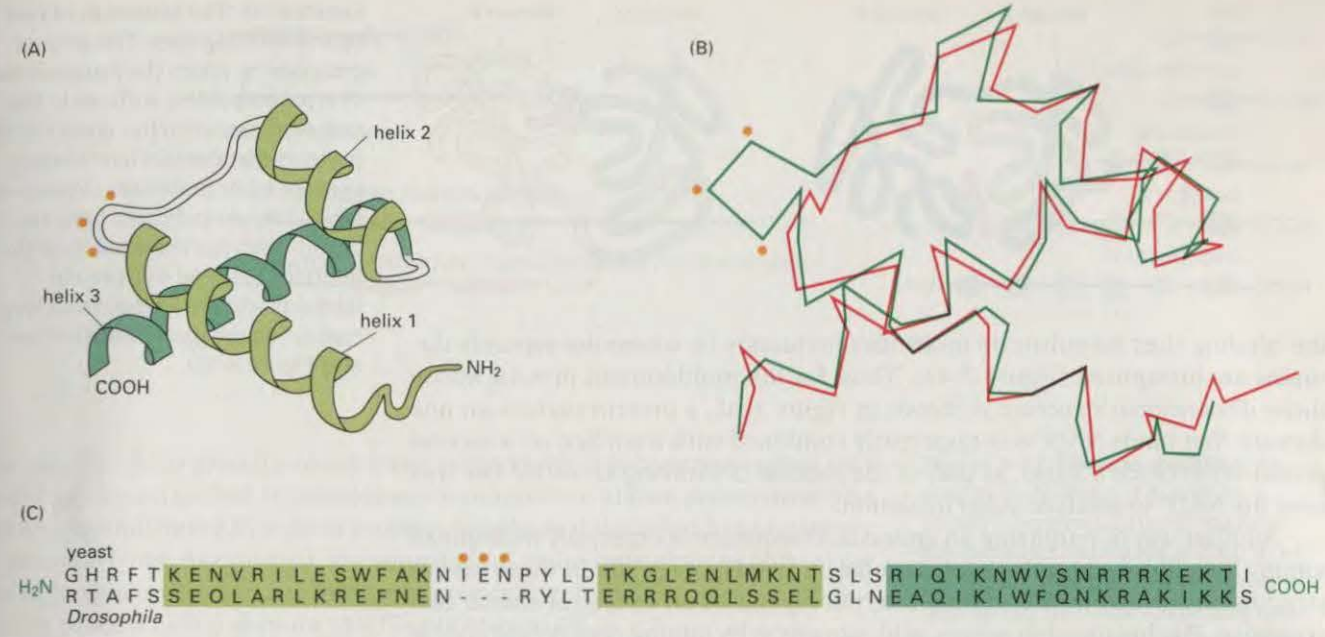


Figure 3–39 Comparison of the conformations of the two serine proteases shown in Figure 3–38. Elastase is shown in (A) and chymotrypsin in (B). Although only those amino acid residues in the polypeptide chain shaded in *green* are the same in the two proteins, their conformations are very similar everywhere. The active site, which is circled in *red,* contains an activated serine residue (see Figure 3–57). Chymotrypsin contains more than two chain termini because it is formed by the proteolytic cleavage of chymotrypsinogen, an inactive precursor.

Sarepta Exhibit 1040, page 37



dimensional structures. The yeast $\alpha 2$ protein and the *Drosophila* engrailed protein, for example, are both gene regulatory proteins in the homeodomain family. Because they are identical in only 17 of their 60 amino acid residues, their relationship became certain only when their three-dimensional structures were compared (Figure 3–40).

The various members of a large protein family will often have distinct functions. Some of the amino acid changes that make these proteins different were no doubt selected in the course of evolution because they resulted in changes in biological activity, giving the individual family members the different functional properties that they have today. Other amino acid changes are likely to be "neutral," having neither a beneficial nor a damaging effect on the basic structure and function of the protein. Since mutation is a random process, there must also have been many deleterious changes that altered the three-dimensional structure of these proteins sufficiently to inactivate them. Such inactive proteins would have been lost whenever the individual organisms making them were at enough of a disadvantage to be eliminated by natural selection. It is not surprising, then, that cells contain whole sets of structurally related polypeptide chains that have a common ancestry but different functions.

Figure 3–40 Comparison of DNA-binding homeodomains from two organisms separated by more than a billion years of evolution. (A) Schematic of structure. (B) Trace of the α-carbon positions. The three-dimensional structures shown were determined by x-ray crystallography for the yeast α2 protein (*green*) and the *Drosophila* engrailed protein (*red*). (C) Comparison of amino acid sequences for the region of the proteins shown in (A) and (B). *Orange dots* demark the position of a three amino acid insert in the α2 protein. (Adapted from C. Wolberger, et al., *Cell* 67:517–528, 1991.)

New Proteins Can Evolve by Recombining Preexisting Polypeptide Domains [26]

Once a number of stable protein surfaces have been made in a cell, new surfaces with different binding properties can be generated by joining two or more proteins together by noncovalent interactions between them, producing a *protein complex*. This combining of proteins to make larger, functional protein assemblies is common. Many protein complexes have molecular weights of a million or more, even though an average polypeptide chain has a molecular weight of 40,000 (about 300 to 400 amino acids), and relatively few polypeptide chains are more than three times this size.

An alternative way of making a new protein from existing chains is to join the corresponding DNA sequences to make a gene that encodes a single large polypeptide chain. Proteins in which different parts of the polypeptide chain fold independently into separate globular domains are believed to have evolved in this way, perhaps after existing for a prolonged period as a protein complex formed from separate polypeptides. Many proteins have such "multidomain" structures, and, as might be expected from the evolutionary considerations discussed above,

Sarepta Exhibit 1040, page 38

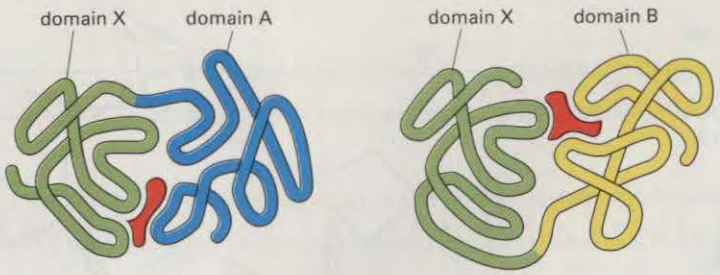


Figure 3–41 The evolution of new ligand-binding sites. The general principle by which the juxtaposition of separate protein surfaces in the course of evolution has given rise to proteins that contain new binding sites for other molecules (*ligands*—see p. 129). As indicated here, the ligand-binding sites often lie at the interface between two protein domains and are formed from loop regions on the protein surface (see also Figure 3–42).

the binding sites for substrate molecules frequently lie where the separate domains are juxtaposed (Figure 3–41). Thus, for the multidomain protein whose three-dimensional structure is shown in Figure 3–42, a protein surface on one domain that binds NAD^+ was apparently combined with a surface on a second domain that binds a sugar, as part of the process of evolving an active site that uses the NAD^+ to catalyze sugar oxidation.

Another way of reutilizing an amino acid sequence is especially widespread among long fibrous proteins such as collagen (see Figure 3–32). In these cases a structure is formed from multiple internal repeats of an ancestral amino acid sequence. Putting together amino acid sequences by joining preexisting coding DNA sequences is clearly a much more efficient strategy for a cell than the alternative of deriving new protein sequences from scratch by random DNA mutation.

Structural Homologies Can Help Assign Functions to Newly Discovered Proteins [27]

The development of techniques for rapidly sequencing DNA molecules has made it possible to determine the amino acid sequences of many thousands of proteins from the nucleotide sequences of their genes. A rapidly enlarging *protein data base* is therefore available that biologists routinely scan by computer to search for possible sequence homologies between a newly sequenced protein and previously studied ones. Although sequences have so far been determined for only

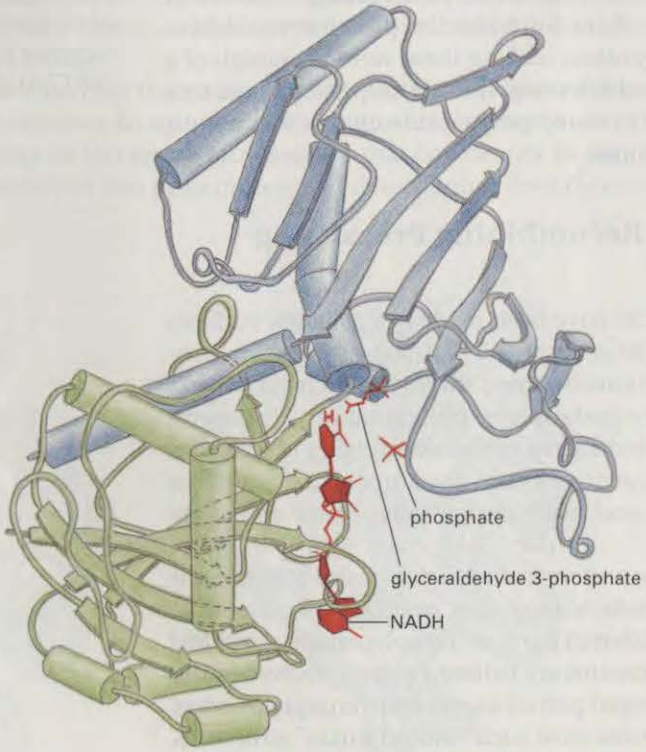


Figure 3–42 The structure of the glycolytic enzyme glyceraldehyde 3-phosphate dehydrogenase. The protein is composed of two domains, each shown in a different color, with regions of α helix represented by cylinders and regions of β sheet represented by arrows. The details of the reaction catalyzed by the enzyme are shown in Figure 2–22. Note that the three bound substrates lie at an interface between the two domains. (Courtesy of Alan J. Wonacott.)

Sarepta Exhibit 1040, page 39

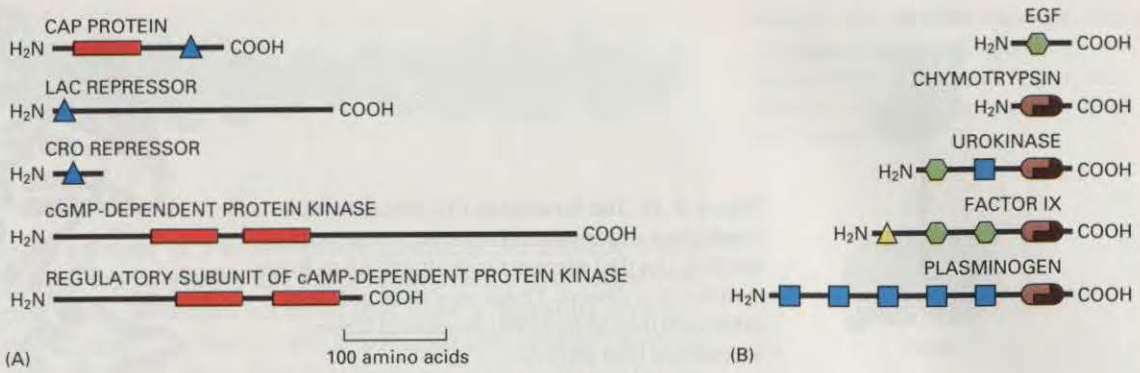


Figure 3–43 Domain shuffling. An extensive shuffling of blocks of protein sequence (protein *modules*) has occurred during the evolution of proteins. Those portions of a protein denoted by the same shape and color are evolutionarily related but not identical. (A) The bacterial catabolite gene activator protein (CAP) contains one domain (*blue triangle*) that binds a specific DNA sequence and a second domain (*red rectangle*) that binds cyclic AMP (see Figure 3–35). The DNA-binding domain here is related to the DNA-binding domains of many other gene regulatory proteins, including the lac repressor and cro repressor proteins. In addition, two copies of the cyclic-AMP-binding domain are found in eucaryotic protein kinases regulated by the binding of cyclic nucleotides. (B) Serine proteases like chymotrypsin are formed from two domains (*brown*). In some related proteases that are highly regulated and more specialized, the two protease domains are connected to one or more domains homologous to domains found in epidermal growth factor (*green hexagon*), to a calcium-binding protein (*yellow triangle*), or to a "kringle" domain (*blue square*) that contains three internal disufide bridges.

a few percent of the proteins in eucaryotic organisms, it is common to find that a newly sequenced protein is homologous to some other, known protein over part of its length, indicating that most proteins may have descended from relatively few ancestral types. As expected, the sequences of many large proteins often show signs of having evolved by the joining of preexisting domains in new combinations—a process called *domain shuffling* (Figure 3–43).

These protein comparisons are important because related structures often imply related functions. Many years of experimentation can be saved by discovering an amino acid sequence homology with a protein of known function. Such sequence homologies, for example, first indicated that certain cell-cycle regulatory genes in yeast cells and certain genes that cause mammalian cells to become cancerous are protein kinases. In the same way many of the proteins that control pattern formation in the fruit fly *Drosophila* were recognized to be gene regulatory proteins, while another protein involved in pattern formation was identified as a serine protease.

The discovery of domain homologies can also be useful in another way. It is much more difficult to determine the three-dimensional structure of a protein than to determine its amino acid sequence. But the conformation of a newly sequenced protein domain can be guessed if it is homologous to a domain of a protein whose conformation has already been determined by x-ray diffraction analysis. By assuming that the twists and turns of the polypeptide chain will be conserved in the two proteins despite the presence of discrepancies in amino acid sequence, one can often sketch the structure of the new protein with reasonable accuracy (see Figure 3–40).

Many new protein sequences are being added to the data base each year, each one increasing the chance of finding useful homologies. Protein-sequence comparisons have therefore become a very important tool in cell biology.

Protein Subunits Can Assemble into Large Structures [28]

The same principles that enable several protein domains to associate to form binding sites for small molecules operate to generate much larger structures in the cell. Supramolecular structures such as enzyme complexes, ribosomes, protein filaments, viruses, and membranes are not made as single, giant, covalently linked molecules; instead they are formed by the noncovalent assembly of many preformed molecules, which are called *subunits* of the final structure.

There are several advantages to the use of smaller subunits to build larger structures: (1) building a large structure from one or a few repeating smaller subunits reduces the amount of genetic information required; (2) both assembly and disassembly can be readily controlled, since the subunits associate through multiple bonds of relatively low energy; and (3) errors in the synthesis of the structure can be more easily avoided, since correction mechanisms can operate during the course of assembly to exclude malformed subunits.

Sarepta Exhibit 1040, page 40

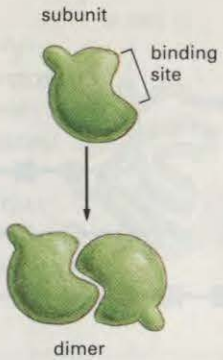


Figure 3–44 The formation of a dimer from a single type of protein subunit. A protein with a binding site that recognizes itself will often form symmetrical dimers. These may then pair with other subunits to form tetramers and larger assemblies (not shown).

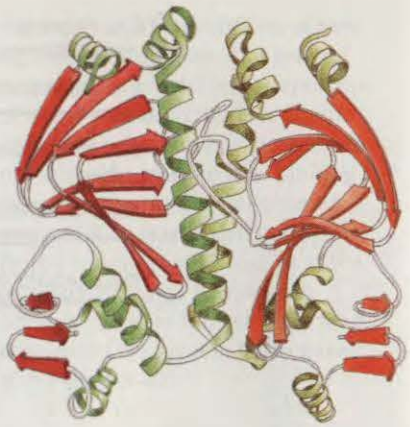


Figure 3–45 Ribbon model of a dimer formed from two identical protein subunits (monomers). The protein shown is the bacterial catabolite gene activator protein (CAP) illustrated previously in Figure 3–35. (Courtesy of Jane Richardson.)

A Single Type of Protein Subunit Can Interact with Itself to Form Geometrically Regular Assemblies [29]

If a protein has a binding site that is complementary to a region of its own surface, it will assemble spontaneously to form a larger structure. In the simplest case, a binding site recognizes itself and forms a symmetrical *dimer.* Many enzymes and other proteins form dimers of this kind, which frequently act as subunits in the formation of larger assemblies (Figures 3–44 and 3–45).

If the binding site of a protein is complementary to a region of its surface that does not include the binding site itself, a chain of subunits will be formed. For certain special orientations of the two binding sites, the chain will soon run into itself and terminate, forming a closed ring of subunits (Figure 3–46). More commonly, an extended polymer of subunits will result, and provided that each subunit is bound to its neighbor in an identical way, the subunits in the polymer will be arranged in a helix that can be extended indefinitely (see Figure 3–5). An *actin filament,* for example, is a helical structure formed from a single globular protein subunit called *actin;* actin filaments are major components in the cytosol of most eucaryotic cells (Figure 3–47). As we discuss below, globular proteins may also associate with like neighbors to form extended sheets or tubes (see Figure 3–49).

Coiled-Coil Proteins Help Build Many Elongated Structures in Cells [30]

Where mechanical strength is of major importance, supramolecular assemblies are usually made from fibrous rather than globular subunits. Such assemblies can be stabilized by extensive regions of protein-protein contact when the subunits

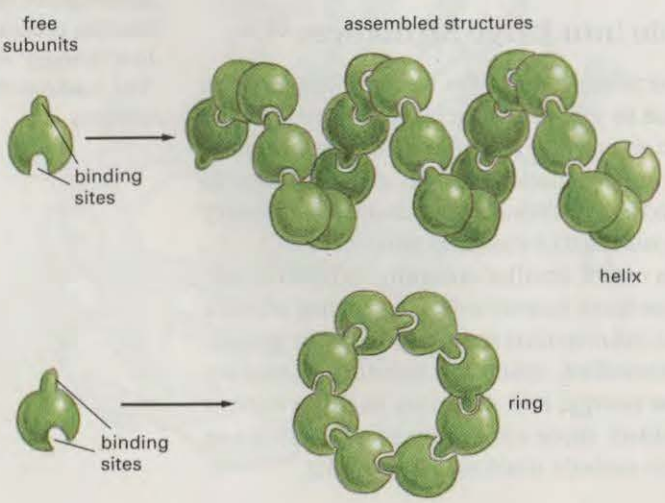


Figure 3–46 Rings or helices can form if a single type of protein subunit interacts with itself repeatedly. The formation of a helix was illustrated in Figure 3–5; a ring forms instead of a helix if the subunits run into one another, stopping further growth of the chain.

Sarepta Exhibit 1040, page 41



Figure 3–47 An actin filament. There are about two globular protein subunits per turn in this important filament, which is discussed in detail in Chapter 16.

are wound around one another as a multistranded helix. A particularly stable structural unit that is used repeatedly for this purpose is known as the **coiled-coil.** It forms by the pairing of two α-helical subunits that have a repeating arrangement of nonpolar side chains. The two α-helical subunits are usually identical and run in parallel (that is, in the same direction from amino to carboxyl terminal). They coil gradually around each other to produce a stiff filament with a diameter of about 2 nm (Figure 3–48). Whereas short coiled-coils serve as dimerization domains in several families of gene regulatory proteins, more commonly a coiled-coil will extend for more than 100 nm and serve as a building block for a large fibrous structure, such as the thick filaments in a muscle cell.

Proteins Can Assemble into Sheets, Tubes, or Spheres [31]

Some protein subunits assemble into flat sheets in which the subunits are arranged in hexagonal arrays. Specialized membrane proteins are sometimes arranged in this way in lipid bilayers. With a slight change in the geometry of the individual subunits, a hexagonal sheet can be converted into a tube (Figure 3–49) or, with more changes, into a hollow sphere. Protein tubes and spheres that bind specific RNA and DNA molecules form the coats of viruses.

The formation of closed structures, such as rings, tubes, or spheres, provides additional stability because it increases the number of bonds that can form between the protein subunits. Moreover, because such a structure is formed by mutually dependent, cooperative interactions between subunits, it can be driven

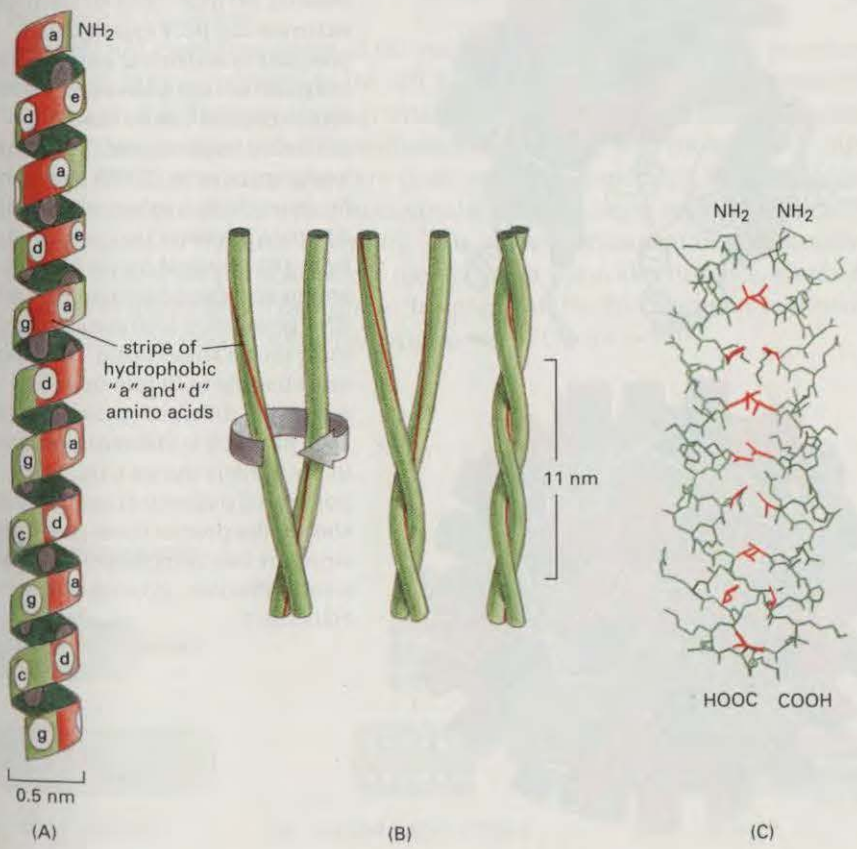


Figure 3–48 The structure of a coiled-coil. In (A) a single α helix is shown, with successive amino acid side chains labeled in a sevenfold sequence "abcdefg" (from bottom to top). Amino acids "a" and "d" in such a sequence lie close together on the cylinder surface, forming a "stripe" (shaded in *red*) that winds slowly around the α helix. Proteins that form coiled-coils typically have hydrophobic amino acids at positions "a" and "d." Consequently, as shown in (B), the two α helices can wrap around each other with the hydrophobic side chains of one α helix interacting with the hydrophobic side chains of the other, while the more hydrophilic amino acid side chains are left exposed to the aqueous environment. (C) The atomic structure of a coiled-coil determined by x-ray crystallography. The *red* side chains are hydrophobic. (C, from T. Alber, *Curr. Opin. Genet. Devel.* 2:205–210, 1992. © Current Science.)

Sarepta Exhibit 1040, page 42

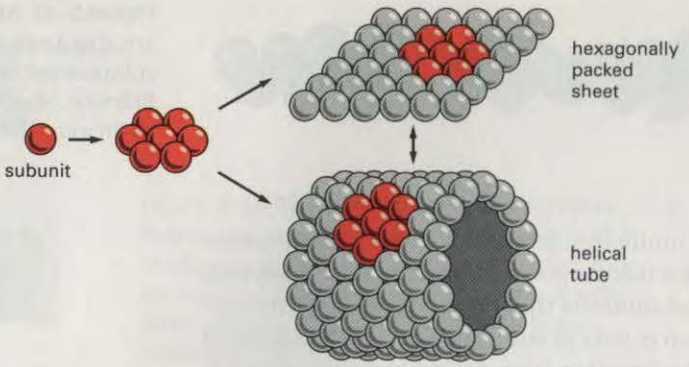


Figure 3–49 Hexagonally packed globular protein subunits can form either a flat sheet or a tube.

to assemble or disassemble by a relatively small change that affects the subunits individually. These principles are dramatically illustrated in the protein *capsid* of many simple viruses, which takes the form of a hollow sphere. These coats are often made of hundreds of identical protein subunits that enclose and protect the viral nucleic acid (Figure 3–50). The protein in such a capsid must have a particularly adaptable structure, since it must make several different kinds of contacts and also change its arrangement to let the nucleic acid out to initiate viral replication once the virus has entered a cell.

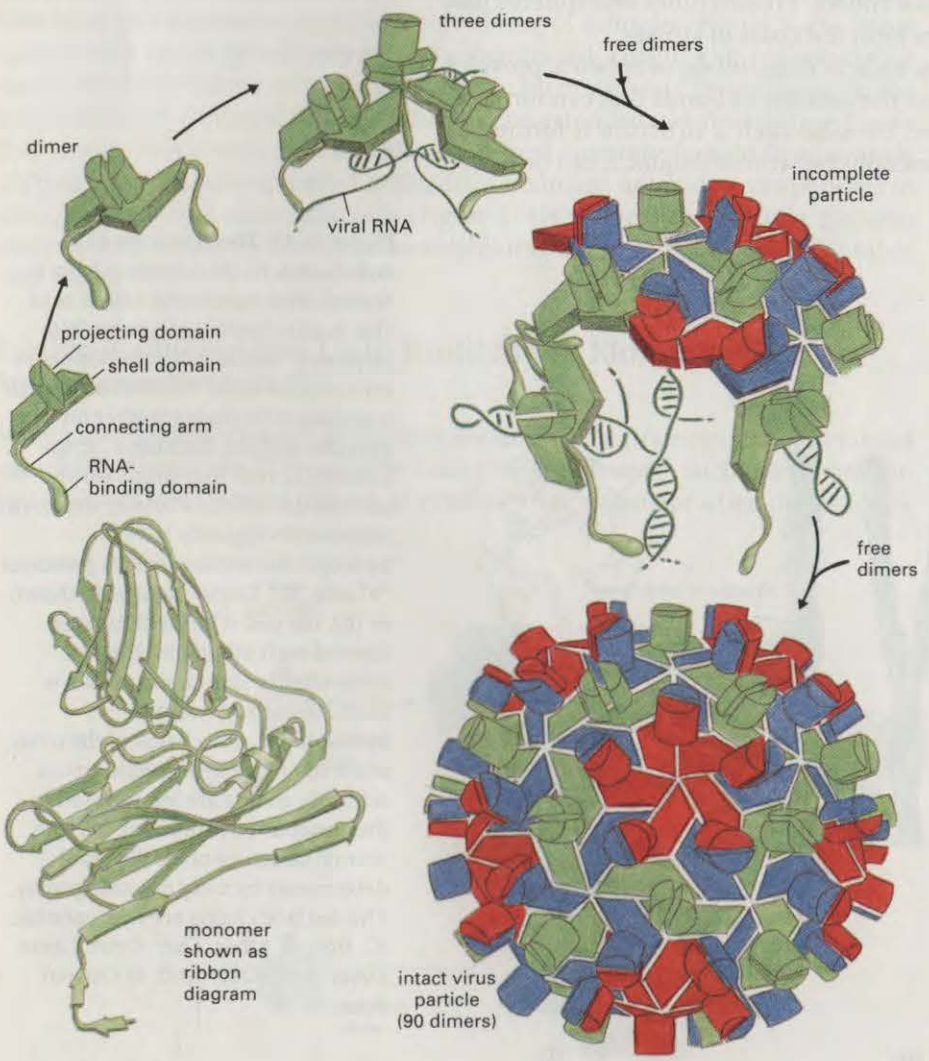
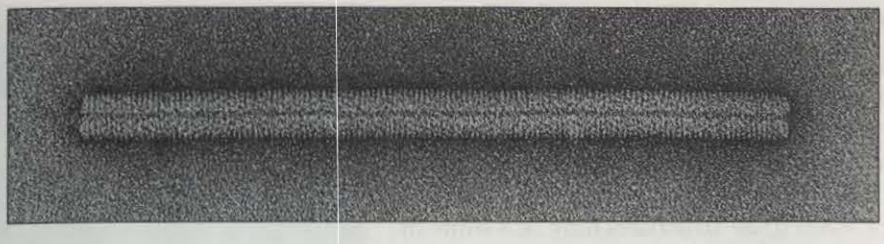
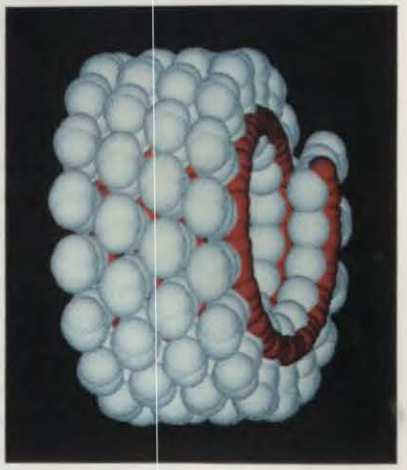


Figure 3–50 The structure of a spherical virus. In many viruses, identical protein subunits pack together to create a spherical shell (a capsid) that encloses the viral genome, composed of either RNA or DNA (see Figure 6–72). For geometric reasons, no more than 60 identical subunits can pack together in a precisely symmetrical way. If slight irregularities are allowed, however, more subunits can be used to produce a larger capsid. The tomato bushy stunt virus (TBSV) shown here, for example, is a spherical virus about 33 nm in diameter that is formed from 180 identical copies of a 386 amino acid capsid protein plus an RNA genome of 4500 nucleotides. To form such a large capsid, the protein must be able to fit into three somewhat different environments, each of which is differently colored in the particle shown here. The postulated pathway of assembly is shown; the precise three-dimensional structure has been determined by x-ray diffraction. (Courtesy of Steve Harrison.)

Sarepta Exhibit 1040, page 43



(A)



(B)

Figure 3–51 The structure of tobacco mosaic virus (TMV). (A) Electron micrograph of a tobacco mosaic virus (TMV), which consists of a single long RNA molecule enclosed in a cylindrical protein coat composed of a tight helical array of identical protein subunits. (B) A model showing part of the structure of TMV. A single-stranded RNA molecule of 6000 nucleotides is packaged in a helical coat constructed from 2130 copies of a coat protein 158 amino acids long. Fully infective virus particles can self-assemble in a test tube from purified RNA and protein molecules. (A, courtesy of Robley Williams; B, courtesy of Richard J. Feldmann.)

Many Structures in Cells Are Capable of Self-assembly [32]

The information for forming many of the complex assemblies of macromolecules in cells must be contained in the subunits themselves, since under appropriate conditions the isolated subunits can spontaneously assemble in a test tube into the final structure. The first large macromolecular aggregate shown to be capable of self-assembly from its component parts was *tobacco mosaic virus (TMV)*. This virus is a long rod in which a cylinder of protein is arranged around a helical RNA core (Figure 3–51). If the dissociated RNA and protein subunits are mixed together in solution, they recombine to form fully active virus particles. The assembly process is unexpectedly complex and includes the formation of double rings of protein, which serve as intermediates that add to the growing virus coat.

Another complex macromolecular aggregate that can reassemble from its component parts is the bacterial ribosome. These ribosomes are composed of about 55 different protein molecules and 3 different rRNA molecules. If the individual components are incubated under appropriate conditions in a test tube, they spontaneously re-form the original structure. Most important, such reconstituted ribosomes are able to carry out protein synthesis. As might be expected, the reassembly of ribosomes follows a specific pathway: certain proteins first bind to the RNA, and this complex is then recognized by other proteins, and so on until the structure is complete.

It is still not clear how some of the more elaborate self-assembly processes are regulated. Many structures in the cell, for example, appear to have a precisely defined length that is many times greater than that of their component macromolecules. How such length determination is achieved is in most cases a mystery. Three possible mechanisms are illustrated in Figure 3–52. In the simplest case a long core protein or other macromolecule provides a scaffold that determines the extent of the final assembly. This is the mechanism that determines the length of the TMV particle, where the RNA chain provides the core. Similarly, a core protein is thought to determine the length of the thin filaments in muscle, as well as the long tails of some bacterial viruses (Figure 3–53).

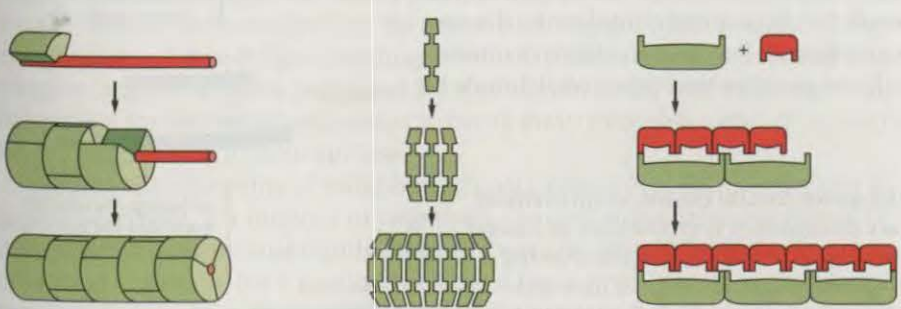


Figure 3–52 Three ways in which a large protein assembly can be made to a fixed length. (A) Coassembly along an elongated core protein or other macromolecule that acts as a measuring device. (B) Termination of assembly because of strain that accumulates in the polymeric structure as additional subunits are added, so that beyond a certain length the energy required to fit another subunit onto the chain becomes excessively large. (C) A vernier type of assembly, in which two sets of rodlike molecules differing in length form a staggered complex that grows until their ends exactly match.

Sarepta Exhibit 1040, page 44

Not All Biological Structures Form by Self-assembly [33]

Some cellular structures held together by noncovalent bonds are not capable of self-assembly. A mitochondrion, a cilium, or a myofibril, for example, cannot form spontaneously from a solution of their component macromolecules because part of the information for their assembly is provided by special enzymes and other cellular proteins that perform the function of jigs or templates but do not appear in the final assembled structure. Even small structures may lack some of the ingredients necessary for their own assembly. In the formation of some bacterial viruses, for example, the head structure, which is composed of a single protein subunit, is assembled on a temporary scaffold composed of a second protein. The second protein is absent from the final virus particle, and so the head structure cannot spontaneously reassemble once it is taken apart. Other examples are known in which proteolytic cleavage is an essential and irreversible step in the assembly process. This is the case for the coats of some bacterial viruses and even for some simple protein assemblies, including the structural protein collagen and the hormone insulin (Figure 3–54). From these relatively simple examples, it seems very likely that the assembly of a structure as complex as a mitochondrion or a cilium will involve both temporal and spatial ordering imparted by other cellular components, as well as irreversible processing steps catalyzed by degradative enzymes.

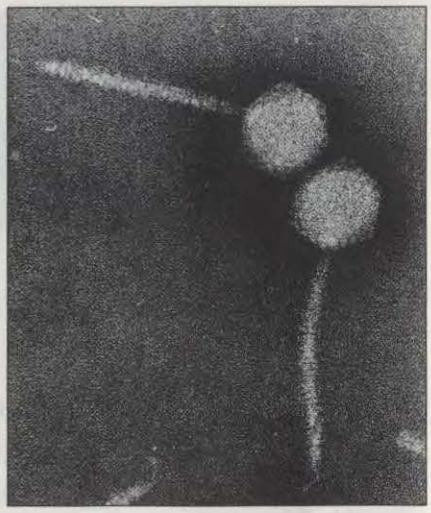


Figure 3–53 Electron micrograph of bacteriophage lambda. The tip of the virus tail attaches to a specific protein on the surface of a bacterial cell, following which the tightly packaged DNA in the head is injected through the tail into the cell. The tail has a precise length, which is determined by the mechanism shown in Figure 3–52A.

Summary

The three-dimensional conformation of a protein molecule is determined by its amino acid sequence. The folded structure is stabilized by noncovalent interactions between different parts of the polypeptide chain. The amino acids with hydrophobic side chains tend to cluster in the interior of the molecule, and local hydrogen-bond interactions between neighboring peptide bonds give rise to α helices and β sheets. Globular regions known as domains are the modular units from which many proteins are constructed; small proteins typically contain only a single domain, while large proteins contain several domains linked together by short lengths of polypeptide chain. As proteins evolved, domains were modified and combined with other domains to construct new proteins.

Proteins are brought together into larger structures by the same noncovalent forces that determine protein folding. Proteins with binding sites for their own surface can assemble into dimers, closed rings, spherical shells, or helical polymers. Although mixtures of proteins and nucleic acids can assemble spontaneously into complex structures in the test tube, many assembly processes involve irreversible steps. Consequently, not all structures in the cell are capable of spontaneous reassembly after they are dissociated into their component parts.

Proteins as Catalysts [34]

The chemical properties of a protein molecule depend almost entirely on its exposed surface residues, which are able to form weak, noncovalent bonds with other molecules. When a protein molecule binds to another molecule, the second molecule is commonly referred to as a **ligand.** Because an effective interaction between a protein molecule and a ligand requires that many weak bonds be

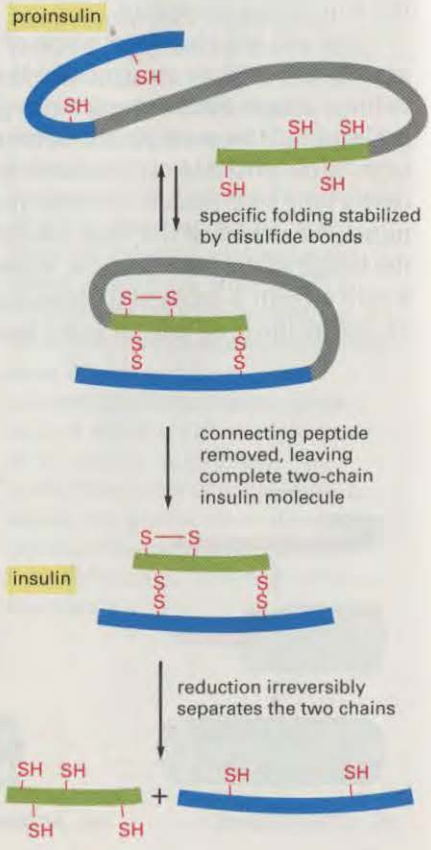


Figure 3–54 The polypeptide hormone insulin cannot spontaneously re-form if its disulfide bonds are disrupted. It is synthesized as a larger protein (*proinsulin*) that is cleaved by a proteolytic enzyme after the protein chain has folded into a specific shape. Excision of part of the proinsulin polypeptide chain causes an irretrievable loss of the information needed for the protein to fold spontaneously into its normal conformation.

Sarepta Exhibit 1040, page 45

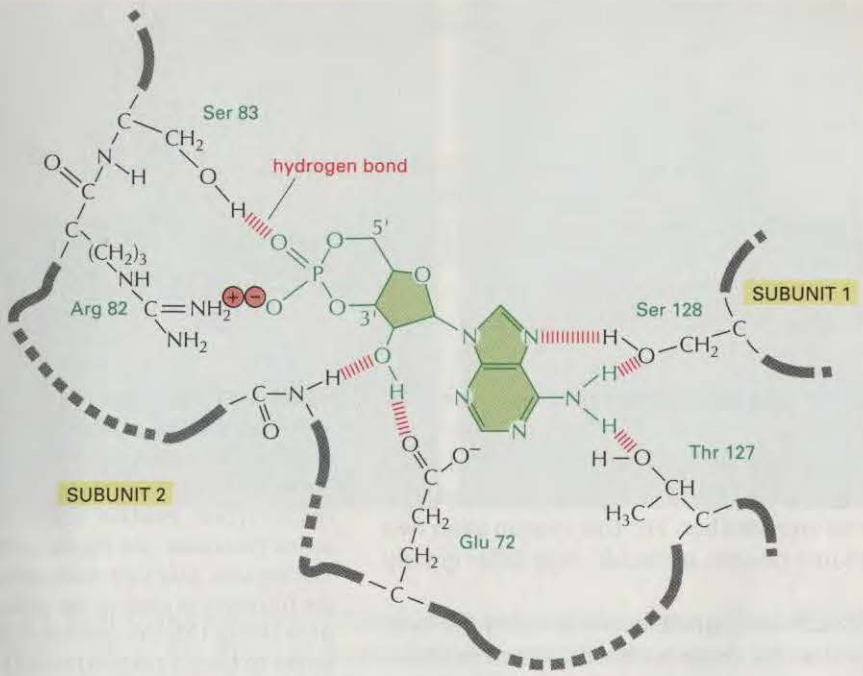


Figure 3–55 The ligand-binding site of the catabolite gene activator protein (CAP). Hydrogen bonding between CAP and its ligand, cyclic AMP (*green*), was determined by x-ray crystallographic analysis of the complex. As indicated, the two identical subunits of the dimer cooperate to form this binding site (see also Figure 3–45). (Courtesy of Tom Steitz.)

formed simultaneously between them, the only ligands that can bind tightly to a protein are those that fit precisely onto its surface.

The region of a protein that associates with a ligand, known as its **binding site,** usually consists of a cavity formed by a specific arrangement of amino acids on the protein surface. These amino acids often belong to widely separated regions of the polypeptide chain (Figure 3–55), and they represent only a minor fraction of the total amino acids present. The rest of the protein molecule is presumably necessary to maintain the polypeptide chain in the correct position and to provide additional binding sites for regulatory purposes; the interior of the protein is often important only insofar as it gives the surface of the molecule the appropriate shape and rigidity.

A Protein's Conformation Determines Its Chemistry [20]

Neighboring surface residues on a protein often interact in a way that alters the chemical reactivity of selected amino acid side chains. These interactions are of several types.

First, neighboring parts of the polypeptide chain may interact in a way that restricts the access of water molecules to other parts of the protein surface. Because water molecules tend to form hydrogen bonds, they compete with ligands for selected side chains on the protein surface (Figure 3–56). The tightness of hydrogen bonds (and ionic interactions) between proteins and their ligands is therefore greatly increased if water molecules are excluded. At first sight it is hard to imagine a mechanism that would exclude a molecule as small as water from a protein surface without affecting the access of the ligand itself. Because of their strong tendency for hydrogen bonding, however, water molecules exist in a large hydrogen-bonded network (see Panel 2–1, pp. 48–49), and it is often energetically unfavorable for individual molecules to break away from this network to reach into a crevice on the protein surface.

Second, the clustering of neighboring polar amino acid side chains can alter their reactivity. If a number of negatively charged side chains are forced together against their mutual repulsion by the way the protein folds, for example, the affinity of the site for a positively charged ion is greatly increased. Selected amino acid side chains can also interact with one another through hydrogen bonds, which can activate normally unreactive side groups (such as the $-CH_2OH$

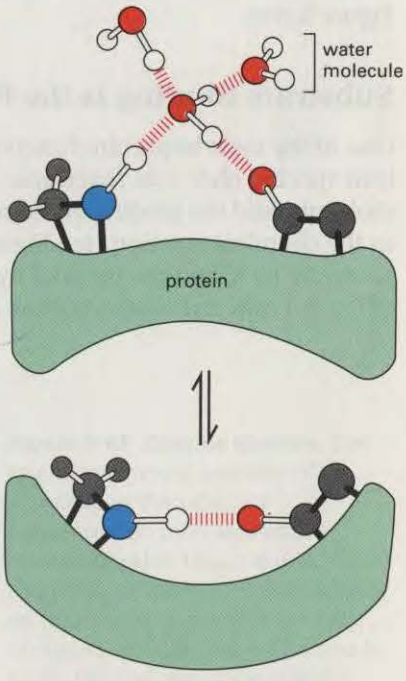


Figure 3–56 Competition for hydrogen bonding. The ability of water molecules to make favorable hydrogen bonds with groups on the protein surface greatly reduces the tendency of these groups to pair with each other.

Sarepta Exhibit 1040, page 46

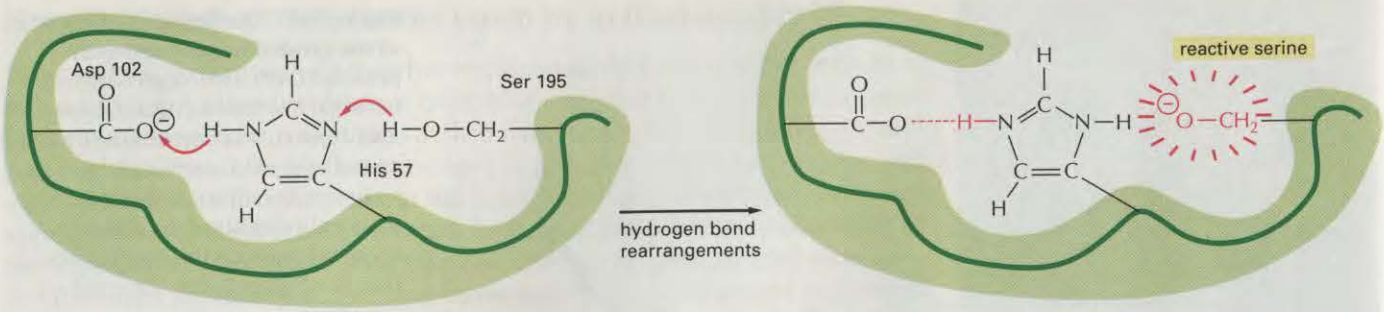


Figure 3–57 An unusually reactive amino acid at the active site of an enzyme. The example shown is the "catalytic triad" found in chymotrypsin, elastase, and other serine proteases (see Figure 3–39). The aspartic acid side chain induces the histidine to remove the proton from serine 195; this activates the serine to form a covalent bond with the enzyme substrate, hydrolyzing a peptide bond as illustrated later in Figure 3–64.

on the serine shown in Figure 3–57) so that they are able to enter into reactions that make or break selected covalent bonds.

The surface of each protein molecule therefore has a unique chemical reactivity that depends not only on which amino acid side chains are exposed, but also on their exact orientation relative to one another. For this reason even two slightly different conformations of the same protein molecule may differ greatly in their chemistry.

Where side-chain reactivities are insufficient, proteins often enlist the help of selected nonpolypeptide molecules that the proteins bind to their surface. These ligands serve as **coenzymes** in enzyme-catalyzed reactions, and they may be so tightly bound to the protein that they are effectively part of the protein itself. Examples are the iron-containing *hemes* in hemoglobin and cytochromes, *thiamine pyrophosphate* in enzymes involved in aldehyde-group transfers, and *biotin* in enzymes involved in carboxyl-group transfers. Most coenzymes are very complex organic molecules that have been selected for the unique chemical reactivity they acquire when bound to a protein surface. Besides its reactive center such a coenzyme has other residues designed to bind it to its host protein (Figure 3–58). A space-filling model of an enzyme bound to a coenzyme is shown in Figure 3–59A.

Substrate Binding Is the First Step in Enzyme Catalysis [35]

One of the most important functions of proteins is to act as enzymes that catalyze specific chemical reactions. The ligand in this case is called a **substrate** molecule, and the binding of the substrate to the enzyme is an essential prelude to the chemical reaction (see Figure 3–59B). If we denote the enzyme by E, the substrate by S, and the product by P, the basic reaction path is $E + S \rightleftharpoons ES \rightleftharpoons EP \rightleftharpoons E + P$. From this simple outline of an enzyme-catalyzed reaction, we see that

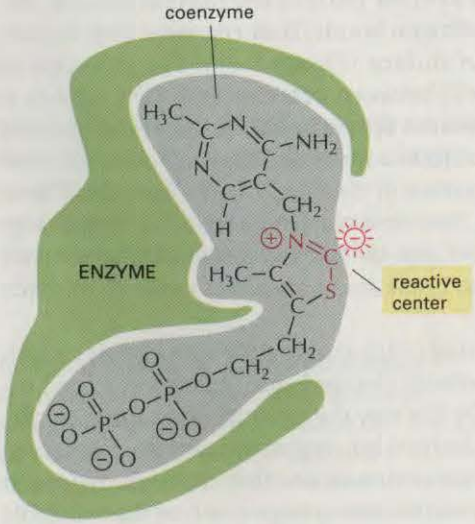
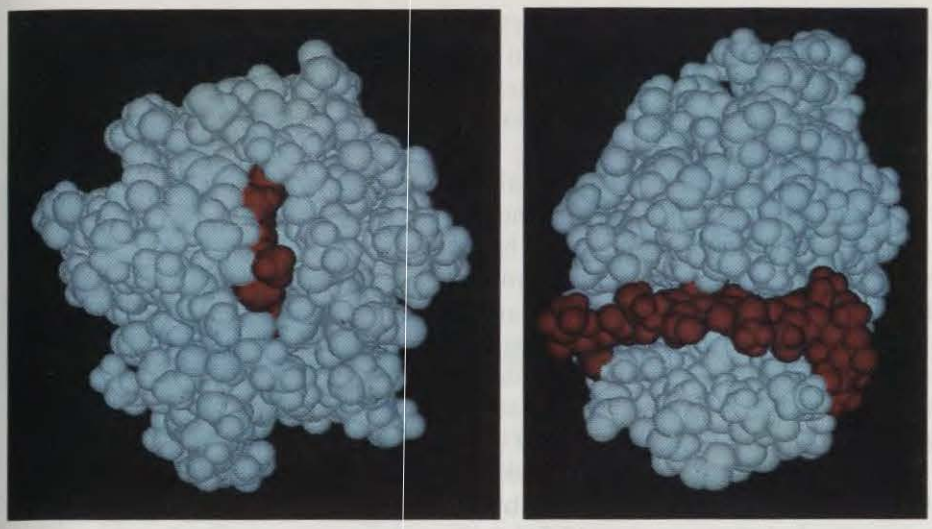


Figure 3–58 Coenzymes. Coenzymes, such as thiamine pyrophosphate (TPP), shown here in *gray*, are small molecules that bind to an enzyme's surface and enable it to catalyze specific reactions. The reactivity of TPP depends on its "acidic" carbon atom, which readily exchanges its hydrogen atom for a carbon atom of a substrate molecule. Other regions of the TPP molecule act as "handles" by which the enzyme holds the coenzyme in the correct position. Coenzymes presumably evolved first in an "RNA world," where they were bound to RNA molecules to help with catalysis (discussed in Chapter 1).

Sarepta Exhibit 1040, page 47



(A) (B)

Figure 3–59 Computer-generated space-filling models of two enzymes. In (A) cytochrome c is shown with its bound heme coenzyme. In (B) egg-white lysozyme is shown with a bound oligosaccharide substrate. In both cases the bound ligand is *red*. (Courtesy of Richard J. Feldmann.)

there is a limit to the amount of substrate that a single enzyme molecule can process in a given time. If the concentration of substrate is increased, the rate at which product is formed also increases, up to a maximum value (Figure 3–60). At that point the enzyme molecule is saturated with substrate and the rate of reaction (denoted V_{max}) depends only on how rapidly the substrate molecule can be processed. This rate divided by the enzyme concentration is called the **turnover number.** The turnover number is often about 1000 substrate molecules processed per second per enzyme molecule, but it can be much greater in extreme cases.

The other kinetic parameter frequently used to characterize an enzyme is its K_M, which is the substrate concentration that allows the reaction to proceed at one-half its maximum rate (see Figure 3–60). A *low* K_M value means that the enzyme reaches its maximum catalytic rate at a *low concentration* of substrate and generally indicates that the enzyme binds its substrate very tightly.

Enzymes Speed Reactions by Selectively Stabilizing Transition States [36]

Extremely high rates of chemical reaction are achieved by enzymes—far higher than for any synthetic catalysts. This efficiency is attributable to several factors. The enzyme serves, first, to increase the local concentration of substrate molecules at the catalytic site and to hold all of the appropriate atoms in the correct orientation for the reaction that is to follow. More important, however, some of the binding energy contributes directly to the catalysis. Substrate molecules pass through a series of intermediate forms of altered geometry and electron distribution before they form the ultimate products of the reaction, and the free energies of these intermediate forms—especially of those in the most unstable

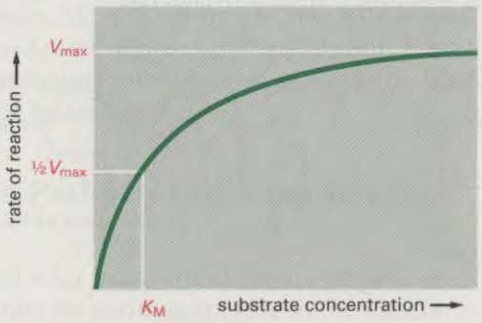


Figure 3–60 Enzyme kinetics. The rate of an enzyme reaction (V) increases as the substrate concentration increases until a maximum value (V_{max}) is reached. At this point all substrate-binding sites on the enzyme molecules are fully occupied, and the rate of reaction is limited by the rate of the catalytic process on the enzyme surface. For most enzymes the concentration of substrate at which the reaction rate is half-maximal (K_M) is a measure of how tightly the substrate is bound, with a large value of K_M corresponding to weak binding.

Sarepta Exhibit 1040, page 48

transition states—are the major determinants of the rate of reaction. Enzymes have a much greater affinity for these transition states of the substrate than they have for the stable forms. Because this binding interaction lowers the energies of crucial transition states, the enzyme greatly accelerates one particular reaction (Figure 3–61).

A dramatic demonstration of how stabilizing a transition state can greatly increase reaction rates is provided by the intentional production of antibodies that act like enzymes. Consider, for example, the hydrolysis of an amide bond, which is similar to the peptide bond that joins adjacent amino acids in a protein. In an aqueous solution an amide bond hydrolyzes very slowly by the mechanism illustrated in Figure 3–62A. In the central intermediate, or transition state, the carbonyl carbon is bonded to four atoms that are arranged at the corners of a tetrahedron. By generating monoclonal antibodies that bind tightly to a stable analogue of this very unstable *tetrahedral intermediate,* as illustrated in Figure 3–62B, an antibody that functions like an enzyme can be obtained. This *catalytic antibody* binds to and stabilizes the tetrahedral intermediate and thereby increases the spontaneous rate of amide-bond hydrolysis more than 10,000-fold.

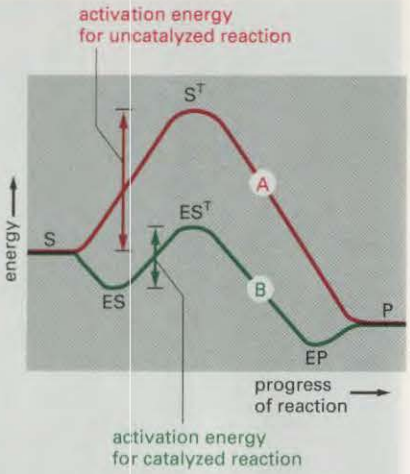


Figure 3–61 Enzymes accelerate chemical reactions by decreasing the activation energy. Often both the uncatalyzed reaction (A) and the enzyme-catalyzed reaction (B) go through several transition states. It is the transition state with the highest energy (S^T and ES^T) that determines the activation energy and limits the rate of the reaction. (S = substrate; P = product of the reaction.)

Enzymes Can Promote the Making and Breaking of Covalent Bonds Through Simultaneous Acid and Base Catalysis [37]

Enzymes are better catalysts than catalytic antibodies. In addition to binding tightly to the transition state, the active site of an enzyme contains precisely positioned atoms that speed up the reaction by altering the distribution of electrons in those atoms involved in the making and breaking of covalent bonds. Peptide bonds, for example, can be hydrolyzed in the absence of an enzyme by exposing a polypeptide to either a strong acid or a strong base, as explained in Figure 3–63B and C. Enzymes are unique, however, in being able to use acid and base catalysis simultaneously, since the acidic and basic residues required are prevented from combining with each other (as they would do in solution) by being tied to the rigid framework of the protein itself (Figure 3–63D).

The fit between an enzyme and its substrate needs to be precise. A small change introduced by genetic engineering in the active site of an enzyme can have a profound effect. Replacing a glutamic acid with an aspartic acid in one enzyme, for example, shifts the position of the catalytic carboxylate ion by only 1 Å (about the radius of a hydrogen atom), and yet this is enough to reduce the activity of the enzyme a thousandfold.

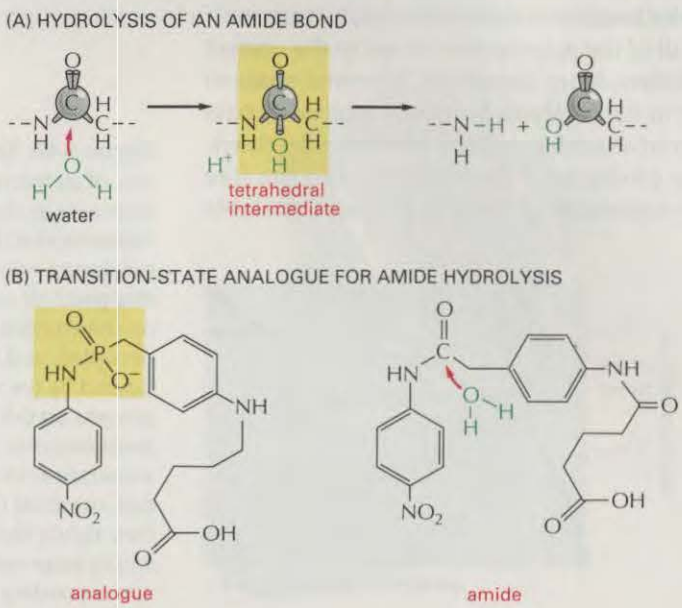


Figure 3–62 Catalytic antibodies. The stabilization of a transition state by an antibody creates an enzyme. (A) The reaction path for hydrolysis of an amide bond goes through a *tetrahedral intermediate,* which is the high-energy transition state for the reaction. (B) The molecule shown on the left was covalently linked to a protein and used as an antigen to generate an antibody that binds tightly to the region of the molecule shown in *yellow.* Because this antibody also bound tightly to the transition state in (A), it was found to function as an enzyme that efficiently catalyzed the hydrolysis of the amide bond in the molecule shown on the right.

Sarepta Exhibit 1040, page 49

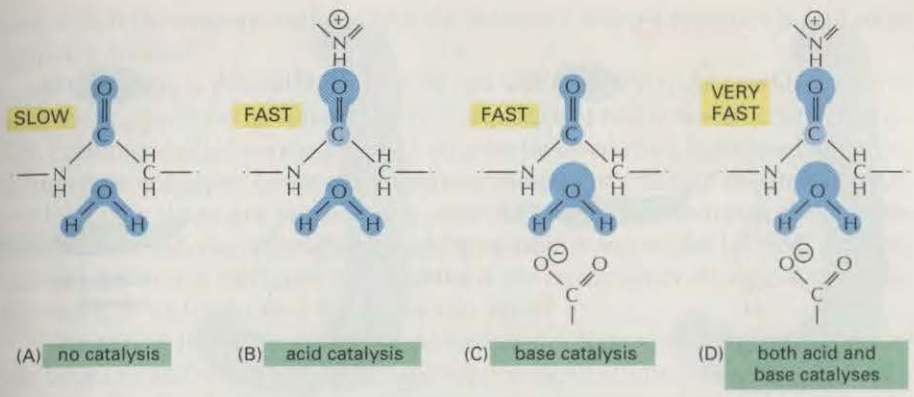


Figure 3–63 Acid catalysis and base catalysis. (A) The start of the uncatalyzed reaction shown in Figure 3–62A is diagrammed, with *blue* shading as a schematic indicator of electron distribution in the water and carbonyl bonds. (B) An acid likes to donate a proton (H^+) to other atoms. By pairing with the carbonyl oxygen, an acid causes electrons to move away from the carbonyl carbon, making this atom much more attractive to the electronegative oxygen of an attacking water molecule. (C) A base likes to take up H^+; by pairing with a hydrogen of the attacking water molecule, a base causes electrons to move toward the water oxygen, making it a better attacking group for the carbonyl carbon. (D) By having appropriately positioned atoms on its surface, an enzyme can carry out both acid catalysis and base catalysis at the same time.

Enzymes Can Further Increase Reaction Rates by Forming Covalent Intermediates with Their Substrates [38]

In addition to the above roles, many enzymes further speed the reaction they catalyze by interacting covalently with one of their substrates, thereby temporarily attaching the substrate to an amino acid or to a coenzyme molecule. Generally, one substrate enters the binding site, becomes covalently bound, and then reacts with a second molecule on the enzyme surface that breaks the covalent attachment just made. At the end of each reaction cycle, the free enzyme is regenerated.

Consider, for example, the mechanism of action of the serine proteases. The reaction they catalyze, the hydrolysis of a peptide bond, is greatly accelerated by the enzymes' affinity for the tetrahedral intermediate of the reaction. But a serine protease does more than a typical catalytic antibody: instead of waiting for an oxygen from a water molecule to attack the carbonyl carbon, it makes the reaction go much more quickly by first using a precisely positioned amino acid side chain for this purpose (the activated serine in Figure 3–57). This step breaks the peptide bond, but it leaves the enzyme covalently linked to the carboxyl group. Then, in a rapid second step, this covalent intermediate is destroyed by the enzyme-catalyzed addition of water, completing the reaction and regenerating the free enzyme (Figure 3–64). Even though this two-step reaction is less direct than a one-step reaction (in which water is added to the peptide bond), it is faster because each step has a relatively low activation energy.

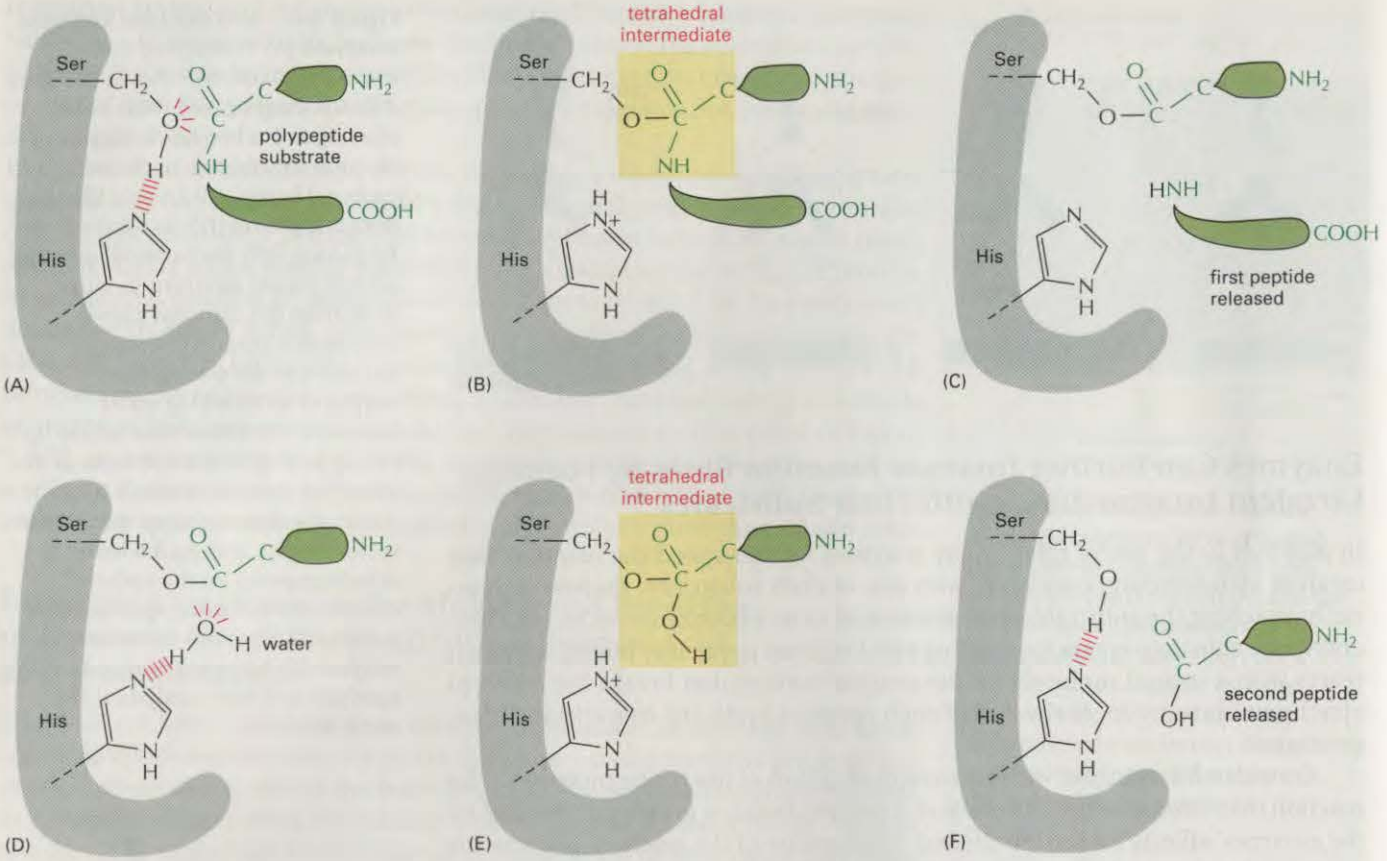
Enzymes Accelerate Chemical Reactions but Cannot Make Them Energetically More Favorable

No matter how sophisticated an enzyme is, it cannot make the chemical reaction it catalyzes either more or less energetically favorable. It cannot alter the free-energy difference between the initial substrates and the final products of the reaction. Like the simple binding interactions already discussed, any given chemical reaction has an *equilibrium point,* at which the backward and forward reaction fluxes are equal, so that no net change occurs (see Figure 3–9). If an enzyme speeds up the rate of the forward reaction, $A + B \rightarrow AB$, by a factor of 10^8, it must speed up the rate of the backward reaction, $AB \rightarrow A + B$, by a factor of 10^8 as well. The *ratio* of the forward to the backward rates of reaction depends only on the concentrations of A, B, and AB. The equilibrium point remains precisely the same whether or not the reaction is catalyzed by an enzyme.

Enzymes Determine Reaction Paths by Coupling Selected Reactions to ATP Hydrolysis [39]

The living cell is a chemical system that is far from equilibrium. The product of each enzyme usually serves as a substrate for another enzyme in the metabolic

Sarepta Exhibit 1040, page 50



pathway and is rapidly consumed. More important, by means of a reaction pathway that is determined by enzymes, many reactions are driven in one direction by being *coupled* to the energetically favorable hydrolysis of ATP to ADP and inorganic phosphate, as previously described in Chapter 2. To make this strategy effective, the ATP pool is itself maintained at a level far from its equilibrium point, with a high ratio of ATP to its hydrolysis products (discussed in Chapter 14). This ATP pool thereby serves as a "storage battery" that keeps energy and atoms continually passing through the cell directed along pathways determined by the enzymes present. For a living system, approaching chemical equilibrium means decay and death.

Multienzyme Complexes Help to Increase the Rate of Cell Metabolism [40]

The efficiency of enzymes in accelerating chemical reactions is crucial to the maintenance of life. Cells, in effect, must race against the unavoidable processes of decay, which run downhill toward chemical equilibrium. If the rates of desirable reactions were not greater than the rates of competing side reactions, a cell would soon die. Some idea of the rate at which cellular metabolism proceeds can be obtained by measuring the rate of ATP utilization. A typical mammalian cell turns over (that is, completely degrades and replaces) its entire ATP pool once every 1 or 2 minutes. For each cell this turnover represents the utilization of roughly 10^7 molecules of ATP per second (or, for the human body, about a gram of ATP every minute).

The rates of cellular reactions are rapid because of the effectiveness of enzyme catalysis. Many important enzymes have become so efficient that there is no possibility of further useful improvement: the factor limiting the reaction rate is no longer the intrinsic speed of action of the enzyme, rather it is the frequency

Figure 3–64 Some enzymes form covalent bonds with their substrates. In the example shown here, a carbonyl group in a polypeptide chain (shown in *green*) forms a covalent bond with a specially activated serine residue (see Figure 3–57) of a serine protease (shown in *gray*), which cleaves the polypeptide chain. When the unbound portion of the polypeptide chain has diffused away, a second step occurs in which a water molecule hydrolyzes the newly formed covalent bond, thereby releasing the portion of the polypeptide bound to the enzyme surface and freeing the serine for another cycle of reaction. Note that two unstable tetrahedral intermediates (shaded in *yellow*) serve as transition states in this reaction and both are stabilized by the enzyme.

Sarepta Exhibit 1040, page 51

with which the enzyme collides with its substrate. Such a reaction is said to be *diffusion-limited.*

If a reaction is diffusion-limited, its rate will depend on the concentration of both the enzyme and its substrate. For a sequence of reactions to occur very rapidly, each metabolic intermediate and enzyme involved must therefore be present in high concentration. Given the enormous number of different reactions carried out by a cell, there are limits to the concentrations of substrates that can be achieved. In fact, most metabolites are present in micromolar (10^{-6} M) concentrations, and most enzyme concentrations are much lower. How is it possible, therefore, to maintain very fast metabolic rates?

The answer lies in the spatial organization of cell components. Reaction rates can be increased without raising substrate concentrations by bringing the various enzymes involved in a reaction sequence together to form a large protein assembly known as a **multienzyme complex.** In this way the product of enzyme A is passed directly to enzyme B and so on to the final product, and diffusion rates need not be limiting even when the concentration of substrate in the cell as a whole is very low. Such enzyme complexes are very common (the structure of one, pyruvate dehydrogenase, was shown in Figure 2–41), and they are involved in nearly all aspects of metabolism, including the central genetic processes of DNA, RNA, and protein synthesis. In fact, it may be that few enzymes in eucaryotic cells diffuse freely in solution; instead, most may have evolved binding sites that concentrate them with other proteins of related function in particular regions of the cell, thereby increasing the rate and efficiency of the reactions that they catalyze.

Cells have another way of increasing the rate of metabolic reactions. It depends on the extensive intracellular membrane systems of eucaryotic cells. These membranes can segregate certain substrates and the enzymes that act on them into the same membrane-bounded compartment, such as the endoplasmic reticulum or the cell nucleus. If, for example, the compartment occupies a total of 10% of the volume of the cell, the concentration of reactants in the compartment can be 10 times greater than in a similar cell with no compartmentalization (Figure 3–65). Reactions that would otherwise be limited by the speed of diffusion can thereby be speeded up by the same factor.

Further details of protein structure and function will be presented in Chapter 5, where we discuss how cells construct tiny machines out of proteins.

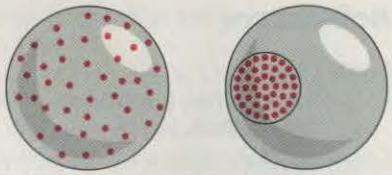


Figure 3–65 Compartmentalization. A large increase in the concentration of interacting molecules can be achieved by confining them to the same membrane-bounded compartment in a eucaryotic cell.

Summary

The biological function of a protein depends on the detailed chemical properties of its surface. Binding sites for ligands are formed as surface cavities in which precisely positioned amino acid side chains are brought together by protein folding. In this way, normally unreactive amino acid side chains can be activated. Enzymes greatly speed up reaction rates by binding the high-energy transition states in a reaction especially tightly; they also carry out acid catalysis and base catalysis simultaneously. The rates of enzyme reactions are often so fast that they are limited only by diffusion; rates can be further increased if enzymes that act sequentially on a substrate are joined into a single multienzyme complex or if the enzymes and their substrates are confined to the same compartment of the cell.

Sarepta Exhibit 1040, page 52

References

General

Branden, C.; Tooze, J. Introduction to Protein Structure. New York: Garland, 1991.

Gesteland, R.F.; Atkins, J.F., eds. The RNA World. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 1993.

Judson, H.F. The Eighth Day of Creation: Makers of the Revolution in Biology. New York: Simon & Schuster, 1979.

Lehninger, A.L.; Nelson, D.L.; Cox, M.M. Principles of Biochemistry, 2nd ed. New York: Worth, 1993.

Mathews, C.K.; van Holde, K.E. Biochemistry. Menlo Park, CA: Benjamin/Cummings, 1990.

Schulz, G.E.; Schirmer, R.H. Principles of Protein Structure. New York: Springer, 1979.

Stryer, L. Biochemistry, 3rd ed. New York: W.H. Freeman, 1988.

Voet, D.; Voet, J.G. Biochemistry. New York: Wiley, 1990.

Watson, J.D., et al. Molecular Biology of the Gene, 3rd ed. Menlo Park, CA: Benjamin-Cummings, 1987.

Cited

1. Cantor, C.R.; Schimmel, P.R. Biophysical Chemistry, Part I and Part III. New York: W.H. Freeman, 1980.

   Eisenberg, D.; Crothers, D. Physical Chemistry with Applications to the Life Sciences. Menlo Park, CA: Benjamin-Cummings, 1979.

   Pauling, L. The Nature of the Chemical Bond, 3rd ed. Ithaca, NY: Cornell University Press, 1960.

   Whitesides, G.M.; Mathias, J.P.; Seto, C.T. Molecular self-assembly and nanochemistry: a chemical strategy for the synthesis of nanostructures. *Science* 254:1312–1319, 1991.

2. Abeles, R.H.; Frey, P.A.; Jencks, W.P. Biochemistry. Boston: Jones and Bartlett, 1992.

   Burley, S.K.; Petsko, G.A. Weakly polar interactions in proteins. *Adv. Prot. Chem.* 39:125–189, 1988.

   Fersht, A.R. The hydrogen bond in molecular recognition. *Trends Biochem. Sci.* 12:301–304, 1987.

3. Cohen, C.; Parry, D.A.D. α-helical coiled coils—a widespread motif in proteins. *Trends Biochem. Sci.* 11:245–248, 1986.

   Dickerson, R.E. The DNA helix and how it is read. *Sci. Am.* 249(6):94–111, 1983.

4. Berg, H.C. Random Walks in Biology. Princeton, NJ: Princeton University Press, 1983.

   Einstein, A. Investigations on the Theory of Brownian Movement. New York: Dover, 1956.

   Lavenda, B.H. Brownian motion. *Sci. Am.* 252(2):70–85, 1985.

5. Lehninger, A.L. Bioenergetics: The Molecular Basis of Biological Energy Transformations, 2nd ed. Menlo Park, CA: Benjamin-Cummings, 1971.

6. Karplus, M.; McCammon, J.A. The dynamics of proteins. *Sci. Am.* 254(4):42–51, 1986.

   Karplus, M.; Petsko, G.A. Molecular dynamics simulations in biology. *Nature* 347:631–639, 1990.

   McCammon, J.A.; Harvey, S.C. Dynamics of Proteins and Nucleic Acids. Cambridge, UK: Cambridge University Press, 1987.

7. Kirkwood, T.B.; Rosenberger, R.F.; Galas, D.J., eds. Accuracy in Molecular Processes: Its Control and Relevance to Living Systems. London: Chapman and Hall, 1986.

8. Berg, P.; Singer, M. Dealing with Genes. The Language of Heredity. Mill Valley, CA: University Science Books, 1992.

   Rosenfield, I.; Ziff, E.; Van Loon, B. DNA for Beginners. London: Writers and Readers Publishing Cooperative. New York: Distributed in the USA by Norton, 1983.

   Saenger, W. Principles of Nucleic Acid Structure. Berlin: Springer, 1984.

9. Moore, J. Heredity and Development, 2nd ed. New York: Oxford University Press, 1992.

   Olby, R. The Path to the Double Helix. Seattle: University of Washington Press, 1974.

   Stent, G.S.; Calendar, A.Z. Molecular Genetics: An Introductory Narrative, 2nd ed. San Francisco: W.H. Freeman, 1978.

10. Watson, J.D.; Crick, F.H.C. Molecular structure of nucleic acids. A structure for deoxyribose nucleic acid. *Nature* 171:737–738, 1953.

11. Felsenfeld, G. DNA. *Sci. Am.* 253(4):58–66, 1985.

    Meselson, M.; Stahl, F.W. The replication of DNA in *E. coli*. *Proc. Natl. Acad. Sci. USA* 44:671–682, 1958.

    Watson, J.D.; Crick, F.H.C. Genetic implications of the structure of deoxyribonucleic acid. *Nature* 171:964–967, 1953.

12. Drake, J.W. Spontaneous mutation. *Annu. Rev. Genet.* 25:125–146, 1991.

    Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* 362:709–715, 1993.

    Wilson, A.C. Molecular basis of evolution. *Sci. Am.* 253(4):164–173, 1985.

13. Sanger, F. Sequences, sequences, and sequences. *Annu. Rev. Biochem.* 57:1–28, 1988.

    Thompson, E.O.P. The insulin molecule. *Sci. Am.* 192(5):36–41, 1955.

    Yanofsky, C. Gene structure and protein structure. *Sci. Am.* 216(5):80–94, 1967.

14. Brenner, S.; Jacob, F.; Meselson, M. An unstable intermediate carrying information from genes to ribosomes for protein synthesis. *Nature* 190:576–581, 1961.

    Darnell, J.E., Jr. RNA. *Sci. Am.* 253(4):68–78, 1985.

15. Chambon, P. Split genes. *Sci. Am.* 244(5):60–71, 1981.

    Steitz, J.A. Snurps. *Sci. Am.* 258(6):58–63, 1988.

    Witkowski, J.A. The discovery of "split" genes: a scientific revolution. *Trends Biochem. Sci.* 13:110–113, 1988.

16. Crick, F.H.C. The genetic code: III. *Sci. Am.* 215(4):55–62, 1966.

    The Genetic Code. *Cold Spring Harbor Symp. Quant. Biol.* 31, 1965.

17. Rich, A.; Kim, S.H. The three-dimensional structure of transfer RNA. *Sci. Am.* 238(1):52–62, 1978.

18. Lake, J.A. The ribosome. *Sci. Am.* 245(2):84–97, 1981.

    Watson, J.D. Involvement of RNA in the synthesis of proteins. *Science* 140:17–26, 1963.

    Zamecnik, P. The machinery of protein synthesis. *Trends Biochem. Sci.* 9:464–466, 1984.

19. Altman, S.; Baer, M.; Guerrier-Takada, C.; Viogue, A. Enzymatic cleavage of RNA by RNA. *Trends Biochem. Sci.* 11:515–518, 1986.

    Cech, T. RNA as an enzyme. *Sci. Am.* 255(5):64–75, 1986.

Sarepta Exhibit 1040, page 53

Cech, T. Fishing for fresh catalysts. *Nature* 365:204–205, 1993.

Michel, F.; Westhof, E. Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J. Mol. Biol.* 216:585–610, 1990.

Noller, H.F. Ribosomal RNA and translation. *Annu. Rev. Biochem.* 60:191–227, 1991.

Noller, H.F.; Hoffarth, V.; Zimniak, L. Unusual resistance of peptidyl transferase to protein extraction procedures. *Science* 256:1416–1419, 1992.

20. Branden, C.; Tooze, J. Introduction to Protein Structure. New York: Garland, 1991.

Creighton, T.E. Proteins: Structure and Molecular Properties, 2nd ed. New York: W.H. Freeman, 1993.

Dickerson, R.E.; Geis, I. The Structure and Action of Proteins. New York: Harper & Row, 1969.

Schulz, G.E.; Schirmer, R.H. Principles of Protein Structure. New York: Springer, 1990.

21. Anfinsen, C.B. Principles that govern the folding of protein chains. *Science* 181:223–230, 1973.

Baldwin, R.L. Seeding protein folding. *Trends Biochem. Sci.* 11:6–9, 1986.

Creighton, T.E. Disulphide bonds and protein stability. *Bioessays* 8:57–63, 1988.

Richards, F.M. The protein folding problem. *Sci. Amer.* 264(1):54–63, 1991.

Rupley, J.A.; Gratton, E.; Careri, G. Water and globular proteins. *Trends Biochem. Sci.* 8:18–22, 1983.

22. Doolittle, R.F. Proteins. *Sci. Am.* 253(4):88–99, 1985.

Milner-White, E.J.; Poet, R. Loops, bulges, turns and hairpins in proteins. *Trends Biochem Sci.* 12:189–192, 1987.

Pauling, L.; Corey, R.B. Configurations of polypeptide chains with favored orientations around single bonds: two new pleated sheets. *Proc. Natl. Acad. Sci. USA* 37:729–740, 1951.

Pauling, L.; Corey, R.B.; Branson, H.R. The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain. *Proc. Natl. Acad. Sci. USA* 27:205–211, 1951.

Richardson, J.S. The anatomy and taxonomy of protein structure. *Adv. Protein Chem.* 34:167–339, 1981.

23. Scott, J.E. Molecules for strength and shape. *Trends Biochem. Sci.* 12:318–321, 1987.

24. Branden, C.; Tooze, J. Introduction to Protein Structure. New York: Garland, 1991.

Hardie, D.G.; Coggins, J.R. Multidomain Proteins—Structure and Evolution. Amsterdam: Elsevier, 1986.

25. Doolittle, R.F. The genealogy of some recently evolved vertebrate proteins. *Trends Biochem. Sci.* 10:233–237, 1985.

Neurath, H. Evolution of proteolytic enzymes. *Science* 224:350–357, 1984.

26. Blake, C. Exons and the evolution of proteins. *Trends Biochem. Sci.* 8:11–13, 1983.

Gilbert, W. Genes-in-pieces revisited. *Science* 228:823–824, 1985.

McCarthy, A.D.; Hardie, D.G. Fatty acid synthase: an example of protein evolution by gene fusion. *Trends Biochem. Sci.* 9:60–63, 1984.

Rossmann, M.G.; Argos, P. Protein folding. *Annu. Rev. Biochem.* 50:497–532, 1981.

27. Gehring, W.J. On the homeobox and its significance. *Bioessays* 5:3–4, 1986.

Hanks, S.K.; Quinn, A.M.; Hunter, T. The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. *Science* 241:42–52, 1988.

Shabb, J.B.; Corbin, J.D. Cyclic nucleotide-binding domains in proteins having diverse functions. *J. Biol. Chem.* 267:5723–5726, 1992.

Thornton, J.M.; Gardner, S.P. Protein motifs and database searching. *Trends Biochem. Sci.* 14:300–304, 1989.

28. Bajaj, M.; Blundell, T. Evolution and the tertiary structure of proteins. *Annu. Rev. Biophys. Bioeng.* 13:453–492, 1984.

Klug, A. From macromolecules to biological assemblies. *Biosci. Rep.* 3:395–430, 1983.

Metzler, D.E. Biochemistry. New York: Academic Press, 1977. (Chapter 4 describes how macromolecules pack together into large assemblies.)

29. Caspar, D.L.D.; Klug, A. Physical principles in the construction of regular viruses. *Cold Spring Harbor Symp. Quant. Biol.* 27:1–24, 1962.

Goodsell, D.S. Inside a living cell. *Trends Biochem. Sci.* 16:203–206, 1991.

30. Cohen, C.; Parry, D.A.D. α-helical coiled coils—a widespread motif in proteins. *Trends Biochem. Sci.* 11:245–248, 1986.

31. Harrison, S.C. Multiple modes of subunit association in the structures of simple spherical viruses. *Trends Biochem. Sci.* 9:345–351, 1984.

Harrison, S.C. Viruses. *Curr. Opin. Struc. Biol.* 2:293–299, 1992.

Hogle, J.M.; Chow, M.; Filman, D.J. The structure of polio virus. *Sci. Am.* 256(3):42–49, 1987.

Rossmann, M.G.; Johnson, J.E. Icosahedral RNA virus structure. *Annu. Rev. Biochem.* 58:533–573, 1989.

32. Fraenkel-Conrat, H.; Williams, R.C. Reconstitution of active tobacco mosaic virus from its inactive protein and nucleic acid components. *Proc. Natl. Acad. Sci. USA* 41:690–698, 1955.

Hendrix, R.W. Tail length determination in double-stranded DNA bacteriophages. *Curr. Top. Microbiol. Immunol.* 136:21–29, 1988.

Namba, K.; Caspar, D.L.D.; Stubbs, G.J. Computer graphics representation of levels of organization in tobacco mosaic virus structure. *Science* 227:773–776, 1985.

Nomura, M. Assembly of bacterial ribosomes. *Science* 179:864–873, 1973.

Trinick, J. Understanding the functions of titin and nebulin. *FEBS Lett.* 307:44–48, 1992.

33. Mathews, C.K., Kutter, E.M.; Mosig, G.; Berget, P.B. Bacteriophage T4, Chaps. 1 and 4. Washington, DC: American Society of Microbiologists, 1983.

Steiner, D.F.; Kemmler, W.; Tager, H.S.; Peterson, J.D. Proteolytic processing in the biosynthesis of insulin and other proteins. *Fed. Proc.* 33:2105–2115, 1974.

34. Dressler, D.; Potter, H. Discovering Enzymes. New York: Scientific American Library, 1991.

Fersht, A. Enzyme Structure and Mechanism, 2nd ed. New York: W.H. Freeman, 1985.

35. Fersht, A.R.; Leatherbarrow, R.J.; Wells, T.N.C. Binding energy and catalysis. *Trends Biochem. Sci.* 11:321–325, 1986.

Sarepta Exhibit 1040, page 54

Hansen, D.E.; Raines, R.T. Binding energy and enzymatic catalysis. *J. Chem. Educ.* 67:483–489, 1990.

36. Lerner, R.A.; Benkovic, S.J.; Schultz, P.G. At the crossroads of chemisry and immunology: catalytic antibodies. *Science* 252:659–667, 1991.

Lerner, R.A.; Tramontano, A. Catalytic antibodies. *Sci. Am.* 258(3):58–70, 1988.

Wolfenden, R. Analog approaches to the structure of the transition state in enzyme reactions. *Accounts Chem. Res.* 5:10–18, 1972.

37. Knowles, J.R. Tinkering with enzymes: what are we learning? *Science* 236:1252–1258, 1987.

Knowles, J.R. Enzyme catalysis: not different, just better. *Nature* 350:121–124, 1991.

38. Stroud, R.M. A family of protein-cutting proteins. *Sci. Am.* 231(1):74–88, 1974.

39. Wood, W.B.; Wilson, J.H.; Benbow, R.M.; Hood, L.E. Biochemistry, A Problems Approach, 2nd ed. Menlo Park, CA: Benjamin-Cummings, 1981. (Chapters 9 and 15 and associated problems.)

40. Barnes, S.J.; Weitzman, P.D.J. Organization of citric acid cycle enzymes into a multienzyme cluster. *FEBS Lett.* 201:267–270, 1986.

Berg, O.G.; von Hippel, P.H. Diffusion-controlled macromolecular interactions. *Annu. Rev. Biophys. Biophys. Biochem.* 14:131–160, 1985.

Reed, L.J.; Cox, D.J. Multienzyme complexes. In The Enzymes, 3rd ed. (P.D. Boyer, ed.), Vol. 1, pp. 213–240. New York: Academic Press, 1970.

Sarepta Exhibit 1040, page 55

MOLECULAR BIOLOGY OF
THE CELL
THIRD EDITION

Bruce Alberts • Dennis Bray
Julian Lewis • Martin Raff • Keith Roberts
James D. Watson



Garland Publishing, Inc.
New York & London

Sarepta Exhibit 1040, page 56

GARLAND STAFF
Text Editor: Miranda Robertson
Managing Editor: Ruth Adams
Illustrator: Nigel Orme
Molecular Model Drawings: Kate Hesketh-Moore
Director of Electronic Publishing: John M-Roblin
Computer Specialist: Chuck Bartelt
Disk Preparation: Carol Winter
Copy Editor: Shirley M. Cobert
Production Editor: Douglas Goertzen
Production Coordinator: Perry Bessas
Indexer: Maija Hinkle

Bruce Alberts received his Ph.D. from Harvard University and is currently President of the National Academy of Sciences and Professor of Biochemistry and Biophysics at the University of California, San Francisco. *Dennis Bray* received his Ph.D. from the Massachusetts Institute of Technology and is currently a Medical Research Council Fellow in the Department of Zoology, University of Cambridge. *Julian Lewis* received his D.Phil. from the University of Oxford and is currently a Senior Scientist in the Imperial Cancer Research Fund Developmental Biology Unit, University of Oxford. *Martin Raff* received his M.D. from McGill University and is currently a Professor in the MRC Laboratory for Molecular Cell Biology and the Biology Department, University College London. *Keith Roberts* received his Ph.D. from the University of Cambridge and is currently Head of the Department of Cell Biology, the John Innes Institute, Norwich. *James D. Watson* received his Ph.D. from Indiana University and is currently Director of the Cold Spring Harbor Laboratory. He is the author of *Molecular Biology of the Gene* and, with Francis Crick and Maurice Wilkins, won the Nobel Prize in Medicine and Physiology in 1962.

© 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 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.].—3rd ed.
p. cm.
Includes bibliographical references and index.
ISBN 0-8153-1619-4 (hard cover).—ISBN 0-8153-1620-8 (pbk.)
1. Cytology. 2. Molecular biology. I. Alberts, Bruce.
[DNLM: 1. Cells. 2. Molecular Biology. QH 581.2 M718 1994]
QH581.2.M64 1994
574.87—dc20
DNLM/DLC
for Library of Congress
93-45907
CIP

Published by Garland Publishing, Inc.
717 Fifth Avenue, New York, NY 10022

Printed in the United States of America

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

Front cover: The photograph shows a rat nerve cell in culture. It is labeled (*yellow*) with a fluorescent antibody that stains its cell body and dendritic processes. Nerve terminals (*green*) from other neurons (not visible), which have made synapses on the cell, are labeled with a different antibody. (Courtesy of Olaf Mundigl and Pietro de Camilli.)

Dedication page: Gavin Borden, late president of Garland Publishing, weathered in during his mid-1980s climb near Mount McKinley with MBoC author Bruce Alberts and famous mountaineer guide Mugs Stump (1940–1992).

Back cover: The authors, in alphabetical order, crossing Abbey Road in London on their way to lunch. Much of this third edition was written in a house just around the corner. (Photograph by Richard Olivier.)

Basic Genetic Mechanisms

6

- RNA and Protein Synthesis
- DNA Repair
- DNA Replication
- Genetic Recombination
- Viruses, Plasmids, and Transposable Genetic Elements

The ability of cells to maintain a high degree of order in a chaotic universe depends on the genetic information that is expressed, maintained, replicated, and occasionally improved by the basic genetic processes—*RNA and protein synthesis, DNA repair, DNA replication,* and *genetic recombination.* In these processes, which produce and maintain the proteins and nucleic acids of a cell (Figure 6–1), the information in a linear sequence of nucleotides is used to specify either another linear chain of nucleotides (a DNA or an RNA molecule) or a linear chain of amino acids (a protein molecule). The framework underlying genetic events is therefore one-dimensional and conceptually simple. In contrast, most other processes in cells result solely from information expressed in the complex three-dimensional surfaces of protein molecules. Perhaps that is why we understand more about genetic mechanisms than about most other biological processes.

In this chapter we examine the molecular machinery that repairs, replicates, and alters on occasion the DNA of the cell. We shall see that the machinery depends on enzymes that cut, copy, and recombine nucleotide sequences. We shall also see that these and other enzymes can be parasitized by viruses, plasmids, and transposable genetic elements, which not only direct their own replication, but also can alter the cell genome by genetic recombination events.

First, however, we reconsider a central topic mentioned briefly in Chapter 3—the mechanisms of RNA and protein synthesis.

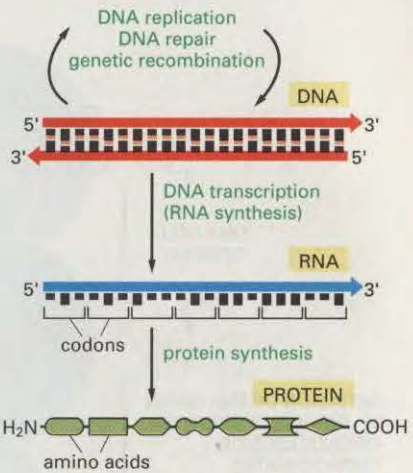


Figure 6–1 The basic genetic processes. The processes shown here are thought to occur in all present-day cells. Very early in the evolution of life, however, much simpler cells probably existed that lacked both DNA and proteins (see Figure 1–11). Note that a sequence of three nucleotides (a codon) in an RNA molecule codes for a specific amino acid in a protein.

RNA and Protein Synthesis

Proteins constitute more than half the total dry mass of a cell, and their synthesis is central to cell maintenance, growth, and development. Protein synthesis occurs on *ribosomes.* It depends on the collaboration of several classes of RNA molecules and begins with a series of preparatory steps. First, a molecule of *messenger RNA (mRNA)* must be copied from the DNA that encodes the protein. Meanwhile, in the cytoplasm, each of the 20 amino acids from which the protein is to be built must be attached to its specific *transfer RNA (tRNA)* molecule, and the subunits of the ribosome on which the new protein is to be made must be preloaded with auxiliary protein factors. Protein synthesis begins when all of these components come together in the cytoplasm to form a functioning ribosome. As a single molecule of mRNA moves stepwise through a ribosome, the sequence of nucleotides in the mRNA molecule is translated into a corresponding sequence of amino acids to produce a distinctive protein chain, as specified by the DNA sequence of

Sarepta Exhibit 1040, page 58

its gene. We begin by considering how the many different RNA molecules in a cell are made.

RNA Polymerase Copies DNA into RNA: The Process of DNA Transcription [1]

RNA is synthesized on a DNA template by a process known as **DNA transcription.** Transcription generates the mRNAs that carry the information for protein synthesis, as well as the transfer, ribosomal, and other RNA molecules that have structural or catalytic functions. All of these RNA molecules are synthesized by **RNA polymerase** enzymes, which make an RNA copy of a DNA sequence. In eucaryotes three kinds of RNA polymerase molecules synthesize different types of RNA, as described in Chapter 8. These RNA polymerases are thought to have derived during evolution from the single enzyme present in bacteria that mediates all bacterial RNA synthesis.

The bacterial RNA polymerase is a large multisubunit enzyme associated with several additional protein subunits that enter and leave the polymerase-DNA complex at different stages of transcription. Free RNA polymerase molecules collide randomly with the bacterial chromosome, sliding along it but sticking only weakly to most DNA. The polymerase binds very tightly, however, when it con-

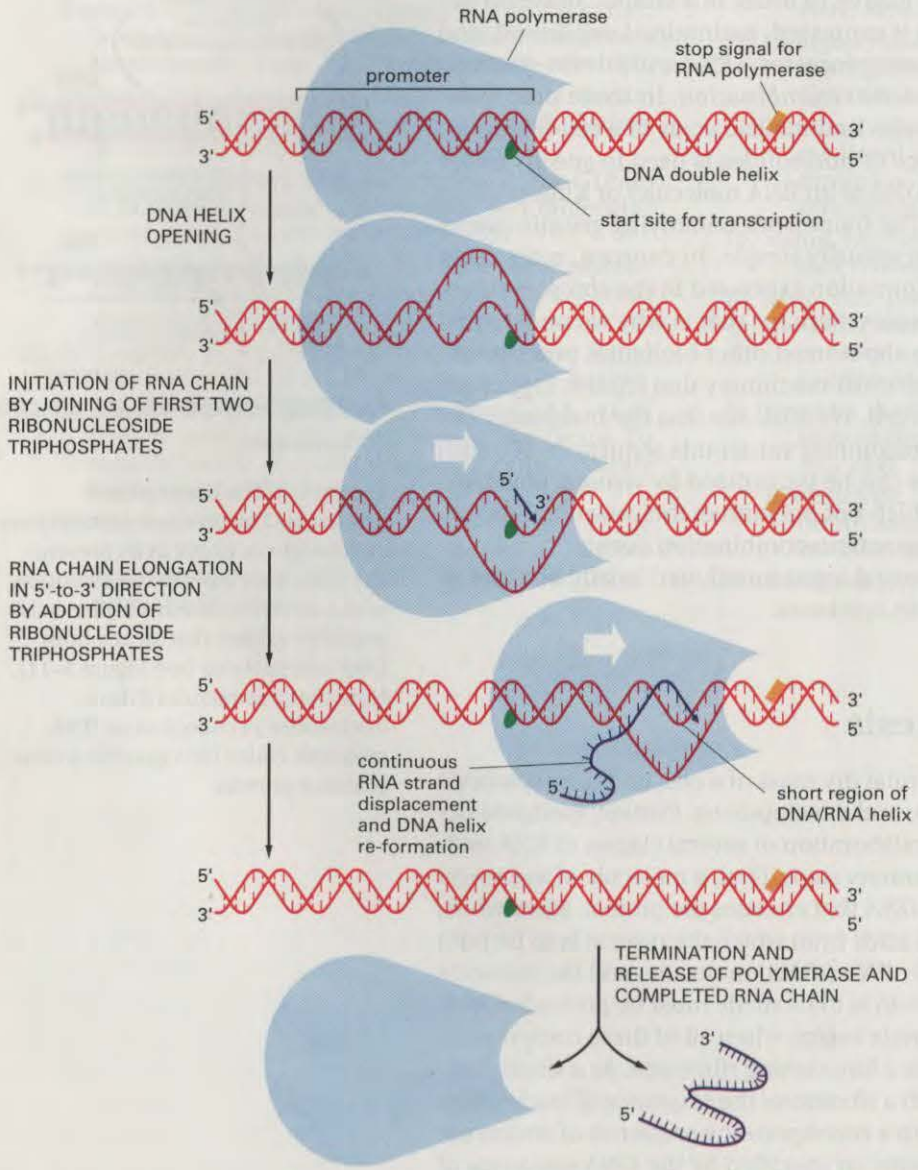


Figure 6–2 The synthesis of an RNA molecule by RNA polymerase. The enzyme binds to the promoter sequence on the DNA and begins its synthesis at a start site within the promoter. It completes its synthesis at a stop (termination) signal, whereupon both the polymerase and its completed RNA chain are released. During RNA chain elongation, polymerization rates average about 30 nucleotides per second at 37°C. Therefore, an RNA chain of 5000 nucleotides takes about 3 minutes to complete.

Sarepta Exhibit 1040, page 59

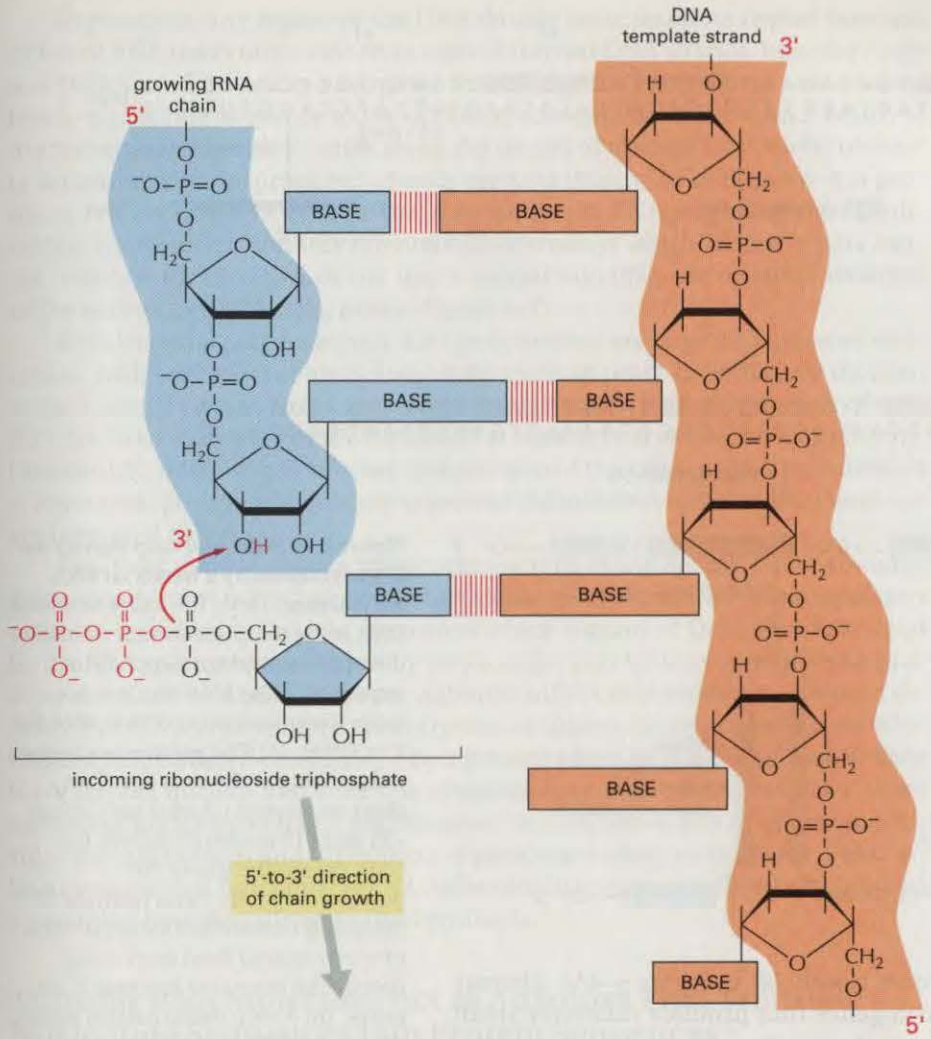


Figure 6–3 The chain elongation reaction catalyzed by an RNA polymerase enzyme. In each step an incoming ribonucleoside triphosphate is selected for its ability to base-pair with the exposed DNA template strand; a ribonucleoside monophosphate is then added to the growing, 3′-OH end of the RNA chain (*red arrow*), and pyrophosphate is released (*red atoms*). The new RNA chain therefore grows by one nucleotide at a time in the 5′-to-3′ direction, and it is complementary in sequence to the DNA template strand. The reaction is driven both by the favorable free-energy change that accompanies the release of pyrophosphate and by the subsequent hydrolysis of the pyrophosphate to inorganic phosphate (see Figure 2–30).

tacts a specific DNA sequence, called the **promoter,** that contains the *start site* for RNA synthesis and signals where RNA synthesis should begin. The reactions that ensue are outlined in Figure 6–2. After binding to the promoter, the RNA polymerase opens up a local region of the double helix to expose the nucleotides on a short stretch of DNA on each strand. One of the two exposed DNA strands acts as a template for complementary base-pairing with incoming ribonucleoside triphosphate monomers, two of which are joined together by the polymerase to begin an RNA chain. The RNA polymerase molecule then moves stepwise along the DNA, unwinding the DNA helix just ahead to expose a new region of the template strand for complementary base-pairing. In this way the growing RNA chain is extended by one nucleotide at a time in the 5′-to-3′ direction (Figure 6–3). The chain elongation process continues until the enzyme encounters a second special sequence in the DNA, the **stop (termination) signal,** where the polymerase halts and releases both the DNA template and the newly made RNA chain.

By convention, when a DNA sequence associated with a gene is specified, it is the sequence of the nontemplate strand that is given, and it is written in the 5′-to-3′ direction. This convention is adopted because the sequence of the nontemplate strand corresponds to the sequence of the RNA that is made.

Nucleotide sequences that act as start sites and stop signals for the bacterial RNA polymerase are illustrated in Figure 6–4. Nucleotide sequences that are found in many examples of a particular type of region in DNA (such as a promoter) are called **consensus sequences.** In bacteria strong promoters (those associated with genes that produce large amounts of mRNA) have sequences that

Sarepta Exhibit 1040, page 60

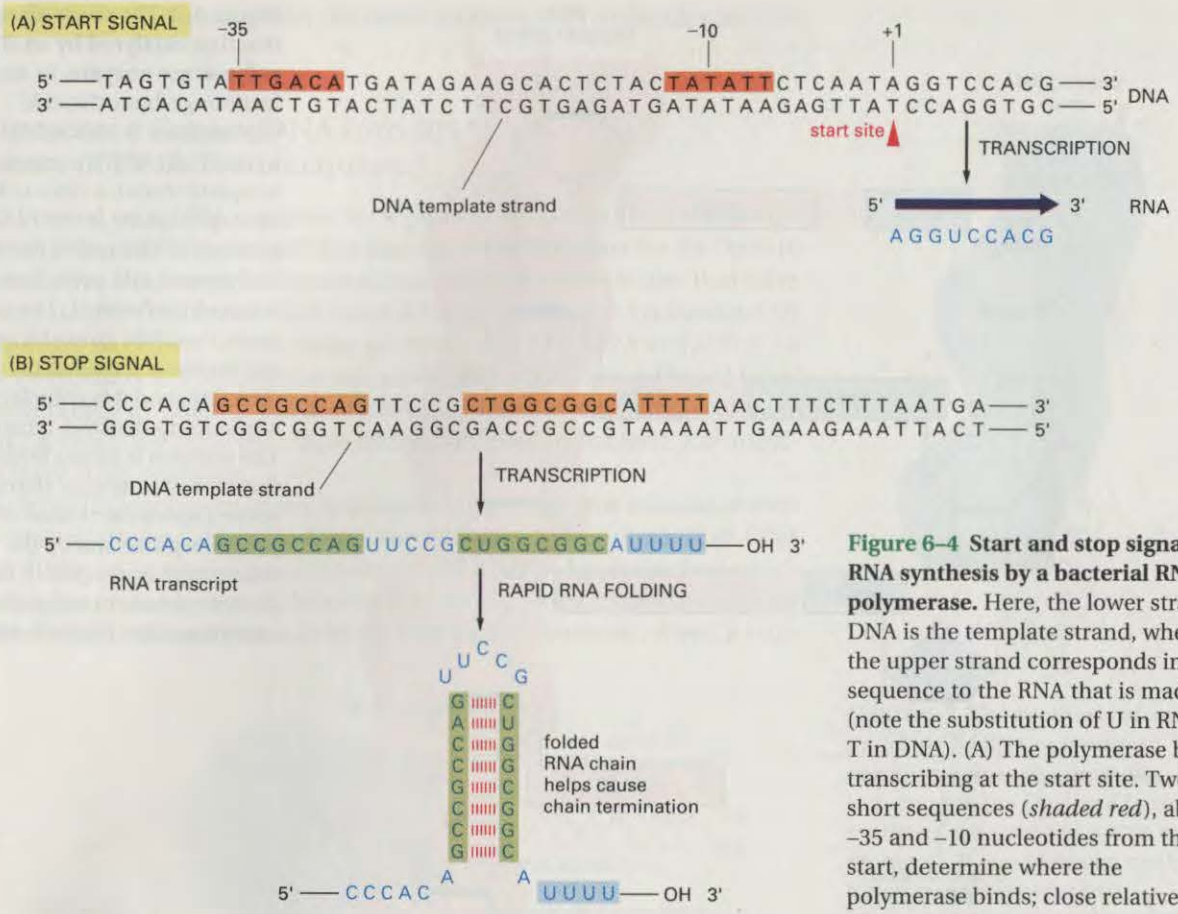


Figure 6–4 Start and stop signals for RNA synthesis by a bacterial RNA polymerase. Here, the lower strand of DNA is the template strand, whereas the upper strand corresponds in sequence to the RNA that is made (note the substitution of U in RNA for T in DNA). (A) The polymerase begins transcribing at the start site. Two short sequences (*shaded red*), about −35 and −10 nucleotides from the start, determine where the polymerase binds; close relatives of these two hexanucleotide sequences, properly spaced from each other, specify the promoter for most *E. coli* genes. (B) A stop (termination) signal. The *E. coli* RNA polymerase stops when it synthesizes a run of U residues (*shaded blue*) from a complementary run of A residues on the template strand, provided that it has just synthesized a self-complementary RNA nucleotide sequence (*shaded green*), which rapidly forms a hairpin helix that is crucial for stopping transcription. The sequence of nucleotides in the self-complementary region can vary widely.

match the promoter consensus sequences closely (as in Figure 6–4A), whereas weak promoters (those associated with genes that produce relatively small amounts of mRNA) match these sequences less well.

Only Selected Portions of a Chromosome Are Used to Produce RNA Molecules [2]

As an RNA polymerase molecule moves along the DNA, an RNA/DNA double helix is formed at the enzyme's active site. This helix is very short because the RNA just made is displaced, allowing the DNA/DNA helix immediately at the rear of the polymerase to rewind (Figure 6–5). As a result, each completed RNA chain is released from the DNA template as a free, single-stranded RNA molecule, typically between 70 and 10,000 nucleotides long.

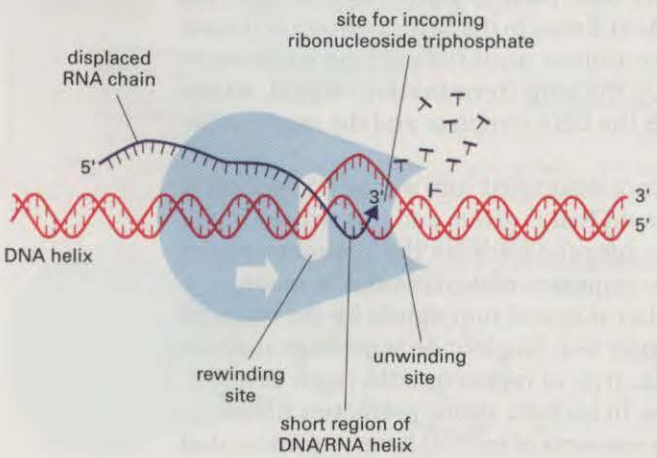


Figure 6–5 DNA unwinding and rewinding by RNA polymerase. A moving RNA polymerase molecule is continuously unwinding the DNA helix ahead of the polymerization site while rewinding the two DNA strands behind this site to displace the newly formed RNA chain. A short region of DNA/RNA helix is therefore formed only transiently, and the final RNA product is released as a single-stranded copy of one of the two DNA strands.

Sarepta Exhibit 1040, page 61

In principle, any region of the DNA double helix could be copied into two different RNA molecules—one from each of the two DNA strands. In reality, only one DNA strand is used as a template in each region. The RNA made is equivalent in nucleotide sequence to the opposite, nontemplate DNA strand. Which of the two strands is copied varies along the length of a single DNA molecule and is determined by the promoter of each gene. As illustrated in Figure 6–4, a promoter is an oriented DNA sequence that points the RNA polymerase in one direction or the other, and this orientation determines which DNA strand is copied (Figure 6–6). The DNA strand that is copied into RNA can be either different or the same for neighboring genes (Figure 6–7).

Both bacterial and eucaryotic RNA polymerases are large, complicated molecules, with multiple subunits and a total mass of more than 500,000 daltons. Some bacterial viruses, in contrast, encode single-chain RNA polymerases of one-fifth this mass that catalyze RNA synthesis at least as well as the host-cell enzyme. Presumably, the multiple subunit composition of the cellular RNA polymerases is important for various regulatory aspects of cellular RNA synthesis that have not yet been well defined.

This brief outline of DNA transcription omits many details. Other complex steps usually must occur before an mRNA molecule is produced. *Gene regulatory proteins,* for example, help to determine which regions of DNA are transcribed by the RNA polymerase and thereby play a major part in determining which proteins are made by a cell. Moreover, although mRNA molecules are produced directly by DNA transcription in procaryotes, in higher eucaryotic cells most RNA transcripts are altered extensively—by a process called *RNA splicing*—before they leave the cell nucleus and enter the cytoplasm as mRNA molecules. All of these aspects of mRNA production are discussed in Chapters 8 and 9, where we consider the cell nucleus and the control of gene expression, respectively. For now, let us assume that functional mRNA molecules have been produced and proceed to examine how they direct protein synthesis.

Transfer RNA Molecules Act as Adaptors That Translate Nucleotide Sequences into Protein Sequences [3]

All cells contain a set of **transfer RNAs (tRNAs),** each of which is a small RNA molecule (most have a length between 70 and 90 nucleotides). The tRNAs, by binding at one end to a specific codon in the mRNA and at their other end to the amino acid specified by that codon, enable amino acids to line up according to the sequence of nucleotides in the mRNA. Each tRNA is designed to carry only one of the 20 amino acids used for protein synthesis: a tRNA that carries glycine is designated $tRNA^{Gly}$ and so on. Each of the 20 amino acids has at least one type of tRNA assigned to it, and most have several tRNAs. Before an amino acid is incorporated into a protein chain, it is attached by its carboxyl end to the 3′ end of an appropriate tRNA molecule. This attachment serves two purposes. First, and most important, it covalently links the amino acid to a tRNA containing the correct **anticodon**—the sequence of three nucleotides that is complementary to the three-nucleotide *codon* that specifies that amino acid on an mRNA molecule. Codon-anticodon pairings enable each amino acid to be inserted into a growing protein chain according to the dictates of the sequence of nucleotides in the mRNA, thereby allowing the genetic code to be used to translate nucleotide se-

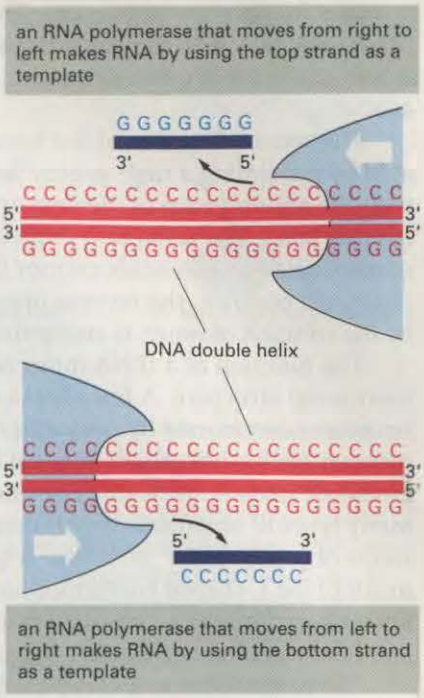


Figure 6–6 RNA polymerase orientation determines which DNA strand serves as template. The DNA strand serving as template must be traversed from its 3′ end to its 5′ end, as illustrated in Figure 6–3. Thus the direction of RNA polymerase movement determines which of the two DNA strands will serve as a template for the synthesis of RNA, as shown here. Polymerase direction is, in turn, determined by the orientation of the promoter sequence, where the RNA polymerase initially binds.

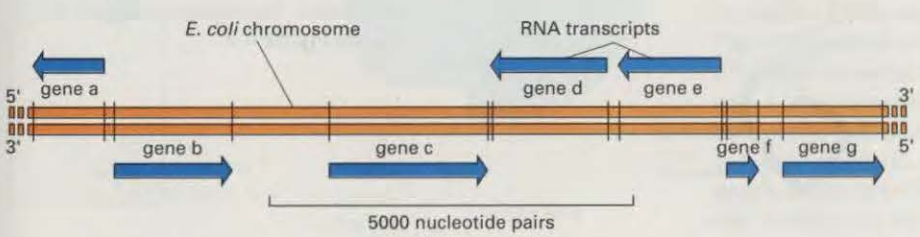


Figure 6–7 Directions of transcription along a short portion of a bacterial chromosome. Note that some genes are transcribed from one DNA strand, while others are transcribed from the other DNA strand. Approximately 0.2% of the *E. coli* chromosome is depicted here. (Adapted from D.L. Daniels et al., *Science* 257:771–777, 1992.)

Sarepta Exhibit 1040, page 62

quences into protein sequences. This is the essential "adaptor" function of the tRNA molecule: with one end attached to an amino acid and the other paired to a codon, the tRNA converts sequences of nucleotides into sequences of amino acids.

The second function of the amino acid attachment is to activate the amino acid by generating a high-energy linkage at its carboxyl end so that it can react with the amino group of the next amino acid in the protein sequence to form a *peptide bond.* The activation process is necessary for protein synthesis because nonactivated amino acids cannot be added directly to a growing polypeptide chain. (In contrast, the reverse process, in which a peptide bond is hydrolyzed by the addition of water, is energetically favorable and can occur spontaneously.)

The function of a tRNA molecule depends on its precisely folded three-dimensional structure. A few tRNAs have been crystallized and their complete structures determined by x-ray diffraction analyses. Both intramolecular complementary base-pairings and unusual base interactions are required to fold a tRNA molecule (see Figure 3–18). The nucleotide sequences of tRNA molecules from many types of organisms reveal that tRNAs can form the loops and base-paired stems of a "cloverleaf" structure (Figure 6–8), and all are thought to fold further to adopt the L-shaped conformation detected in crystallographic analyses. In the native structure the amino acid is attached to one end of the "L," while the anticodon is located at the other (Figure 6–9).

The nucleotides in a completed nucleic acid chain (like the amino acids in proteins) can be covalently modified to modulate the biological activity of the nucleic acid molecule. Such posttranscriptional modifications are especially common in tRNA molecules, which contain a variety of modified nucleotides (Figure 6–10). Some of the modified nucleotides affect the conformation and base-pairing of the anticodon and thereby facilitate the recognition of the appropriate mRNA codon by the tRNA molecule.

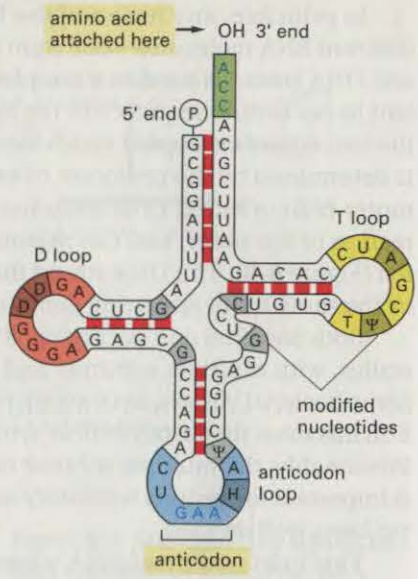


Figure 6–8 The "cloverleaf" structure of tRNA. This is a view of the molecule shown in Figure 6–9 after it has been partially unfolded. There are many different tRNA molecules, including at least one for each kind of amino acid. Although they differ in nucleotide sequence, they all have the three stem loops shown plus an amino acid-accepting arm. The particular tRNA molecule shown binds phenylalanine and is therefore denoted $tRNA^{Phe}$. In all tRNA molecules the amino acid is attached to the A residue of a CCA sequence at the 3′ end of the molecule. Complementary base-pairings are shown by *red bars.*

Specific Enzymes Couple Each Amino Acid to Its Appropriate tRNA Molecule [4]

Only the tRNA molecule, and not its attached amino acid, determines where the amino acid is added during protein synthesis. This was established by an ingenious experiment in which an amino acid (cysteine) was chemically converted into a different amino acid (alanine) after it was already attached to its specific tRNA. When such "hybrid" tRNA molecules were used for protein synthesis in a cell-free system, the wrong amino acid was inserted at every point in the protein chain where that tRNA was used. Thus the accuracy of protein synthesis is crucially dependent on the accuracy of the mechanism that normally links each activated amino acid specifically to its corresponding tRNA molecules.

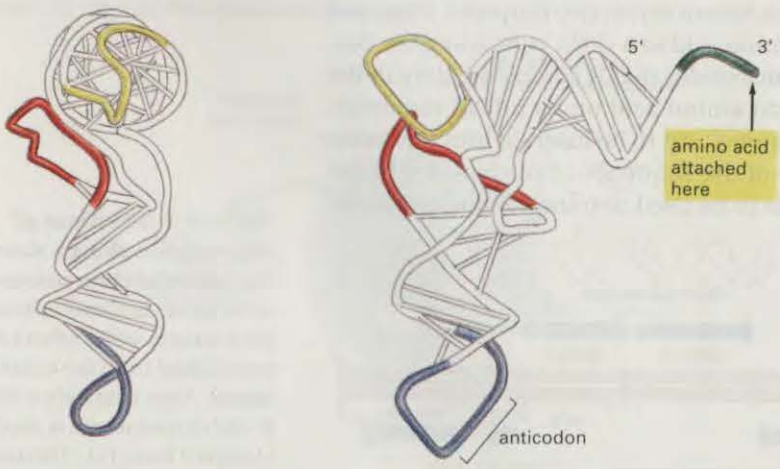
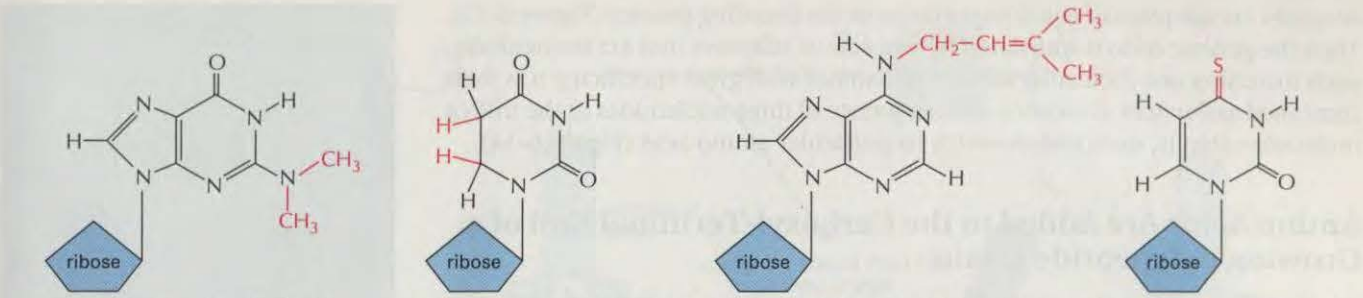


Figure 6–9 The folded structure of a typical tRNA molecule. Two views of the three-dimensional conformation determined by x-ray diffraction are shown. Note that the molecule is L-shaped; one end is designed to accept the amino acid, while the other end contains the three nucleotides of the anticodon. Each loop is colored to match Figure 6–8.

Sarepta Exhibit 1040, page 63



How does a tRNA molecule become covalently linked to the one amino acid in 20 that is its appropriate partner? The mechanism depends on enzymes called **aminoacyl-tRNA synthetases,** which couple each amino acid to its appropriate set of tRNA molecules. There is a different synthetase enzyme for every amino acid (20 synthetases in all): one attaches glycine to all $tRNA^{Gly}$ molecules, another attaches alanine to all $tRNA^{Ala}$ molecules, and so on. The coupling reaction that creates an **aminoacyl-tRNA** molecule is catalyzed in two steps, as illustrated in Figure 6–11. The structure of the amino acid-RNA linkage is shown in Figure 6–12.

Although the tRNA molecules serve as the final adaptors in converting nucleotide sequences into amino acid sequences, the aminoacyl-tRNA synthetase

Figure 6–10 A few of the unusual nucleotides found in tRNA molecules. These nucleotides are produced by covalent modification of a normal nucleotide after it has been incorporated into an RNA chain. In most tRNA molecules about 10% of the nucleotides are modified (see Figure 6–8).

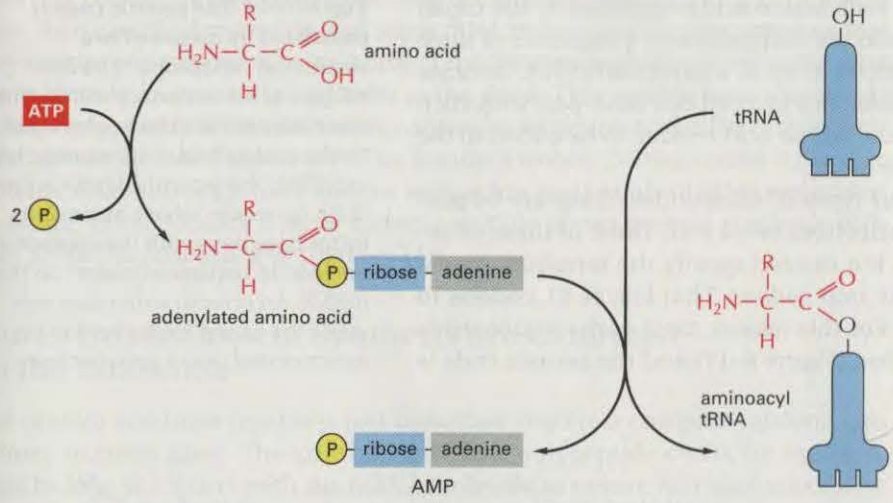


Figure 6–11 Amino acid activation. The two-step process in which an amino acid (with its side chain denoted by R) is activated for protein synthesis by an aminoacyl-tRNA synthetase enzyme is shown. As indicated, the energy of ATP hydrolysis is used to attach each amino acid to its tRNA molecule in a high-energy linkage. The amino acid is first activated through the linkage of its carboxyl group directly to an AMP moiety, forming an *adenylated amino acid;* the linkage of the AMP, normally an unfavorable reaction, is driven by the hydrolysis of the ATP molecule that donates the AMP. Without leaving the synthetase enzyme, the AMP-linked carboxyl group on the amino acid is then transferred to a hydroxyl group on the sugar at the 3′ end of the tRNA molecule. This transfer joins the amino acid by an activated ester linkage to the tRNA and forms the final aminoacyl-tRNA molecule. The synthetase enzyme is not shown in these diagrams.

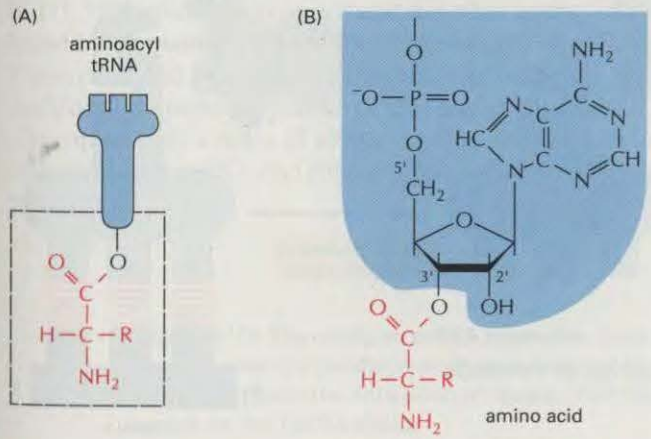


Figure 6–12 The structure of the aminoacyl-tRNA linkage. The carboxyl end of the amino acid forms an ester bond to ribose. Because the hydrolysis of this ester bond is associated with a large favorable change in free energy, an amino acid held in this way is said to be *activated.* (A) Schematic drawing of the structure. (B) Actual structure corresponding to boxed region in (A). As in Figure 6–11, the "R-group" indicates the side chain of the amino acid (see Panel 2–5, pp. 56–57).

Sarepta Exhibit 1040, page 64

enzymes are adaptors of equal importance to the decoding process (Figure 6–13). Thus the genetic code is translated by two sets of adaptors that act sequentially, each matching one molecular surface to another with great specificity; it is their combined action that associates each sequence of three nucleotides in the mRNA molecule—that is, each codon—with its particular amino acid (Figure 6–14).

Amino Acids Are Added to the Carboxyl-Terminal End of a Growing Polypeptide Chain

The fundamental reaction of protein synthesis is the formation of a peptide bond between the carboxyl group at the end of a growing polypeptide chain and a free amino group on an amino acid. Consequently, a protein is synthesized stepwise from its amino-terminal end to its carboxyl-terminal end. Throughout the entire process the growing carboxyl end of the polypeptide chain remains activated by its covalent attachment to a tRNA molecule (a *peptidyl-tRNA* molecule). This high-energy covalent linkage is disrupted in each cycle but is immediately replaced by the identical linkage on the most recently added amino acid (Figure 6–15). In this way each amino acid added carries with it the activation energy for the addition of the *next* amino acid rather than the energy for its own addition—an example of the "head growth" type of polymerization described in Chapter 2 (see Figure 2–36).

The Genetic Code Is Degenerate [5]

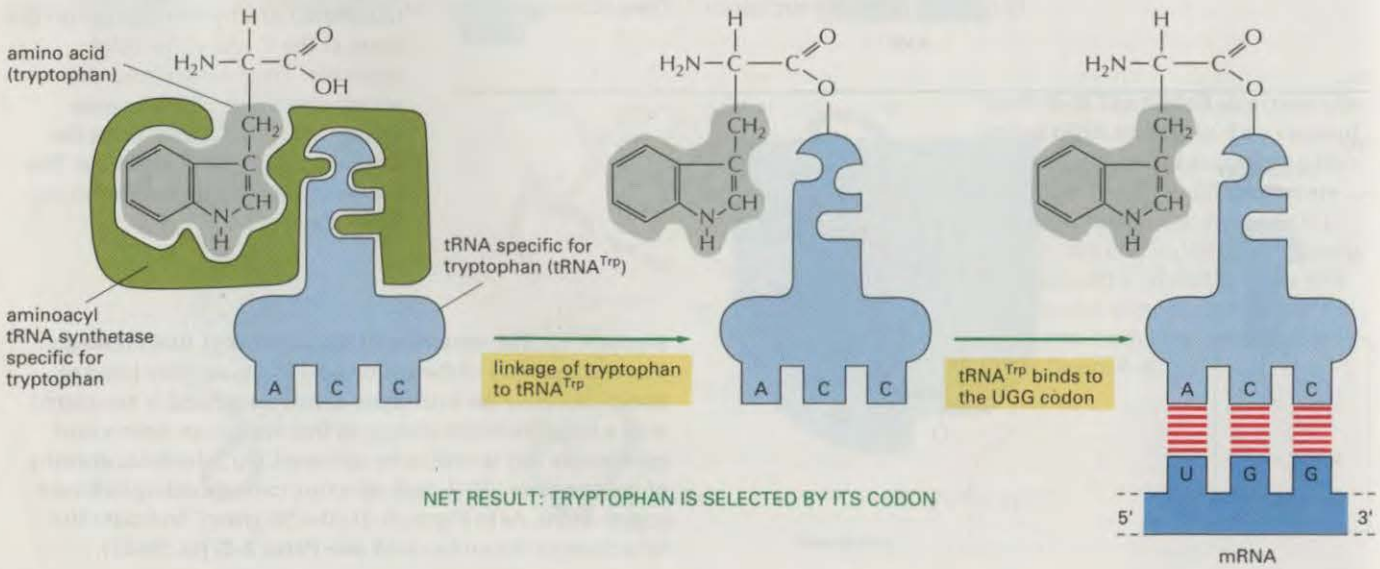
In the course of protein synthesis, the translation machinery moves in the 5′-to-3′ direction along an mRNA molecule and the mRNA sequence is read three nucleotides at a time. As we have seen, each amino acid is specified by the triplet of nucleotides (**codon**) in the mRNA molecule that pairs with a sequence of three complementary nucleotides at the anticodon tip of a particular tRNA. Because only one of the many types of tRNA molecules in a cell can base-pair with each codon, the codon determines the specific amino acid residue to be added to the growing polypeptide chain end (Figure 6–16).

Since RNA is constructed from four types of nucleotides, there are 64 possible sequences composed of three nucleotides ($4 \times 4 \times 4$). Three of these 64 sequences do not code for amino acids but instead specify the termination of a polypeptide chain; they are known as *stop codons*. That leaves 61 codons to specify only 20 different amino acids. For this reason, most of the amino acids are represented by more than one codon (Figure 6–17) and the genetic code is



Figure 6–13 The recognition of a tRNA molecule by its aminoacyl-tRNA synthetase. For this tRNA ($tRNA^{Gln}$), specific nucleotides in both the anticodon (*bottom*) and the amino acid-accepting arm allow the correct tRNA to be recognized by the synthetase enzyme (*blue*). (Courtesy of Tom Steitz.)

Figure 6–14 The genetic code is translated by means of two sequential "adaptors." The first adaptor is the aminoacyl-tRNA synthetase enzyme, which couples a particular amino acid to its corresponding tRNA; the second adaptor is the tRNA molecule, whose *anticodon* forms base pairs with the appropriate nucleotide sequence (*codon*) on the mRNA. An error in either step will cause the wrong amino acid to be incorporated into a protein chain.



Sarepta Exhibit 1040, page 65

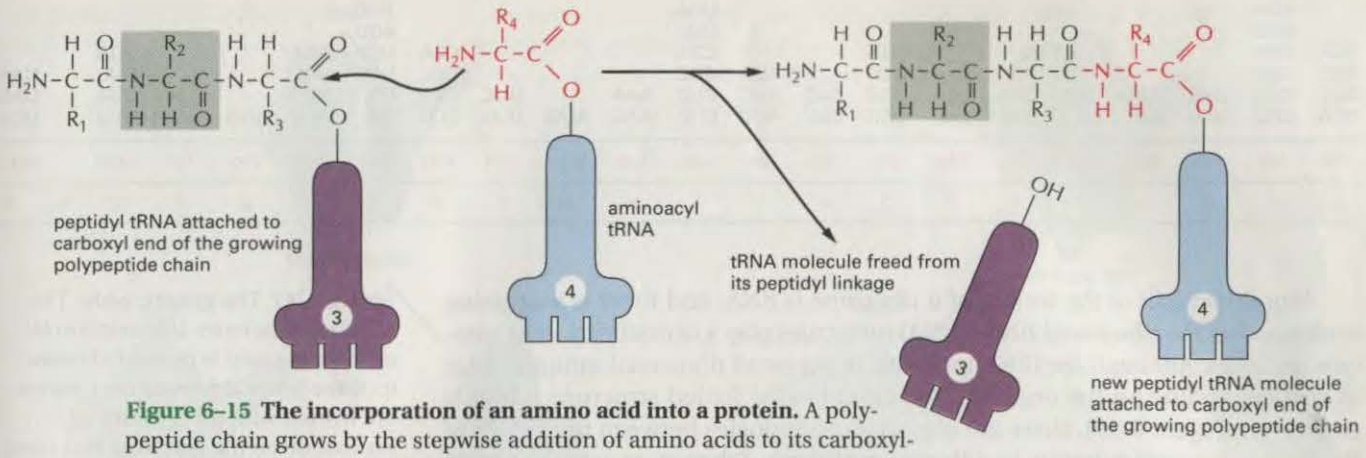


Figure 6–15 The incorporation of an amino acid into a protein. A polypeptide chain grows by the stepwise addition of amino acids to its carboxyl-terminal end. The formation of each peptide bond is energetically favorable because the growing carboxyl terminus has been activated by the covalent attachment of a tRNA molecule. The peptidyl-tRNA linkage that activates the growing end is regenerated in each cycle. The amino acid side chains have been abbreviated as R_1, R_2, R_3, and R_4; as a reference point, all of the atoms in the second amino acid in the polypeptide chain are shaded *gray*.

said to be *degenerate*. Two amino acids, methionine and tryptophan, have only one codon each, and they are the least abundant amino acids in proteins.

The degeneracy of the genetic code implies either that there is more than one tRNA for each amino acid or that a single tRNA molecule can base-pair with more than one codon. In fact, both situations occur. For some amino acids there is more than one tRNA molecule, and some tRNA molecules are constructed so that they require accurate base-pairing only at the first two positions of the codon and can tolerate a mismatch (or wobble) at the third. This *wobble base-pairing* explains why so many of the alternative codons for an amino acid differ only in their third nucleotide (see Figure 6–17). The standard wobble pairings make it possible to fit the 20 amino acids to 61 codons with as few as 31 kinds of tRNA molecules; in animal mitochondria a more extreme wobble allows protein synthesis with only 22 tRNAs (discussed in Chapter 14).

The Events in Protein Synthesis Are Catalyzed on the Ribosome [6]

The protein synthesis reactions just described require a complex catalytic machinery to guide them. The growing end of the polypeptide chain, for example, must be kept in register with the mRNA molecule to ensure that each successive codon in the mRNA engages precisely with the anticodon of a tRNA molecule and does not slip by one nucleotide, thereby changing the reading frame (see Figure 3–17). This precise movement and the other events in protein synthesis are catalyzed by **ribosomes,** which are large complexes of RNA and protein molecules. Eucaryotic and procaryotic ribosomes are very similar in design and function. Both are composed of one large and one small subunit that fit together to form a complex with a mass of several million daltons (Figure 6–18). The small subunit binds the mRNA and tRNAs, while the large subunit catalyzes peptide bond formation.

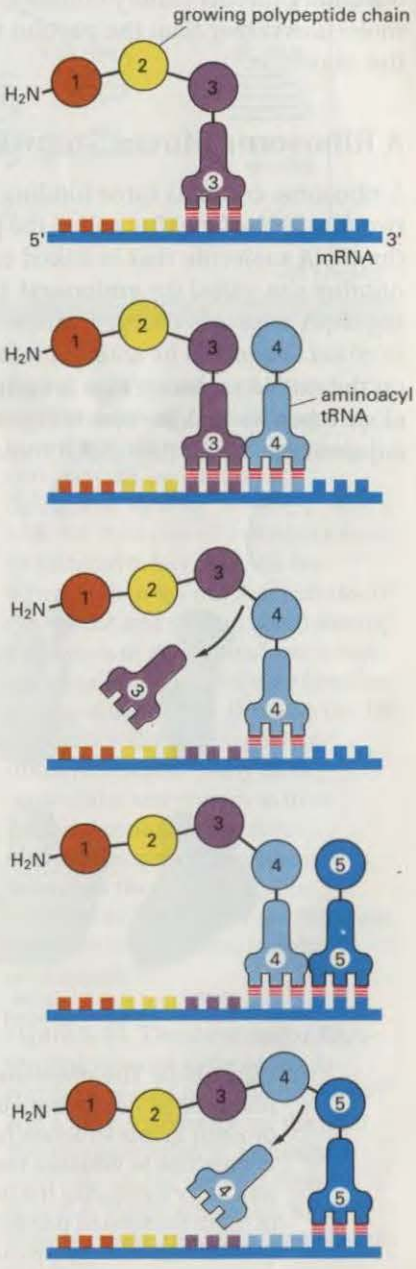


Figure 6–16 Decoding an mRNA molecule. Each amino acid added to the growing end of a polypeptide chain is selected by complementary base-pairing between the anticodon on its attached tRNA molecule and the next codon on the mRNA chain.

Sarepta Exhibit 1040, page 66

GCA GCC GCG GCU	AGA AGG CGA CGC CGG CGU	GAC GAU	AAC AAU	UGC UGU	GAA GAG	CAA CAG	GGA GGC GGG GGU	CAC CAU	AUA AUC AUU	UUA UUG CUA CUC CUG CUU	AAA AAG	AUG	UUC UUU	CCA CCC CCG CCU	AGC AGU UCA UCC UCG UCU	ACA ACC ACG ACU	UGG	UAC UAU	GUA GUC GUG GUU	UAA UAG UGA
Ala	Arg	Asp	Asn	Cys	Glu	Gln	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val	stop
A	R	D	N	C	E	Q	G	H	I	L	K	M	F	P	S	T	W	Y	V	

Figure 6–17 The genetic code. The standard one-letter abbreviation for each amino acid is presented below its three-letter abbreviation. Codons are written with the 5′-terminal nucleotide on the left. Note that most amino acids are represented by more than one codon and that variation is common at the third nucleotide (see also Figure 3–16).

More than half of the weight of a ribosome is RNA, and there is increasing evidence that the **ribosomal RNA (rRNA)** molecules play a central part in its catalytic activities. Although the rRNA molecule in the small ribosomal subunit varies in size depending on the organism, its complicated folded structure is highly conserved (Figure 6–19); there are also close homologies between the rRNAs of the large ribosomal subunits in different organisms. Ribosomes contain a large number of proteins (Figure 6–20), but many of these have been relatively poorly conserved in sequence during evolution, and a surprising number seem not to be essential for ribosome function. Therefore, it has been suggested that the ribosomal proteins mainly enhance the function of the rRNAs and that the RNA molecules rather than the protein molecules catalyze many of the reactions on the ribosome.

A Ribosome Moves Stepwise Along the mRNA Chain [7]

A ribosome contains three binding sites for RNA molecules: one for mRNA and two for tRNAs. One site, called the **peptidyl-tRNA-binding site,** or **P-site,** holds the tRNA molecule that is linked to the growing end of the polypeptide chain. Another site, called the **aminoacyl-tRNA-binding site,** or **A-site,** holds the incoming tRNA molecule charged with an amino acid. A tRNA molecule is held tightly at either site only if its anticodon forms base pairs with a complementary codon on the mRNA molecule that is bound to the ribosome. The A- and P-sites are so close together that the two tRNA molecules are forced to form base pairs with adjacent codons in the mRNA molecule (Figure 6–21).

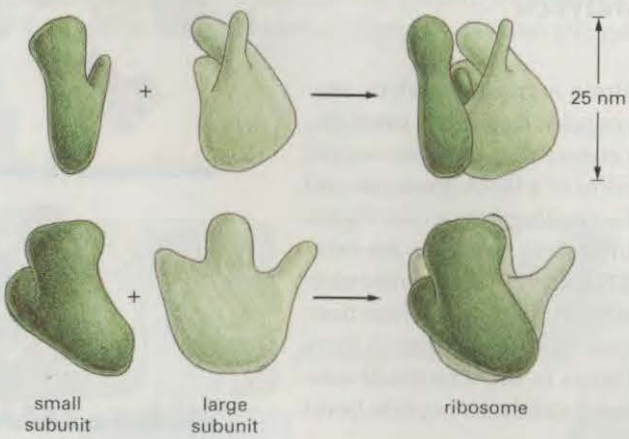


Figure 6–18 The ribosome. A three-dimensional model of the bacterial ribosome as viewed from two angles. The positions of many ribosomal proteins in this structure have been determined by using an electron microscope to visualize the positions where specific antibodies bind, as well as by measuring the neutron scattering from ribosomes containing one or more deuterated proteins. (After J.A. Lake, *Annu. Rev. Biochem.* 54:507–530, 1985. © 1985 by Annual Reviews Inc.)



Figure 6–19 The structure of the rRNA in the small subunit. This model of *E. coli* 16S rRNA is indicative of the complex folding that underlies the catalytic activities of the RNAs in the ribosome. The 16S rRNA molecule contains 1540 nucleotides, and it is folded into three domains: 5′ (*blue*), central (*red*), and 3′ (*green*). (Adapted from S. Stern, B. Weiser, and H.F. Noller, *J. Mol. Biol.* 204:447–481, 1988.)

Sarepta Exhibit 1040, page 67

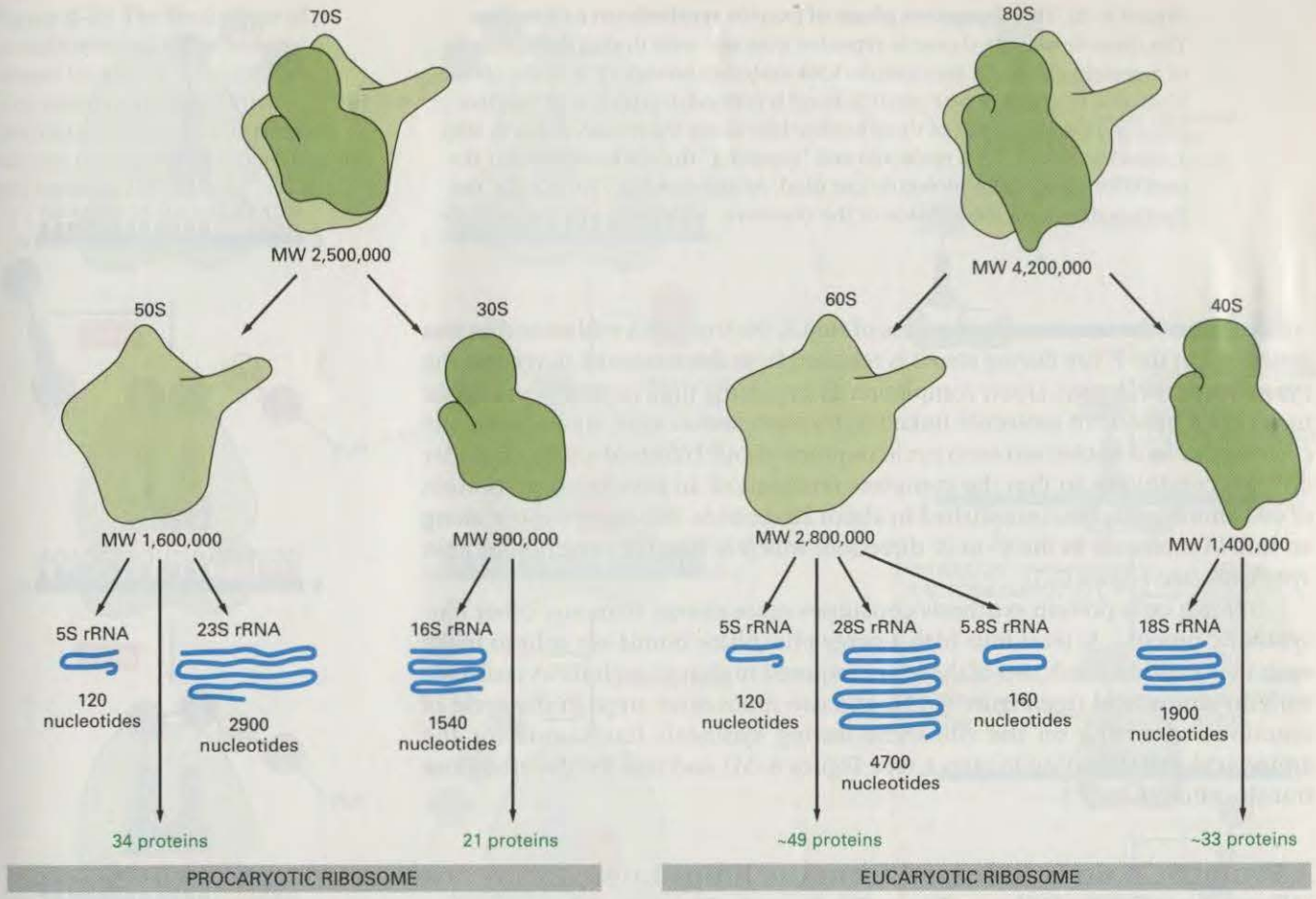


Figure 6–20 A comparison of the structures of procaryotic and eucaryotic ribosomes. Ribosomal components are commonly designated by their "S values," which indicate their rate of sedimentation in an ultracentrifuge. Despite the differences in the number and size of their rRNA and protein components, both types of ribosomes have nearly the same structure and they function in very similar ways. Although the 18S and 28S rRNAs of the eucaryotic ribosome contain many extra nucleotides not present in their bacterial counterparts, these nucleotides are present as multiple insertions that are thought to protrude as loops and leave the basic structure of each rRNA largely unchanged.

The process of polypeptide chain elongation on a ribosome can be thought of as a cycle with three discrete steps (Figure 6–22):

1. In step 1, an aminoacyl-tRNA molecule becomes bound to a vacant ribosomal A-site (adjacent to an occupied P-site) by forming base pairs with the three mRNA nucleotides (*codon*) exposed at the A-site.
2. In step 2, the carboxyl end of the polypeptide chain is uncoupled from the tRNA molecule in the P-site and joined by a peptide bond to the amino acid linked to the tRNA molecule in the A-site. This central reaction of protein synthesis (see Figure 6–15) is catalyzed by a **peptidyl transferase** enzyme. Recent experiments with ribosomes that have been experimentally stripped of proteins show that this catalysis is mediated not by a protein but by a specific region of the major rRNA molecule in the large subunit (see Figure 3–23).
3. In step 3, the new peptidyl-tRNA in the A-site is translocated to the P-site as the ribosome moves exactly three nucleotides along the mRNA molecule. This step requires energy and is driven by a series of conformational changes induced in one of the ribosomal components by the hydrolysis of a GTP molecule.

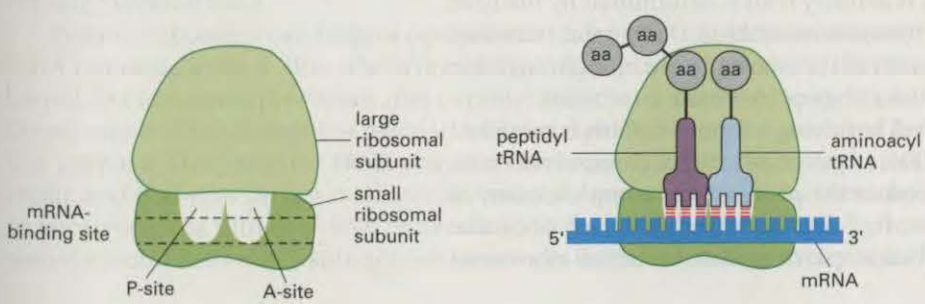


Figure 6–21 The three major RNA-binding sites on a ribosome. An empty ribosome is shown on the left and a loaded ribosome on the right. The representation of a ribosome used here and in the next three figures is highly schematic; for a more accurate view, see Figures 6–18 and 6–25.

Sarepta Exhibit 1040, page 68

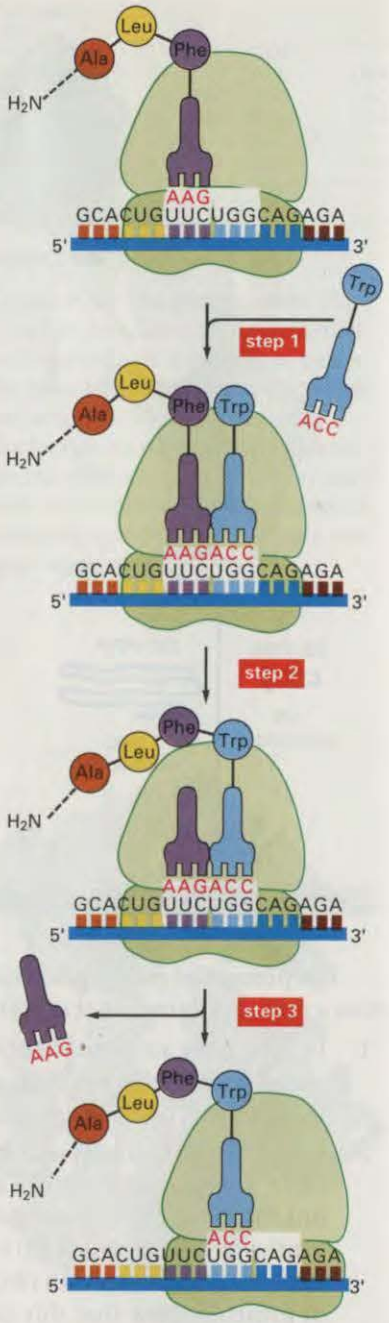


Figure 6–22 The elongation phase of protein synthesis on a ribosome. The three-step cycle shown is repeated over and over during the synthesis of a protein chain. An aminoacyl-tRNA molecule binds to the A-site on the ribosome in step 1, a new peptide bond is formed in step 2, and the ribosome moves a distance of three nucleotides along the mRNA chain in step 3, ejecting an old tRNA molecule and "resetting" the ribosome so that the next aminoacyl-tRNA molecule can bind. As indicated in Figure 6–21, the P-site is drawn on the left side of the ribosome, with the A-site on the right.

As part of the translocation process of step 3, the free tRNA molecule that was generated in the P-site during step 2 is released from the ribosome to reenter the cytoplasmic tRNA pool. Upon completion of step 3, the unoccupied A-site is free to accept a new tRNA molecule linked to the next amino acid, which starts the cycle again. In a bacterium each cycle requires about 1/20th of a second under optimal conditions, so that the complete synthesis of an average-sized protein of 400 amino acids is accomplished in about 20 seconds. Ribosomes move along an mRNA molecule in the 5′-to-3′ direction, which is also the direction of RNA synthesis (see Figure 6–3).

In most cells protein synthesis consumes more energy than any other biosynthetic process. At least four high-energy phosphate bonds are split to make each new peptide bond: two of these are required to charge each tRNA molecule with an amino acid (see Figure 6–11), and two more drive steps in the cycle of reactions occurring on the ribosome during synthesis itself—one for the aminoacyl-tRNA binding in step 1 (see Figure 6–31) and one for the ribosome translocation in step 3.

A Protein Chain Is Released from the Ribosome When Any One of Three Stop Codons Is Reached [8]

Of the 64 possible codons in an mRNA molecule, 3 (UAA, UAG, and UGA) are **stop codons,** which terminate the translation process. Cytoplasmic proteins called *release factors* bind directly to any stop codon that reaches the A-site on the ribosome. This binding alters the activity of the peptidyl transferase, causing it to catalyze the addition of a water molecule instead of an amino acid to the peptidyl-tRNA. This reaction frees the carboxyl end of the growing polypeptide chain from its attachment to a tRNA molecule, and since only this attachment normally holds the growing polypeptide to the ribosome, the completed protein chain is immediately released into the cytoplasm. The ribosome releases the mRNA and dissociates into its two separate subunits (Figure 6–23), which can assemble on another mRNA molecule to begin a new round of protein synthesis by the process to be described next.

The Initiation Process Sets the Reading Frame for Protein Synthesis [9]

In principle, an RNA sequence can be translated in any one of three **reading frames,** each of which will specify a completely different polypeptide chain (see Figure 3–17). Which of the three frames is actually read is determined by the RNA sequence, which determines how the ribosome assembles. During the *initiation* phase of protein synthesis, the two subunits of the ribosome are brought together at the exact spot on the mRNA where the polypeptide chain is to begin.

The initiation process is complicated, involving a number of steps catalyzed by proteins called **initiation factors (IFs),** many of which are themselves composed of several polypeptide chains. Because the process is so complex, many of the details of initiation are still uncertain. It is clear, however, that each ribosome is assembled onto an mRNA chain in two steps: only after the small ribosomal

Sarepta Exhibit 1040, page 69

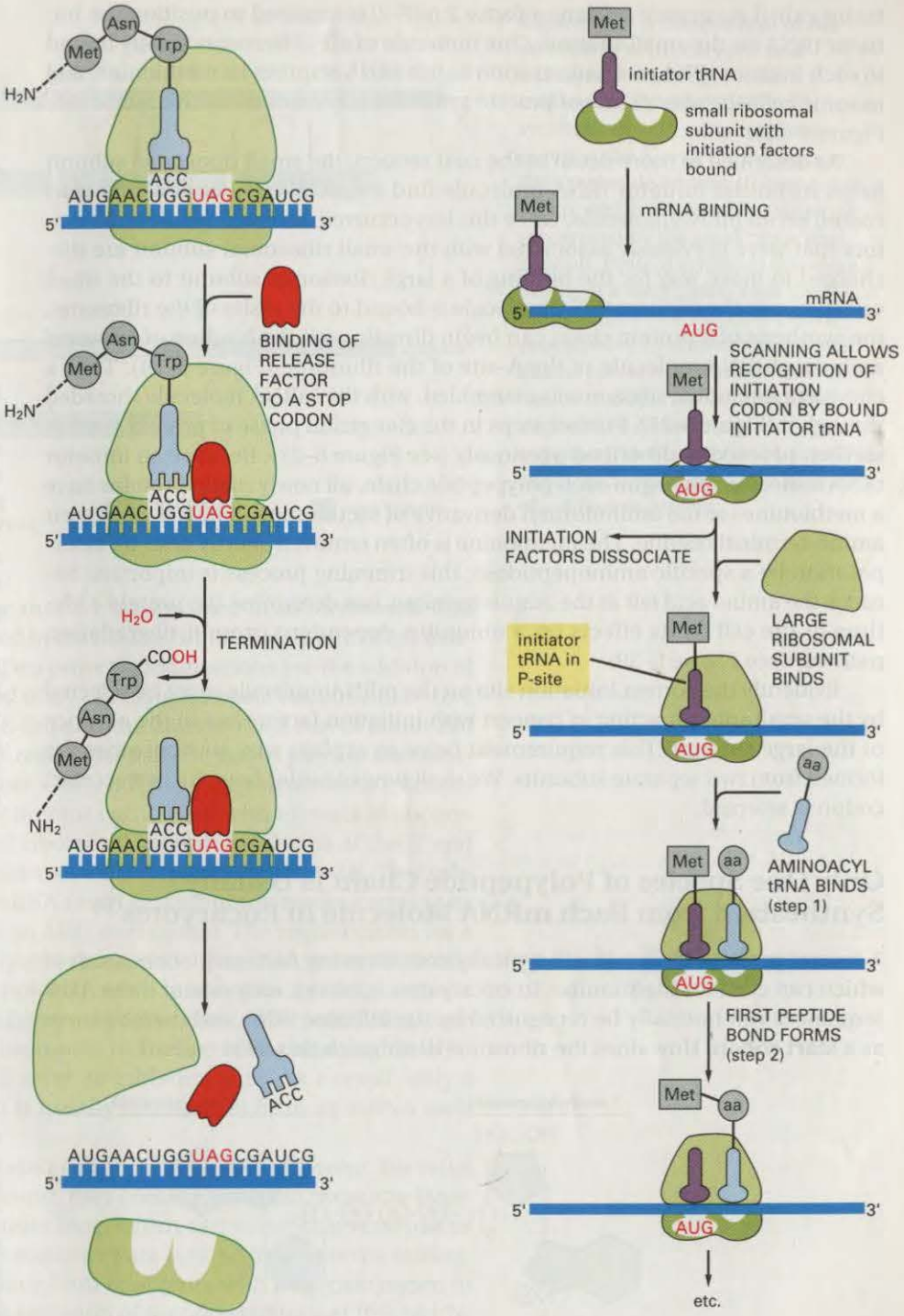


Figure 6–23 The final phase of protein synthesis. The binding of release factor to a stop codon terminates translation. The completed polypeptide is released, and the ribosome dissociates into its two separate subunits.

subunit loaded with initiation factors finds the start codon (AUG, see below) does the large subunit bind.

Before a ribosome can begin a new protein chain, it must bind an aminoacyl tRNA molecule in its P-site, where normally only peptidyl tRNA molecules are bound. (As explained previously, the peptidyl tRNA is translocated to the P-site during step 3 of the elongation reaction.) A special tRNA molecule is required for this purpose. This **initiator tRNA** provides the amino acid that starts a protein chain, and it always carries methionine (aminoformyl methionine in bacteria). In eucaryotes the initiator tRNA molecule must be loaded onto the small ribosomal subunit before this subunit can bind to an mRNA molecule. An initiation

Figure 6–24 The initiation phase of protein synthesis in eucaryotes. Step 1 and step 2 refer to steps in the elongation reaction shown in Figure 6–22.

Sarepta Exhibit 1040, page 70

factor called *eucaryotic initiation factor 2 (eIF-2)* is required to position the initiator tRNA on the small subunit. One molecule of eIF-2 becomes tightly bound to each initiator tRNA molecule as soon as this tRNA acquires its methionine, and in some cells the overall rate of protein synthesis is controlled by this factor (see Figure 9–82).

As described in more detail in the next section, the small ribosomal subunit helps its bound initiator tRNA molecule find a special AUG codon (the *start codon*) on an mRNA molecule. Once this has occurred, the several initiation factors that were previously associated with the small ribosomal subunit are discharged to make way for the binding of a large ribosomal subunit to the small one. Because the initiator tRNA molecule is bound to the P-site of the ribosome, the synthesis of a protein chain can begin directly with the binding of a second aminoacyl-tRNA molecule to the A-site of the ribosome (Figure 6–24). Thus a complete functional ribosome is assembled, with the mRNA molecule threaded through it (Figure 6–25). Further steps in the elongation phase of protein synthesis then proceed as described previously (see Figure 6–22). Because an initiator tRNA molecule has begun each polypeptide chain, all newly made proteins have a methionine (or the aminoformyl derivative of methionine in bacteria) as their amino-terminal residue. The methionine is often removed shortly after its incorporation by a specific aminopeptidase; this trimming process is important because the amino acid left at the amino terminus can determine the protein's lifetime in the cell by its effects on a ubiquitin-dependent protein-degradation pathway (see Figure 5–39).

Evidently the correct initiation site on the mRNA molecule must be selected by the small subunit acting in concert with initiation factors but in the absence of the large subunit. This requirement helps to explain why all ribosomes are formed from two separate subunits. We shall now consider how the correct start codon is selected.

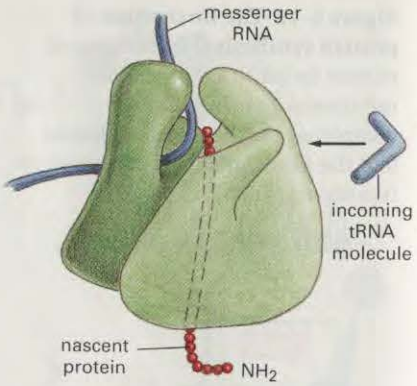


Figure 6–25 A three-dimensional model of a functioning bacterial ribosome. The small (*dark green*) subunit and the large (*light green*) subunit form a complex through which the messenger RNA is threaded. Although the exact paths of the mRNA and the nascent polypeptide chain are unknown, the addition of amino acids occurs in the general region shown, with the tRNAs held in the pocket formed between the large and small subunit. (Modified from J.A. Lake, *Annu. Rev. Biochem.* 54:507–530, 1985. © 1985 by Annual Reviews Inc.)

Only One Species of Polypeptide Chain Is Usually Synthesized from Each mRNA Molecule in Eucaryotes [10]

A messenger RNA molecule will typically contain many AUG sequences, each of which can code for methionine. In eucaryotes, however, only one of these AUG sequences will normally be recognized by the initiator tRNA and thereby serve as a **start codon.** How does the ribosome distinguish this start codon?

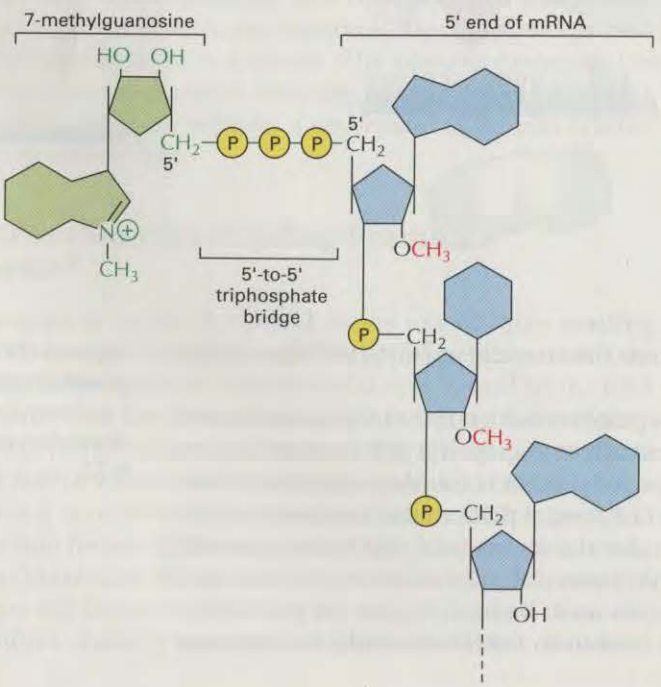


Figure 6–26 The structure of the cap at the 5′ end of eucaryotic mRNA molecules. Note the unusual 5′-to-5′ linkage to the positively charged 7-methylguanosine and the methylation of the 2′ hydroxyl group on the first ribose sugar in the RNA. (The second sugar may or may not be methylated.)

Sarepta Exhibit 1040, page 71

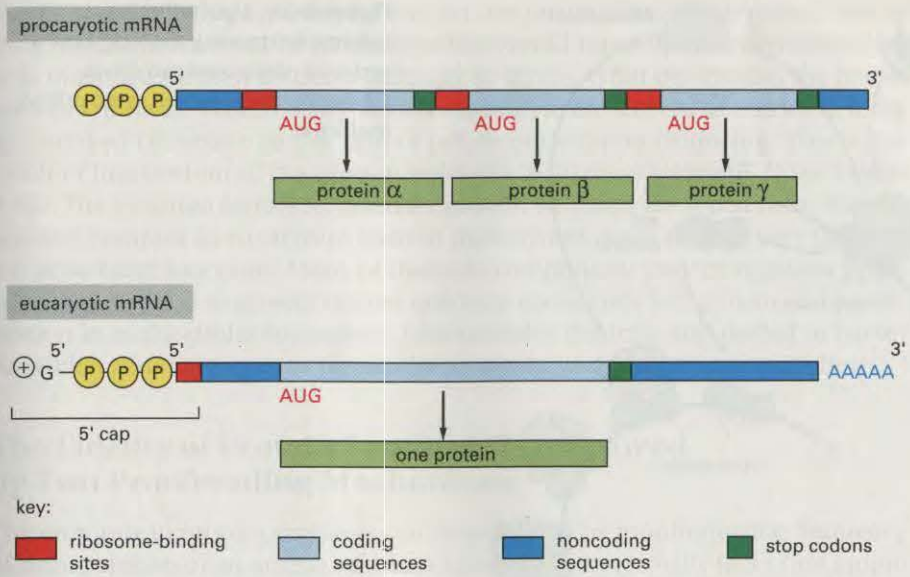


Figure 6–27 A comparison of the structures of procaryotic and eucaryotic messenger RNA molecules. Although both mRNAs are synthesized with a triphosphate group at the 5′ end, the eucaryotic RNA molecule immediately acquires a 5′ cap, which is part of the structure recognized by the small ribosomal subunit. Protein synthesis therefore begins at a start codon near the 5′ end of the mRNA (see Figure 6–24). In procaryotes, by contrast, the 5′ end has no special significance, and there can be multiple ribosome-binding sites (called *Shine-Dalgarno sequences*) in the interior of an mRNA chain, each resulting in the synthesis of a different protein.

Eucaryotic RNAs (except those that are synthesized in mitochondria and chloroplasts) are extensively modified in the nucleus immediately after their transcription (discussed in Chapter 8). Two general modifications are the addition of a unique "cap" structure, composed of a 7-methylguanosine residue linked to a triphosphate at the 5′ end (Figure 6–26) and the addition of a run of about 200 adenylic residues ("poly A") at the 3′ end. What part the poly A plays in the translation process is uncertain (see Figure 9–87), but the 5′ cap structure is essential for efficient protein synthesis. Experiments carried out with extracts of eucaryotic cells have shown that the small ribosomal subunit first binds at the 5′ end of an mRNA chain, aided by recognition of the 5′ cap (see Figure 6–24). This subunit then propels itself along the mRNA chain in a scanning mode, carrying its bound initiator tRNA in a search of an AUG start codon. The requirements for a start codon apparently are not very stringent, since the small subunit usually selects the first AUG it encounters; however, a few nucleotides in addition to the AUG are also important for the selection process. For most eucaryotic RNAs, once a start codon near the 5′ end has been selected, none of the many other AUG codons farther down the chain will serve as initiation sites. As a result, only a single species of polypeptide chain is usually synthesized from an mRNA molecule (for exceptions see p. 467).

The mechanism for selecting a start codon in bacteria is different. Bacterial mRNAs have no 5′ cap structure. Instead, they contain a specific ribosome-binding site sequence, up to six nucleotides long, which can occur at several places in the same mRNA molecule. These sequences are located four to seven nucleotides upstream from an AUG, and they form base pairs with a specific region of the rRNA in a ribosome to signal the initiation of protein synthesis at this nearby start codon. Bacterial ribosomes, unlike eucaryotic ribosomes, bind directly to start codons in the interior of an mRNA molecule to initiate protein synthesis. As a result, bacterial messenger RNAs are commonly *polycistronic*—that is, they encode multiple proteins that are separately translated from the same mRNA molecule. Eucaryotic mRNAs, in contrast, are typically *monocistronic*, with only one species of polypeptide chain being translated per messenger molecule (Figure 6–27).

The Binding of Many Ribosomes to an Individual mRNA Molecule Generates Polyribosomes [11]

The complete synthesis of a protein takes 20 to 60 seconds on average. But even during this very short period, multiple initiations usually take place on each

Sarepta Exhibit 1040, page 72

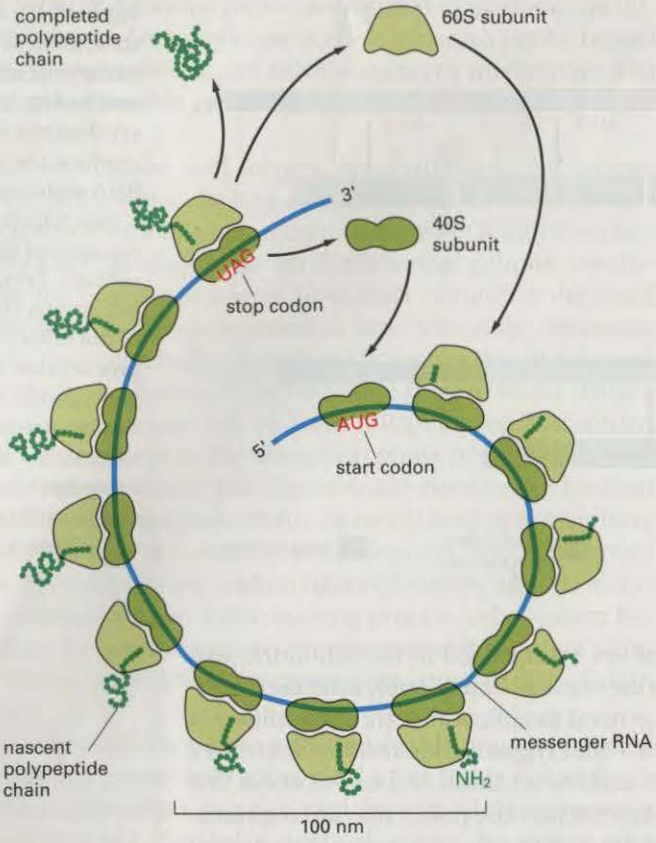


Figure 6–28 A polyribosome. Schematic drawing showing how a series of ribosomes can simultaneously translate the same mRNA molecule.

mRNA molecule being translated. A new ribosome hops onto the 5′ end of the mRNA molecule almost as soon as the preceding ribosome has translated enough of the amino acid sequence to be out of the way. Such mRNA molecules are thus found in the cell as **polyribosomes,** or **polysomes,** formed by several ribosomes spaced as close as 80 nucleotides apart along a single messenger molecule (Figures 6–28 and 6–29). Polyribosomes are a common feature of cells. They can be isolated and separated from single ribosomes in the cytosol by ultracentrifugation after cell lysis (Figure 6–30). The mRNA purified from these polyribosomes can be used to determine if the protein encoded by a particular DNA sequence is being actively synthesized in the cells used to prepare the polyribosomes. These mRNA molecules can also serve as the starting material for the preparation of specialized cDNA libraries (discussed in Chapter 7).

In eucaryotes the nuclear envelope keeps transcription and protein synthesis separate. But in procaryotes, RNA is accessible to ribosomes as soon as it is made. Thus, ribosomes will begin synthesizing a polypeptide chain at the 5′ end of a nascent mRNA molecule and then follow behind the RNA polymerase as it completes an mRNA chain.

The Overall Rate of Protein Synthesis in Eucaryotes Is Controlled by Initiation Factors [12]

As we discuss in Chapter 17, the cells in a multicellular organism proliferate only when they are stimulated to do so by specific growth factors. Although the

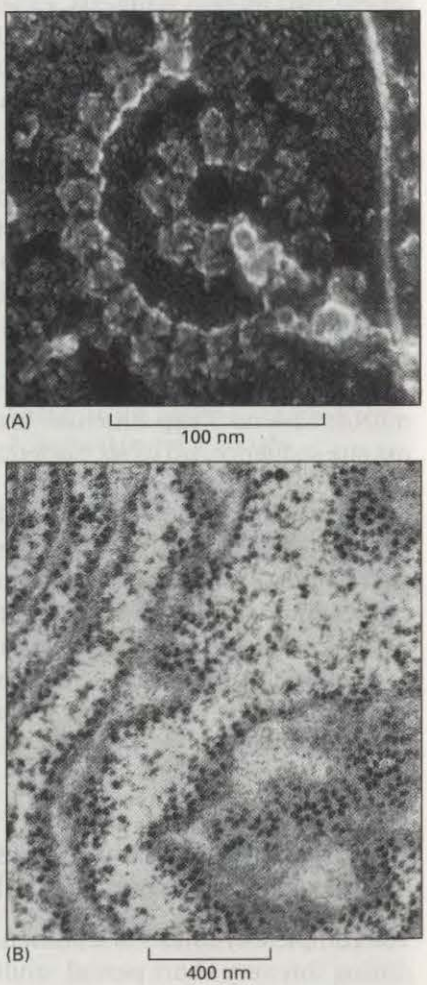


Figure 6–29 Freeze-etch (A) and transmission (B) electron micrographs of typical polyribosomes in a eucaryotic cell. The cell cytoplasm is generally crowded with such polyribosomes, some free in the cytosol and some membrane-bound. (A, courtesy of John Heuser; B, courtesy of George Palade.)

Sarepta Exhibit 1040, page 73

mechanisms by which growth factors act are incompletely understood, one of their major effects must be to increase the overall rate of protein synthesis, for cells must double their contents before they divide. What determines the rate of protein synthesis? When eucaryotic cells in culture are starved of nutrients, there is a marked reduction in the rate of polypeptide chain *initiation*. This is the result of inactivation of the protein synthesis initiation factor eIF-2 (see Figure 9–82). The initiation factors required for protein synthesis are much more numerous and complex in eucaryotes than in procaryotes, even though they perform the same basic functions. Many of the extra components may be regulatory proteins that respond to growth factors and help coordinate cell growth and proliferation in multicellular organisms. Less complex controls are needed in bacteria, which generally grow as fast as the nutrients in their environment allow.

The Fidelity of Protein Synthesis Is Improved by Two Proofreading Mechanisms [13]

The error rate in protein synthesis can be estimated by monitoring the frequency of incorporation of an amino acid into a protein that normally lacks that amino acid. Error rates of about 1 amino acid misincorporated for every 10^4 amino acids polymerized are observed, which means that only about 1 in every 25 protein molecules of average size (400 amino acids) should contain an error. The fidelity of protein synthesis depends on the accuracy of the two adaptor mechanisms previously discussed: the linking of each amino acid to its corresponding tRNA molecule and the base-pairing of the codons in mRNA to the anticodons in tRNA (see Figure 6–14). Not surprisingly, cells have evolved "proofreading" mechanisms to reduce the number of errors in both these crucial steps of protein synthesis.

Two fundamentally different proofreading mechanisms are used, each representative of strategies used in other processes in the cell. Both involve expenditure of free energy, since, as discussed in Chapter 2, a price must be paid for any increase in order in the cell. A relatively simple mechanism is used to improve the accuracy of amino acid attachment to tRNA. Many aminoacyl tRNA synthetases have two active sites, one that carries out the loading reaction shown earlier (Figure 6–11) and one that recognizes an incorrect amino acid attached to its tRNA molecule and removes it by hydrolysis. The correction process is energetically costly because to be effective it must remove an appreciable fraction of correctly attached amino acids as well. The same type of costly two-step proofreading process is used in DNA replication (see Figure 6–42).

A more subtle "kinetic proofreading" mechanism is used to improve the fidelity of codon-anticodon pairing. Thus far we have given a simplified account of this pairing. In fact, once tRNA molecules have acquired an amino acid, they form a complex with an abundant protein called an *elongation factor (EF)*, which binds tightly to both the amino acid end of a tRNA and to a molecule of GTP. It is this complex, and not free tRNA, that pairs with the appropriate codon in an mRNA molecule. The bound elongation factor allows correct codon-anticodon pairing to occur but prevents the amino acid from being incorporated into the growing polypeptide chain. The initial codon recognition, however, triggers the elongation factor to hydrolyze its bound GTP (to GDP and inorganic phosphate), whereupon the factor dissociates from the ribosome without its tRNA, allowing protein synthesis to proceed. The elongation factor thereby introduces a short delay between codon-anticodon base-pairing and polypeptide chain elongation, which provides an opportunity for the bound tRNA molecule to exit from the ribosome. An incorrect tRNA molecule forms a smaller number of codon-anticodon hydrogen bonds than a correct one; it therefore binds more weakly to the ribosome and is more likely to dissociate during this period. Because the delay introduced by the elongation factor causes most incorrectly bound tRNA molecules to leave the ribosome without being used for protein synthesis, this factor increases the ratio of correct to incorrect amino acids incorporated into protein (Figure 6–31).

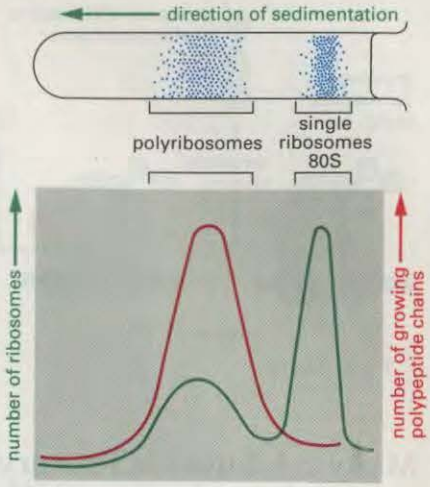


Figure 6–30 The isolation of polyribosomes. Polyribosomes are separated from single ribosomes (and their subunits) by sedimentation in a centrifuge. This method is based on the fact that large molecular aggregates move faster than small ones in a strong gravitational field. Generally, the sedimentation is done through a gradient of sucrose to stabilize the solution against convective mixing. Note that most of the growing polypeptide chains (*red line*) are associated with the polyribosomes.

Sarepta Exhibit 1040, page 74

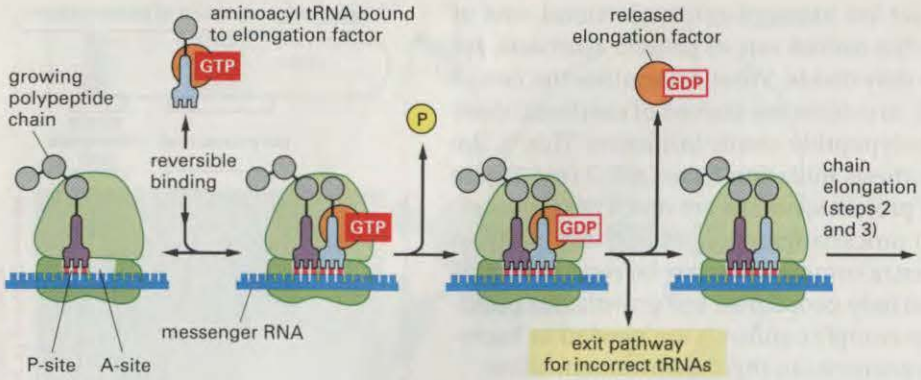


Figure 6–31 Kinetic proofreading selects for the correct tRNA molecule on the ribosome. This more detailed view of step 1 of the elongation phase of protein synthesis shows how, in the initial binding event, an aminoacyl-tRNA molecule that is tightly bound to an elongation factor pairs transiently with the codon at the A-site. This pairing triggers GTP hydrolysis by the elongation factor, enabling the factor to dissociate from the aminoacyl-tRNA molecule, which can now participate in chain elongation (see Figure 6–22). A delay between aminoacyl tRNA binding and its availability for protein synthesis is thereby inserted into the protein synthesis mechanism. As a result, only those tRNAs with the correct anticodon are likely to remain paired to the mRNA long enough to be added to the growing polypeptide chain.

The elongation factor, which is an abundant protein, is called EF-Tu in procaryotes and EF-1 in eucaryotes. The dramatic change in the three-dimensional structure of EF-Tu that is caused by GTP hydrolysis is illustrated in Figure 5–20.

Many Inhibitors of Procaryotic Protein Synthesis Are Useful as Antibiotics [14]

Many of the most effective antibiotics used in modern medicine are compounds made by fungi that act by inhibiting bacterial protein synthesis. A number of these drugs exploit the structural and functional differences between procaryotic and eucaryotic ribosomes so as to interfere with the function of procaryotic ribosomes preferentially. Thus some of these compounds can be taken in high doses without undue toxicity to humans. Because different antibiotics bind to different regions of bacterial ribosomes, they often inhibit different steps in the synthetic process. Some of the more common antibiotics of this kind are listed in Table 6–1 along with several other commonly used inhibitors of protein synthesis, some of which act on eucaryotic cells and therefore cannot be used as antibiotics.

Table 6–1 Inhibitors of Protein or RNA Synthesis

Inhibitor	Specific Effect
*Acting Only on Procaryotes**	
Tetracycline	blocks binding of aminoacyl-tRNA to A-site of ribosome
Streptomycin	prevents the transition from initiation complex to chain-elongating ribosome and also causes miscoding
Chloramphenicol	blocks the peptidyl transferase reaction on ribosomes (step 2 in Figure 6–22)
Erythromycin	blocks the translocation reaction on ribosomes (step 3 in Figure 6–22)
Rifamycin	blocks initiation of RNA chains by binding to RNA polymerase (prevents RNA synthesis)
Acting on Procaryotes and Eucaryotes	
Puromycin	causes the premature release of nascent polypeptide chains by its addition to growing chain end
Actinomycin D	binds to DNA and blocks the movement of RNA polymerase (prevents RNA synthesis)
Acting Only on Eucaryotes	
Cycloheximide	blocks the translocation reaction on ribosomes (step 3 in Figure 6–22)
Anisomycin	blocks the peptidyl transferase reaction on ribosomes (step 2 in Figure 6–22)
α-Amanitin	blocks mRNA synthesis by binding preferentially to RNA polymerase II

*The ribosomes of eucaryotic mitochondria (and chloroplasts) often resemble those of procaryotes in their sensitivity to inhibitors. Therefore, some of these antibiotics can have a deleterious effect on human mitochondria.

Sarepta Exhibit 1040, page 75

Because they block specific steps in the processes that lead from DNA to protein, many of the compounds listed in Table 6–1 are useful for cell biological studies. Among the most commonly used drugs in such experimental studies are *chloramphenicol, cycloheximide,* and *puromycin,* all of which specifically inhibit protein synthesis. In a eucaryotic cell, for example, chloramphenicol inhibits protein synthesis on ribosomes only in mitochondria (and in chloroplasts in plants), presumably reflecting the procaryotic origins of these organelles (discussed in Chapter 14). Cycloheximide, on the other hand, affects only ribosomes in the cytosol. The difference in the sensitivity of protein synthesis to these two drugs provides a powerful way to determine in which cell compartment a particular protein is translated. Puromycin is especially interesting because it is a structural analogue of a tRNA molecule linked to an amino acid; the ribosome mistakes it for an authentic amino acid and covalently incorporates it at the carboxyl terminus of the growing polypeptide chain, thereby causing the premature termination and release of the polypeptide (see Figure 3–23). As might be expected, puromycin inhibits protein synthesis in both procaryotes and eucaryotes.

How Did Protein Synthesis Evolve? [15]

The molecular processes underlying protein synthesis seem inexplicably complex. Although we can describe many of them, they do not make conceptual sense in the way that DNA transcription, DNA repair, and DNA replication do. As we have seen, protein synthesis in present-day organisms centers on the ribosome, which consists of proteins arranged around a core of rRNA molecules. Why should rRNA molecules exist at all, and how did they come to play such a dominant part in the structure and function of the ribosome?

Before the discovery of mRNA in the early 1960s, it was suspected that the large amount of RNA in ribosomes served a "messenger" function, carrying genetic information from DNA to proteins. Now we know, however, that all of the ribosomes in a cell contain an identical set of rRNA molecules that have no such informational role. In bacterial ribosomes, rRNA molecules have been shown to have catalytic functions in protein synthesis. As mentioned earlier, the major rRNA of the large ribosomal subunit appears to be the peptidyl transferase; in addition, the rRNA of the small ribosomal subunit forms a short base-paired helix with the initiation site sequence on bacterial mRNA molecules, positioning the neighboring AUG start codon at the P-site. A variety of specific base-pair interactions likewise form between tRNA molecules and bacterial rRNAs, although these interactions involve individual bases on the rRNA that are far apart in the nucleotide sequence, suggesting complex sets of interactions that depend on the tertiary structure of the rRNA.

Protein synthesis also relies heavily on a large number of proteins that are bound to the rRNAs in a ribosome (see Figure 6–20). The complexity of a process with so many interacting components has made many biologists despair of ever understanding the pathway by which protein synthesis evolved. The discovery that RNA molecules can act as enzymes, however, has provided a new way of viewing the pathway. As discussed in Chapter 1, early biological reactions probably used RNA molecules rather than protein molecules as catalysts. In the earliest cells tRNA molecules on their own may have formed catalytic surfaces that allowed them to bind and activate specific amino acids without requiring aminoacyl-tRNA synthetase enzymes. Likewise, rRNA molecules may have served by themselves as the entire "ribosome," folding up in complex ways to generate an intricate set of surfaces that both guided tRNA pairings with mRNA codons and catalyzed the polymerization of the tRNA-linked amino acids (see Figure 1–7). Over the course of evolution individual proteins have been added to this machinery, each one making the process a little more accurate and efficient, or adding regulatory controls. In this view the large amount of RNA in present-day ribosomes is a remnant of a very early stage in evolution, before proteins dominated biological catalysis.

Sarepta Exhibit 1040, page 76

Summary

Before the synthesis of a particular protein can begin, the corresponding mRNA molecule must be produced by DNA transcription. Then a small ribosomal subunit binds to the mRNA molecule at a start codon (AUG) that is recognized by a unique initiator tRNA molecule. A large ribosomal subunit binds to complete the ribosome and initiate the elongation phase of protein synthesis. During this phase aminoacyl tRNAs, each bearing a specific amino acid, sequentially bind to the appropriate codon in mRNA by forming complementary base pairs with the tRNA anticodon. Each amino acid is added to the carboxyl-terminal end of the growing polypeptide by means of a cycle of three sequential steps: aminoacyl-tRNA binding, followed by peptide bond formation, followed by ribosome translocation. The ribosome progresses from codon to codon in the 5′-to-3′ direction along the mRNA molecule until one of three stop codons is reached. A release factor then binds to the stop codon, terminating translation and releasing the completed polypeptide from the ribosome.

Eucaryotic and procaryotic ribosomes are highly homologous, despite substantial differences in the number and size of their rRNA and protein components. The predominant role of rRNA in ribosome structure and function is likely to reflect the ancient origin of protein synthesis, which is thought to have evolved in an environment dominated by RNA-mediated catalysis.

DNA Repair [16]

The long-term survival of a species may be enhanced by genetic changes, but the survival of the individual demands genetic stability. Maintaining genetic stability requires not only an extremely accurate mechanism for replicating the DNA before a cell divides, but also mechanisms for repairing the many accidental lesions that occur continually in DNA. Most such spontaneous changes in DNA are temporary because they are immediately corrected by processes collectively called *DNA repair*. Only rarely do the cell's DNA maintenance processes fail and allow a permanent change in the DNA. Such a change is called a **mutation,** and it can destroy an organism if the change occurs in a vital position in the DNA sequence.

Before examining the mechanisms of DNA repair, we briefly discuss the maintenance of DNA sequences from one generation to the next.

DNA Sequences Are Maintained with Very High Fidelity [17]

The rate at which stable changes occur in DNA sequences (the *mutation rate*) can be estimated only indirectly. One way is to compare the amino acid sequence of the same protein in several species. The fraction of the amino acids that are different can then be compared with the estimated number of years since each pair of species diverged from a common ancestor, as determined from the fossil record. In this way one can calculate the number of years that elapse, on average, before an inherited change in the amino acid sequence of a protein becomes fixed in the species. Because each such change will commonly reflect the alteration of a single nucleotide in the DNA sequence of the gene encoding that protein, this value can be used to estimate the average number of years required to produce a single, stable mutation in the gene.

Such calculations always will substantially underestimate the actual mutation rate because most mutations will spoil the function of the protein and vanish from the population through natural selection. But there is one family of proteins whose sequence does not seem to matter, and so the genes that encode them can accumulate mutations without being selected against. These proteins are the **fibrinopeptides**—20-residue-long fragments that are discarded from the protein *fibrinogen* when it is activated to form *fibrin* during blood clotting. Since the function of fibrinopeptides apparently does not depend on their amino acid se-

Sarepta Exhibit 1040, page 77

quence, they can tolerate almost any amino acid change. Sequence analysis of the fibrinopeptides indicates that an average-sized protein 400 amino acids long would be randomly altered by an amino acid change roughly once every 200,000 years. More recently, DNA sequencing technology has made it possible to compare corresponding nucleotide sequences in regions of the genome that do not code for protein. Comparisons of such sequences in several mammalian species produce estimates of the mutation rate during evolution that are in excellent agreement with those obtained from the fibrinopeptide studies.

The Observed Mutation Rates in Proliferating Cells Are Consistent with Evolutionary Estimates [18]

The mutation rate can be estimated more directly by observing the rate at which spontaneous genetic changes arise in a large population of cells followed over a relatively short period of time. This can be done either by estimating the frequency with which new mutants arise in very large animal populations (in a colony of fruit flies or mice, for example) or by screening for changes in specific proteins in cells growing in culture. Although they are only approximate, the numbers obtained in both cases are consistent with an error frequency of 1 base-pair change in roughly 10^9 base pairs for each cell generation. Consequently, a single gene that encodes an average-sized protein (containing about 10^3 coding base pairs) would suffer a mutation once in about 10^6 cell generations. This number is at least roughly consistent with the evolutionary estimate described above, in which one mutation appears in an average gene in the germ line every 200,000 years.

Most Mutations in Proteins Are Deleterious and Are Eliminated by Natural Selection [19]

When the number of amino acid differences in a particular protein is plotted for several pairs of species against the time since the species diverged, the result is a reasonably straight line. That is, the longer the period since divergence, the larger the number of differences. For convenience, the slope of this line can be expressed in terms of the "unit evolutionary time" for that protein, which is the average time required for 1 amino acid change to appear in a sequence of 100 amino acid residues. When various proteins are compared, each shows a different but characteristic rate of evolution (Figure 6–32). Since all DNA base pairs are thought to be subject to roughly the same rate of random mutation, these different rates must reflect differences in the probability that an organism with a random mutation over the given protein will survive and propagate. Changes in amino acid sequence are evidently much more harmful for some proteins than for others. From Table 6–2 we can estimate that about 6 of every 7 random amino acid changes are harmful over the long term in hemoglobin, about 29 of every 30 amino acid changes are harmful in cytochrome c, and virtually all amino acid changes are harmful in histone H4. We assume that individuals who carried such harmful mutations have been eliminated from the population by natural selection.

Low Mutation Rates Are Necessary for Life as We Know It [19]

Since most mutations are deleterious, no species can afford to allow them to accumulate at a high rate in its germ cells. We discuss later why the observed mutation frequency, low though it is, nevertheless, is thought to limit the number of essential proteins that any organism can encode in its germ line to about 60,000. By an extension of the same arguments, a mutation frequency tenfold higher would limit an organism to about 6000 essential proteins. In this case evolution would probably have stopped at an organism no more complex than a fruit fly.

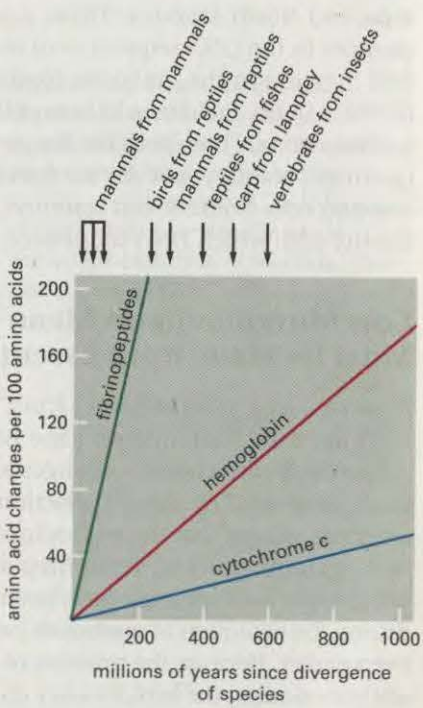


Figure 6–32 Different proteins evolve at very different rates. A comparison of the rates of amino acid change found in hemoglobin, cytochrome c, and the fibrinopeptides. Hemoglobin and cytochrome c have changed much more slowly during evolution than the fibrinopeptides. In determining rates of change per year (as in Table 6–2), it is important to realize that two species that diverged from a common ancestor 100 million years ago are separated by 200 million years of evolutionary time.

Sarepta Exhibit 1040, page 78

Table 6–2 Observed Rates of Change of the Amino Acid Sequences in Four Proteins over Evolutionary Time

Protein	Unit Evolutionary Time* (in millions of years)
Fibrinopeptide	0.7
Hemoglobin	5
Cytochrome c	21
Histone H4	500

*The "unit evolutionary time" is defined as the average time required for one *acceptable* amino acid change to appear in the indicated protein for every 100 amino acids that it contains.

While *germ* cells must be protected against high rates of mutation in order to maintain the species, the other cells of a multicellular organism (its *somatic* cells) must be protected from genetic change to safeguard the individual. Nucleotide changes in somatic cells can give rise to variant cells, some of which, through a process of natural selection, grow rapidly at the expense of the rest of the organism. In the extreme case the uncontrolled cell proliferation known as cancer results, which is responsible for about 30% of the deaths that occur in Europe and North America. These deaths are due largely to the accumulation of changes in the DNA sequences of somatic cells (discussed in Chapter 24). A tenfold increase in the mutation frequency would presumably cause a disastrous increase in the incidence of cancer by accelerating the rate at which somatic cell variants arise. Thus, both for the perpetuation of a species with 60,000 proteins (germ cell stability) and for the prevention of cancer resulting from mutations in somatic cells (somatic cell stability), eucaryotes depend on the remarkably high fidelity with which DNA sequences are maintained.

Low Mutation Rates Mean That Related Organisms Must Be Made from Essentially the Same Proteins [20]

Humans, as a genus distinct from the great apes, have existed for only a few million years. Each human gene has therefore had the chance to accumulate relatively few nucleotide changes since our inception, and most of these have been eliminated by natural selection. A comparison of humans and monkeys, for example, shows that their cytochrome c molecules differ in about 1% and their hemoglobins in about 4% of their amino acid positions. Clearly, a great deal of our genetic heritage must have been formed long before *Homo sapiens* appeared, during the evolution of mammals (which started about 300 million years ago) and even earlier. Because the proteins of mammals as different as whales and humans are very similar, the evolutionary changes that have produced such striking morphological differences must involve relatively few changes in the molecules from which we are made. Instead, it is thought that the morphological differences arise from differences in the temporal and spatial pattern of gene expression during embryonic development, which then determine the size, shape, and other characteristics of the adult. At the end of Chapter 8 we discuss the mechanisms that are thought to underlie such evolutionary changes in gene expression.

If Left Uncorrected, Spontaneous DNA Damage Would Rapidly Change DNA Sequences [21]

The physicist Erwin Schroedinger pointed out in 1945 that, whatever its chemical nature (at that time unknown), a gene must be extremely small and composed

Sarepta Exhibit 1040, page 79

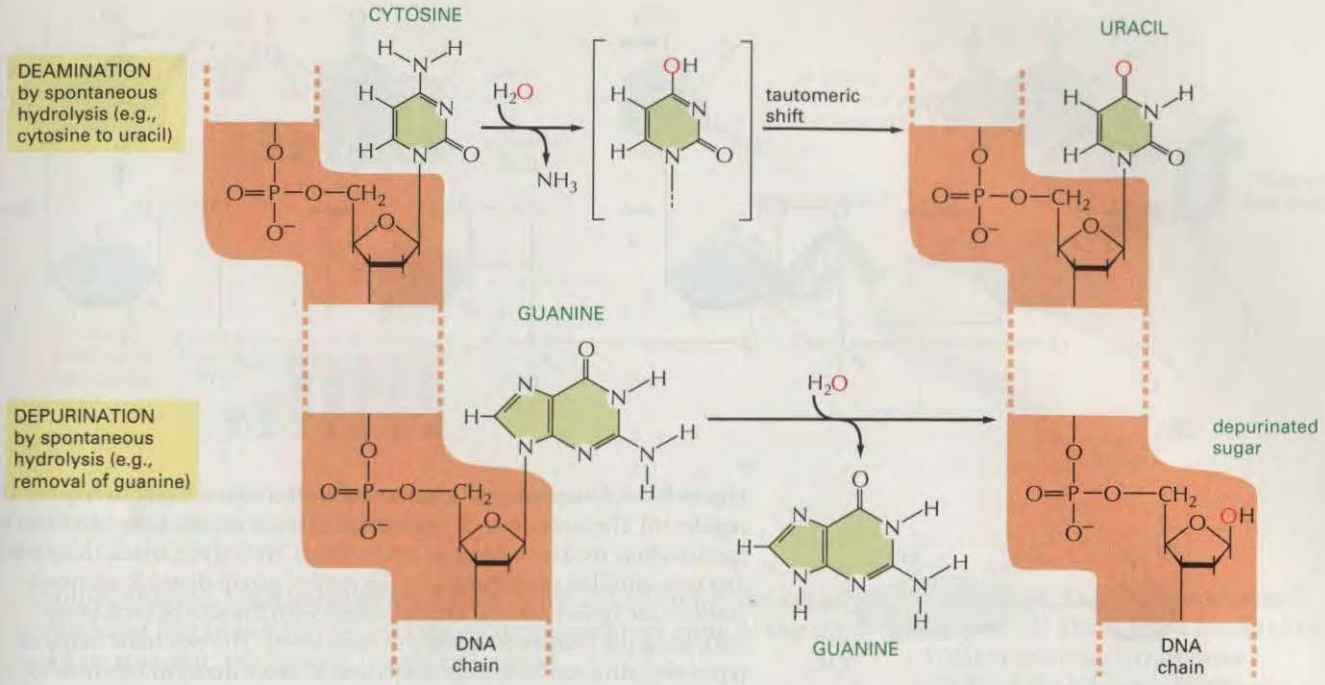


Figure 6–33 Deamination and depurination. These hydrolytic reactions are the two most frequent spontaneous chemical reactions known to create serious DNA damage in cells. Only a single example is shown for each type of reaction. (See also Figure 6–39.)

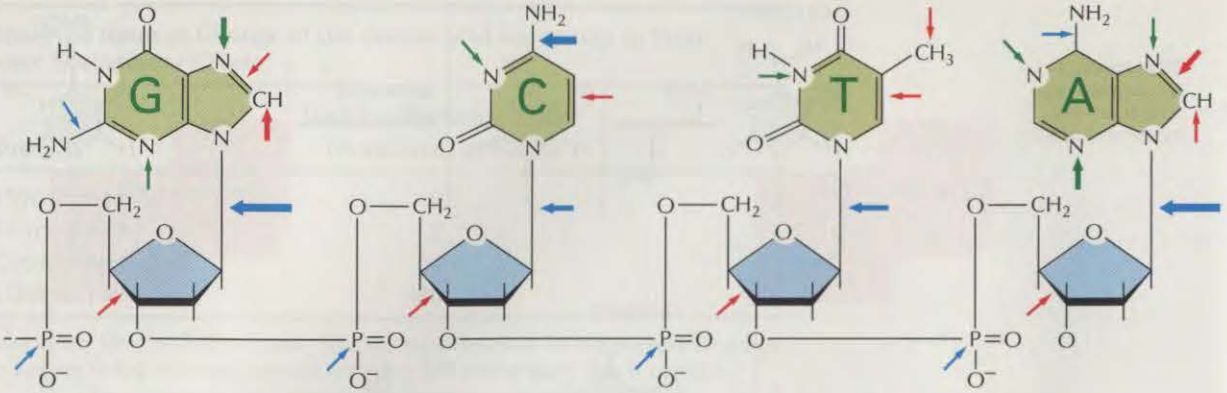
of few atoms. Otherwise the very large number of genes thought to be necessary to generate an organism would not fit in the cell nucleus. On the other hand, because it was so small, a gene would be expected to undergo significant changes as a result of spontaneous reactions induced by random thermal collisions with solvent molecules. This poses a serious dilemma, since genetic data imply that genes are composed of a remarkably stable substance in which spontaneous changes (mutations) occur rarely.

This dilemma is real. DNA does undergo major changes as a result of thermal fluctuations. We now know, for example, that about 5000 purine bases (adenine and guanine) are lost per day from the DNA of each human cell because of the thermal disruption of their *N*-glycosyl linkages to deoxyribose (*depurination*). Similarly, spontaneous *deamination* of cytosine to uracil in DNA is estimated to occur at a rate of 100 bases per genome per day (Figure 6–33). DNA bases are also subject to change by reactive metabolites (including reactive forms of oxygen) that can alter their base-pairing abilities and by ultraviolet light from the sun, which promotes a covalent linkage of two adjacent pyrimidine bases in DNA (forming, for example, the *thymine dimers* shown in Figure 6–34B). These are only a few of many changes that can occur in our DNA (Figure 6–34A). Most of them would be expected to lead either to deletion of one or more base pairs in the daughter DNA chain after DNA replication or to a base-pair substitution (each C → U deamination, for example, would eventually change a C-G base pair to a T-A base pair, since U closely resembles T and forms a complementary base pair with A). As we have seen, a high rate of such random changes would have disastrous consequences for an organism.

The Stability of Genes Depends on DNA Repair [22]

Despite the thousands of random changes created every day in the DNA of a human cell by heat energy and metabolic accidents, only a few stable changes (mutations) accumulate in the DNA sequence of an average cell in a year. We now know that fewer than one in a thousand accidental base changes in DNA causes a mutation; the rest are eliminated with remarkable efficiency by **DNA repair.**

Sarepta Exhibit 1040, page 80



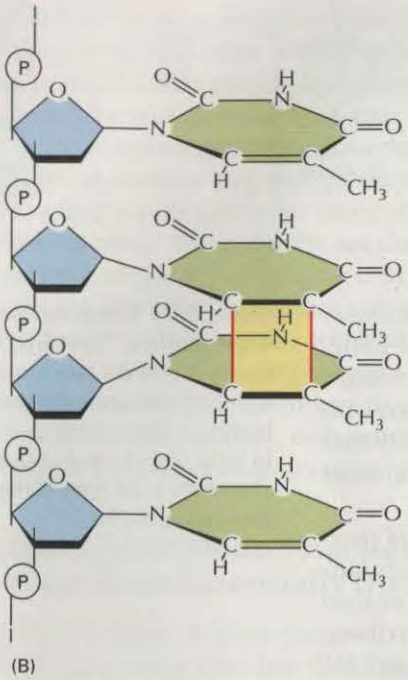


Figure 6–34 A summary of spontaneous alterations likely to require DNA repair. (A) The sites on each nucleotide that are known to be modified by spontaneous oxidative damage (*red arrows*), hydrolytic attack (*blue arrows*), and uncontrolled methylation by the methyl group donor S-adenosyl-methionine (*green arrows*) are indicated, with the size of each arrow indicating the relative frequency of each event. The two most frequent types of hydrolytic events are illustrated in more detail in Figure 6–33. (B) The thymine dimer, a type of damage introduced into DNA in cells that are exposed to ultraviolet irradiation (as in sunlight). A similar dimer will form between any two neighboring pyrimidine bases (C or T residues) in DNA. (A, after T. Lindahl, *Nature* 362:709–715, 1993. © 1993 Macmillan Magazines Ltd.)

There are a variety of repair mechanisms, each catalyzed by a different set of enzymes. Nearly all of these mechanisms depend on the existence of two copies of the genetic information, one in each strand of the DNA double helix: if the sequence in one strand is accidentally changed, information is not lost irretrievably because a complementary copy of the altered strand remains in the sequence of nucleotides in the other strand. The basic pathway for DNA repair is illustrated schematically in Figure 6–35. As indicated, it involves three steps:

1. The altered portion of a damaged DNA strand is recognized and removed by enzymes called *DNA repair nucleases,* which hydrolyze the phosphodiester bonds that join the damaged nucleotides to the rest of the DNA molecule, leaving a small gap in the DNA helix in this region.

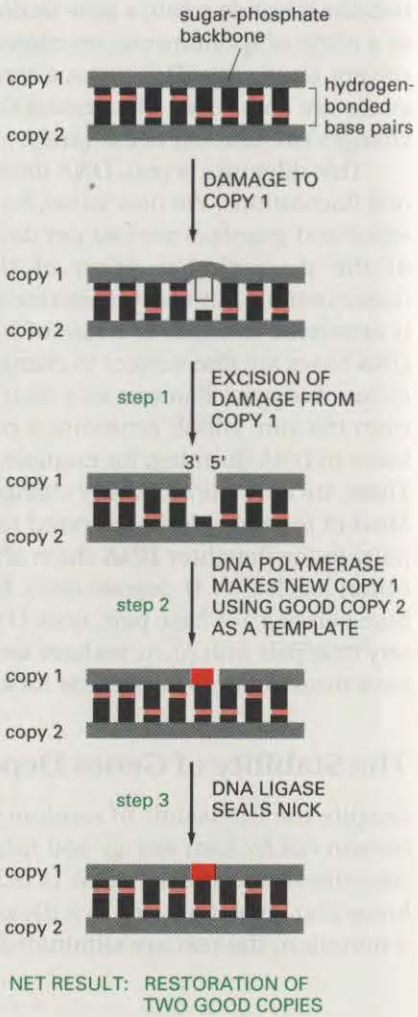
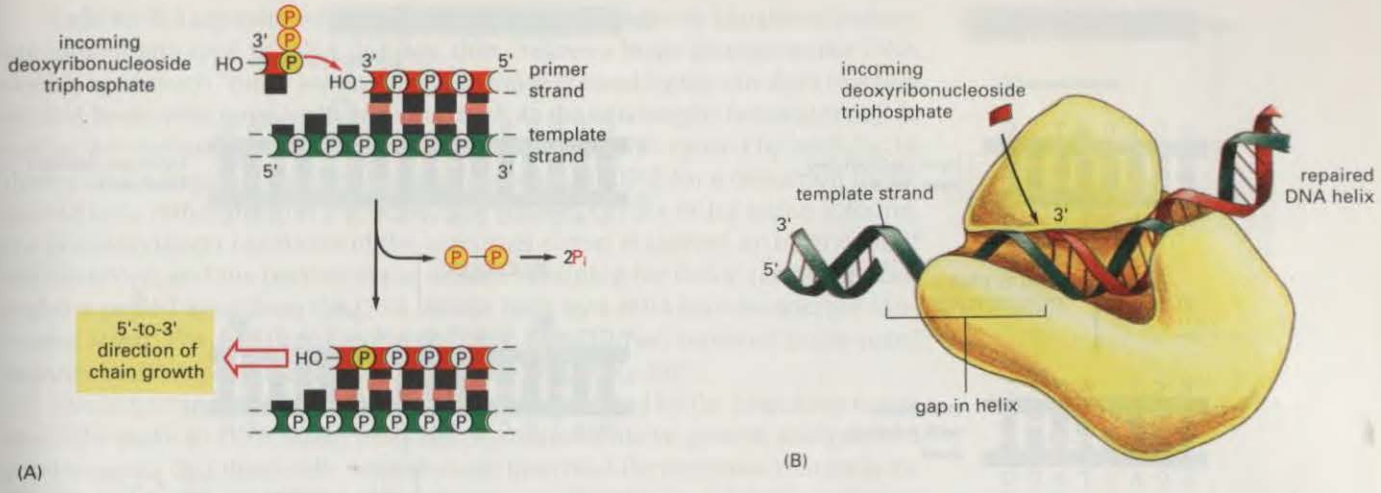


Figure 6–35 DNA repair. The three steps common to most types of repair are excision (step 1), resynthesis (step 2), and ligation (step 3). In step 1 the damage is excised; in steps 2 and 3 the original DNA sequence is restored. *DNA polymerase* fills in the gap created by the excision events, and *DNA ligase* seals the nick left in the repaired strand. Nick sealing consists of the re-formation of a broken phosphodiester bond (see Figure 6–37).

Sarepta Exhibit 1040, page 81



2. Another enzyme, **DNA polymerase,** binds to the 3′-OH end of the cut DNA strand and fills in the gap by making a complementary copy of the information stored in the "good" (*template*) strand.
3. The break or "nick" in the damaged strand left when the DNA polymerase has filled in the gap is sealed by a third type of enzyme, **DNA ligase,** which completes the restoration process.

Both DNA polymerase and DNA ligase have important general roles in DNA metabolism; both function in DNA replication as well as in DNA repair, for example. The reactions that these two enzymes catalyze are illustrated in Figures 6–36 and 6–37, respectively.

Figure 6–36 The DNA polymerase enzyme. (A) The reaction catalyzed by DNA polymerase. This enzyme catalyzes the stepwise addition of a deoxyribonucleotide to the 3′-OH end of a polynucleotide chain (the *primer strand*) that is paired to a second, *template strand.* The new DNA strand therefore grows in the 5′-to-3′ direction. Because each incoming deoxyribonucleoside triphosphate must pair with the template strand in order to be recognized by the polymerase, this strand determines which of the four possible deoxyribonucleotides (A, C, G, or T) will be added. As in the case of RNA polymerase, the reaction is driven by a large favorable free-energy change (see Figure 6–3). (B) The structure of an *E. coli* DNA polymerase molecule has been determined by x-ray crystallography. This drawing illustrates how the polymerase is thought to function during the DNA synthesis involved in DNA repair. (B, adapted from L.S. Beese, V. Derbyshire, and T.A. Steitz, *Science* 260:352–355, 1993. © 1993 the AAAS.)

DNA Damage Can Be Removed by More Than One Pathway [23]

The details of the excision step in DNA repair depend on the type of damage. Depurination, for example, which is by far the most frequent lesion that occurs in DNA, leaves a deoxyribose sugar with a missing base (see Figure 6–33). This exposed sugar is rapidly recognized by the enzyme *AP endonuclease,* which cuts the DNA phosphodiester backbone at the 5′ side of the altered site. After excision of the sugar phosphate residue by a phosphodiesterase enzyme, an undamaged DNA sequence is restored by DNA polymerase and DNA ligase (see Figure 6–35).

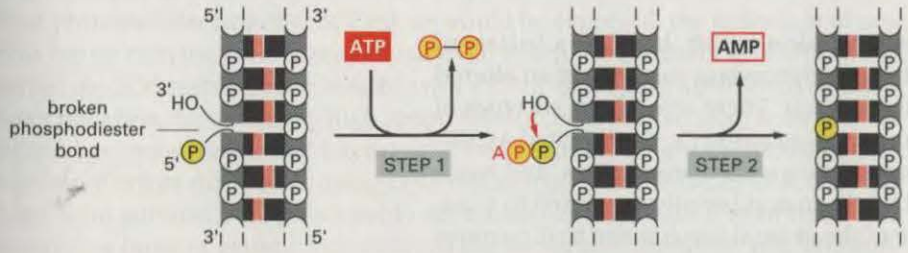


Figure 6–37 The reaction catalyzed by DNA ligase. This enzyme seals a broken phosphodiester bond. As shown, DNA ligase uses a molecule of ATP to activate the 5′ end at the nick (step 1) before forming the new bond (step 2). In this way the energetically unfavorable nick-sealing reaction is driven by being coupled to the energetically favorable process of ATP hydrolysis. In *Bloom's syndrome,* an inherited human disease, individuals are partially defective in DNA ligation and consequently are deficient in DNA repair; as a consequence, they have a dramatically increased incidence of cancer.

Sarepta Exhibit 1040, page 82

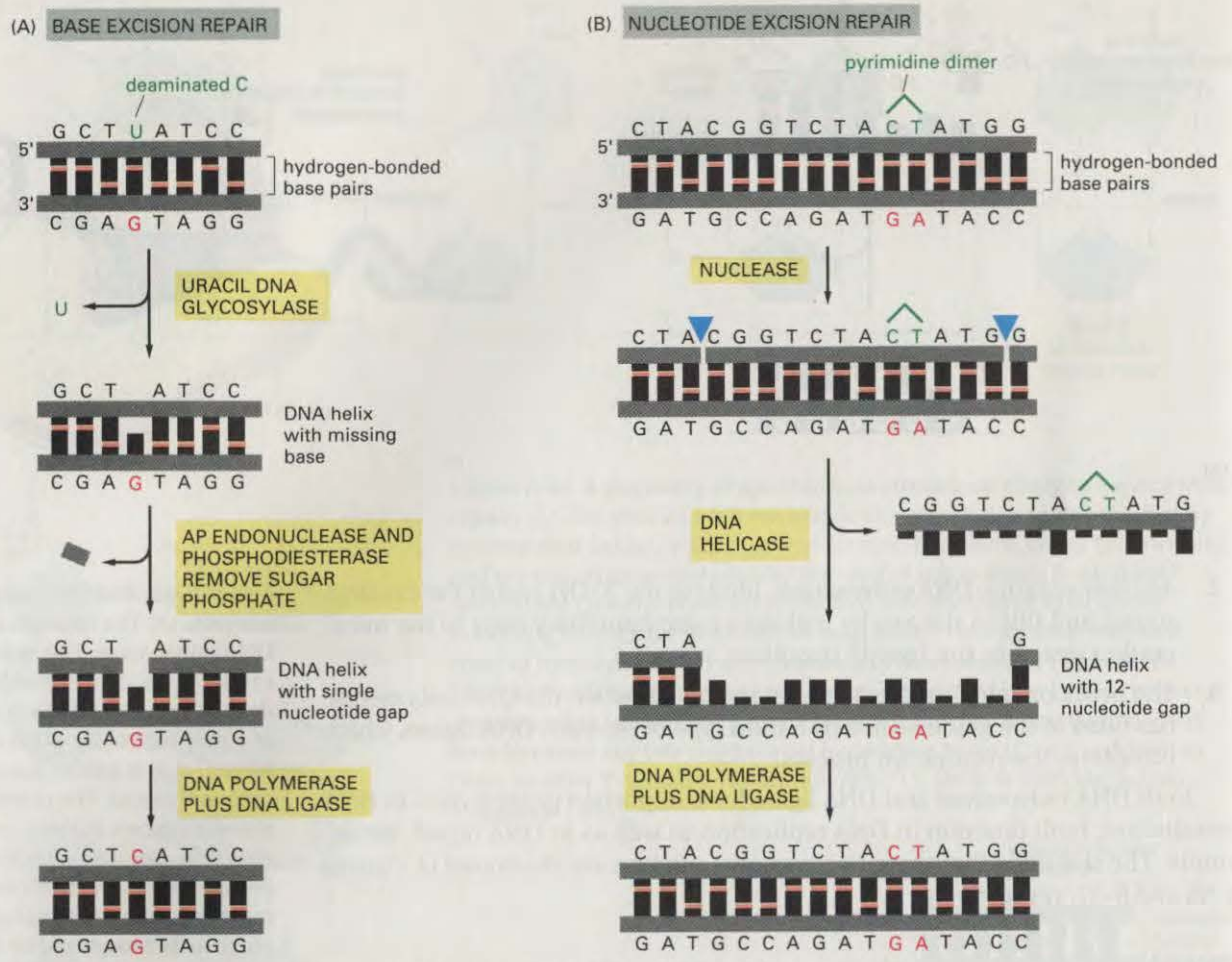


Figure 6–38 Comparison of two major DNA repair pathways. (A) *Base excision repair.* This pathway starts with a DNA glycosylase. Here the enzyme uracil DNA glycosylase removes an accidentally deaminated cytosine in DNA. After the action of this glycosylase (or another DNA glycosylase that recognizes a different kind of damage) the sugar phosphate with the missing base is cut out by the sequential action of AP endonuclease and a phosphodiesterase, the same enzymes that initiate the repair of depurinated sites. The gap of a single nucleotide is then filled by DNA polymerase and DNA ligase. The net result is that the U that was created by accidental deamination is restored to a C. The AP endonuclease derives its name from the fact that it recognizes any site in the DNA helix that contains a deoxyribose sugar with a missing base; such sites can arise either by the loss of a purine (*ap*urinic sites) or by the loss of a pyrimidine (*ap*yriminic sites). (B) *Nucleotide excision repair.* After a multienzyme complex recognizes a bulky lesion such as a pyrimidine dimer (see Figure 6–34B), one cut is made on each side of the lesion, and an associated DNA helicase then removes the entire portion of the damaged strand. The multienzyme complex in bacteria leaves the gap of 12 nucleotides shown; the gap produced in human DNA is more than twice this size.

A related repair pathway, called **base excision repair,** involves a battery of enzymes called *DNA glycosylases.* Each DNA glycosylase recognizes an altered base in DNA and catalyzes its hydrolytic removal. There are at least six types of these enzymes, including those that remove deaminated Cs, deaminated As, different types of alkylated or oxidized bases, bases with opened rings, and bases in which a carbon-carbon double bond has been accidentally converted to a carbon-carbon single bond. As an example of the general mechanism that operates in all cases, the removal of a deaminated C by uracil DNA glycosylase is shown in Figure 6–38A. The DNA glycosylase reaction produces a deoxyribose sugar with a missing base. Because this sugar phosphate is the same substrate recognized by the AP endonuclease, the subsequent steps in the repair process proceed in the same way as for depurinated sites. The importance of removing accidentally deaminated DNA bases has been directly demonstrated. In mutant bacteria that lack the enzyme uracil DNA glycosylase, the normally low spontaneous rate of change of a C-G to a T-A base pair is increased about twentyfold.

Sarepta Exhibit 1040, page 83

Cells have a separate **nucleotide excision repair** pathway capable of removing almost any type of DNA damage that creates a large change in the DNA double helix. Such "bulky lesions" include those created by the covalent reaction of DNA bases with large hydrocarbons (such as the carcinogen benzopyrene), as well as the various pyrimidine dimers (T-T, T-C, and C-C) caused by sunlight. In these cases a large multienzyme complex scans the DNA for a distortion in the double helix rather than for a specific base change. Once a bulky lesion is found, the phosphodiester backbone of the abnormal strand is cleaved on both sides of the distortion, and the portion of the strand containing the lesion (an oligonucleotide) is peeled away from the DNA double helix by a *DNA helicase* enzyme (discussed later). The gap produced in the DNA helix is then repaired in the usual manner by DNA polymerase and DNA ligase (Figure 6–38B).

The importance of these repair processes is indicated by the large investment that cells make in DNA repair enzymes. A comprehensive genetic analysis of a yeast suggests that these cells contain more than 50 different genes that code for DNA repair functions. DNA repair pathways are likely to be at least as complex in humans. Individuals with the genetic disease *xeroderma pigmentosum,* for example, are defective in a nucleotide excision repair process that can be shown by genetic analysis to require at least seven different gene products. Such individuals develop severe skin lesions, including skin cancer, because of the accumulation of pyrimidine dimers in cells that are exposed to sunlight.

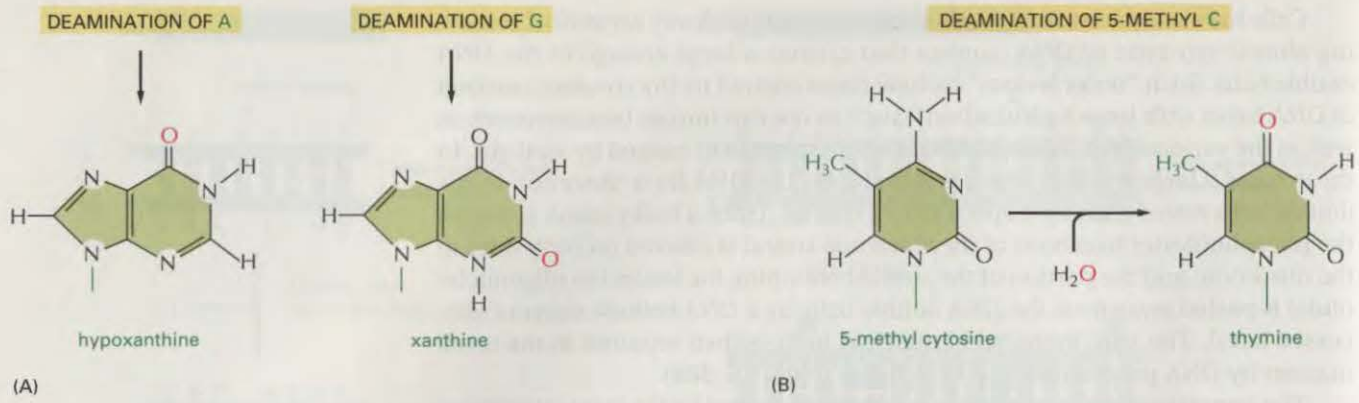
Cells Can Produce DNA Repair Enzymes in Response to DNA Damage [24]

Cells have evolved a number of mechanisms to help them survive in a hazardous world. Often an extreme environmental insult activates a battery of genes whose products protect the cell from its effects. One such mechanism shared by all cells is the *heat-shock response,* which is evoked by the exposure of cells to unusually high temperatures. The induced "heat-shock proteins" include some that are thought to help stabilize and repair partially denatured cell proteins (see Figure 5–29).

Many cells also have mechanisms that enable them to synthesize DNA repair enzymes as an emergency response to severe DNA damage. The best-studied example is the **SOS response** in *E. coli.* In this bacterium any block to DNA replication caused by DNA damage produces a signal (thought to be an excess of single-stranded DNA) that induces an increase in the transcription of more than 15 genes, many of which code for proteins that function in DNA repair. The signal first activates the *E. coli* RecA protein (discussed later), which then destroys a negatively acting gene regulatory protein (a *repressor*) that normally suppresses the transcription of the entire set of SOS response genes. Studies of mutant bacteria deficient in different parts of the response indicate that the newly synthesized proteins have two effects. First, as would be expected, the induction of new DNA repair enzymes increases cell survival. When the mutants deficient in this part of the SOS response are treated with a DNA-damaging agent such as ultraviolet radiation, an unusually high proportion of them die. Second, several of the induced proteins transiently increase the mutation rate by greatly increasing the number of errors made in copying DNA sequences. While this has little effect on short-term survival, it is presumably advantageous in the long term because it produces a burst of genetic variability in the bacterial population and hence increases the chance that a mutant cell with increased fitness will arise.

The DNA repair system activated in the SOS response is not the only inducible DNA repair system known. Bacteria have another system that is activated specifically by the presence of methylated nucleotides in DNA, and there is at least one inducible DNA repair system in yeast cells. Some higher eucaryotic cells have been reported to adapt to DNA damage in similar ways.

Sarepta Exhibit 1040, page 84



The Structure and Chemistry of the DNA Double Helix Make It Easy to Repair [25]

The DNA double helix seems to be optimally constructed for repair. As discussed in Chapter 1, RNA is thought to have evolved before DNA, and it seems likely that the genetic code was initially carried in the four nucleotides A, C, G, and U. This raises the question of why the U in RNA has been replaced in DNA by T (which is 5-methyl U). We have seen that spontaneous C deamination converts C to U but that this event is rendered harmless by uracil DNA glycosylase (see Figure 6–38A). One can imagine how any repair enzyme designed to recognize and excise such accidents would be confused by the normal U nucleotides in a U-containing DNA molecule. Thus it is not surprising that U is not used in DNA.

This line of argument is strengthened by the observation that every possible deamination event in DNA yields an unnatural base, which can therefore be directly recognized and removed by a specific DNA glycosylase. Hypoxanthine, for example, is the simplest purine base capable of pairing specifically with C, but hypoxanthine is the direct deamination product of A. The addition of a second amino group to hypoxanthine produces G, which cannot be formed from A by spontaneous deamination and whose deamination product is likewise unique (Figure 6–39A).

A special situation occurs in vertebrate DNA, where selected C nucleotides are methylated at specific CG sequences associated with inactive genes (discussed in Chapter 9). As illustrated in Figure 6–39B, the accidental deamination of these methylated C nucleotides produces the natural nucleotide T, which forms a mismatched base pair with a G on the opposite DNA strand. To help protect methylated C nucleotides against such mutations, a special DNA glycosylase recognizes a mismatched base pair involving T in the sequence TG and removes the T. This DNA repair mechanism must be relatively ineffective, however, as methylated C nucleotides are common sites for mutations in vertebrate DNA. Even though only about 3% of the C nucleotides in human DNA are methylated, mutations in these methylated nucleotides account for about one-third of the single-base mutations that have been observed in inherited human diseases (see also Figure 9–71).

Whereas the chemistry of the bases ensures that deamination will be detected, accurate repair—and the fundamental answer to Schroedinger's dilemma—depends on the existence of separate copies of the genetic information in the two strands of the double helix. Only in the very unlikely event that both strands are damaged simultaneously at the same base pair is the cell left without one good copy to serve as a template for DNA repair. Even in this case mechanisms have evolved that are sometimes able to repair the damage. These repair mechanisms require that a second DNA helix of the same sequence be present in the cell, and they use genetic recombination mechanisms to transfer the missing information from one DNA helix to another—a process called *gene conversion,* which we discuss later.

Figure 6–39 The deamination of DNA nucleotides. In each case the oxygen atom added from the reaction with water is colored *red.* (A) The spontaneous deamination products of A and G are recognizable as unnatural when they occur in DNA and thus are readily recognized and repaired. The deamination of C to U was illustrated in Figure 6–33, and T has no amino group to deaminate. (B) A few percent of the C nucleotides in vertebrate DNAs are methylated to help control gene expression. When these 5-methyl C nucleotides are accidentally deaminated, they form T. This T will be paired with a G on the opposite strand, forming a mismatched base pair.

Sarepta Exhibit 1040, page 85

Genetic information is stored in single-stranded DNA or RNA molecules only in some very small viruses with genomes of a few thousand nucleotides. The types of repair processes that we have described cannot operate on such nucleic acids, and the chance of a nucleotide change occurring in these viruses is very high. It seems that only organisms with tiny genomes can afford to encode their genetic information in a structure other than a DNA double helix.

Summary

The fidelity with which DNA sequences are maintained in higher eucaryotes can be estimated from the rates at which changes have occurred in nonessential protein and DNA sequences over evolutionary time. This fidelity is so high that a mammalian germ-line cell with a genome of 3×10^9 base pairs is subjected on average to only about 10 to 20 base-pair changes per year. But unavoidable chemical processes damage thousands of DNA nucleotides in a typical mammalian cell every day. Genetic information can be stored stably in DNA sequences only because a large variety of DNA repair enzymes continuously scan the DNA and replace the damaged nucleotides.

The process of DNA repair depends on the presence of a separate copy of the genetic information in each strand of the DNA double helix. An accidental lesion on one strand can therefore be cut out by a repair enzyme and a good strand resynthesized from the information in the undamaged strand. Most of the damage to DNA bases is excised by one of two major pathways. In base excision repair an altered base is removed by a DNA glycosylase enzyme, followed by excision of the resulting sugar phosphate. In nucleotide excision repair a small region of the strand surrounding the damage is removed from the DNA helix as an oligonucleotide. In both cases the small gap left in the DNA helix is filled in by the sequential action of DNA polymerase and DNA ligase.

DNA Replication [26]

Besides maintaining the integrity of DNA sequences by DNA repair, all organisms must duplicate their DNA accurately before every cell division. *DNA replication* occurs at polymerization rates of about 500 nucleotides per second in bacteria and about 50 nucleotides per second in mammals. Clearly, the proteins that catalyze this process must be both accurate and fast. Speed and accuracy are achieved by means of a multienzyme complex that guides the process and constitutes an elaborate "replication machine."

Base-pairing Underlies DNA Replication as well as DNA Repair [27]

DNA templating is the process in which the nucleotide sequence of a DNA strand (or selected portions of a DNA strand) is copied by complementary base-pairing (A with T or U, and G with C) into a complementary nucleic acid sequence (either DNA or RNA). The process entails the recognition of each nucleotide in the DNA strand by an unpolymerized complementary nucleotide and requires that the two strands of the DNA helix be separated, at least transiently, so that the hydrogen bond donor and acceptor groups on each base become exposed for base-pairing. The appropriate incoming single nucleotides are thereby aligned for their enzyme-catalyzed polymerization into a new nucleic acid chain. In 1957 the first such nucleotide polymerizing enzyme, *DNA polymerase*, was discovered. The substrates for this enzyme were found to be deoxyribonucleoside triphosphates, which are polymerized on a single-stranded DNA template. The stepwise mechanism of this reaction is the one previously illustrated in Figure 6–36 in connection with DNA repair. The discovery of DNA polymerase led to the isola-

Sarepta Exhibit 1040, page 86

tion of *RNA polymerase*, which was correctly inferred to use ribonucleoside triphosphates as its substrates.

During DNA replication each of the two old DNA strands serves as a template for the formation of an entire new strand. Because each of the two daughters of a dividing cell inherits a new DNA double helix containing one old and one new strand (see Figure 3–13), DNA is said to be replicated "semiconservatively" by DNA polymerase.

The DNA Replication Fork Is Asymmetrical [28]

Autoradiographic analyses carried out in the early 1960s on whole replicating chromosomes labeled with a short pulse of the radioactive DNA precursor ^{3}H-thymidine revealed a localized region of replication that moves along the parental DNA double helix. Because of its Y-shaped structure, this active region is called a **DNA replication fork.** At a replication fork the DNA of both new daughter strands is synthesized by a multienzyme complex that contains the DNA polymerase.

Initially, the simplest mechanism of DNA replication appeared to be continuous growth of both new strands, nucleotide by nucleotide, at the replication fork as it moves from one end of a DNA molecule to the other. But because of the antiparallel orientation of the two DNA strands in the DNA double helix (see Figure 3–10 and Panel 3–2, pp. 100–101), this mechanism would require one daughter strand to grow in the 5′-to-3′ direction and the other in the 3′-to-5′ direction. Such a replication fork would require two different DNA polymerase enzymes. One would polymerize in the 5′-to-3′ direction (see Figure 6–36), where each incoming deoxyribonucleoside triphosphate carries the triphosphate activation needed for its own addition. The other would move in the 3′-to-5′ direction and work by so-called "head growth," in which the end of the growing DNA chain carries the triphosphate activation required for the addition of each subsequent nucleotide. Although head-growth polymerization occurs elsewhere in biochemistry (see Figure 2–36), it does not occur in DNA synthesis; no 3′-to-5′ DNA polymerase has ever been found (Figure 6–40).

How, then, is 3′-to-5′ DNA synthesis achieved? The answer was first suggested in the late 1960s by experiments in which highly radioactive ^{3}H-thymidine was added to dividing bacteria for a few seconds so that only the most recently replicated DNA, just behind the replication fork, became radiolabeled. This selective labeling method revealed the transient existence of pieces of DNA that were 1000 to 2000 nucleotides long, now commonly known as *Okazaki fragments*, at the bacterial growing fork. (Such replication intermediates were later found in eucaryotes, where they are only 100 to 200 nucleotides long.) The Okazaki fragments were shown to be synthesized only in the 5′-to-3′ chain direction and to

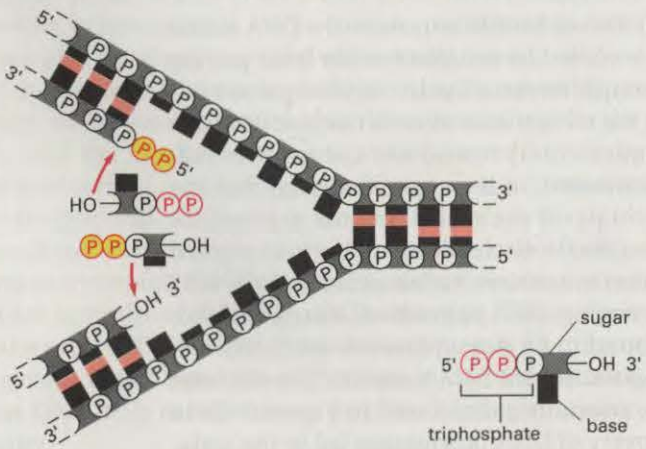


Figure 6–40 An incorrect model for DNA replication. Although it might appear to be the simplest mechanism for DNA replication, the mechanism illustrated here is not the one that cells use. Note that in this scheme both daughter DNA strands would grow continuously, using the energy of hydrolysis of the *yellow* phosphates to add the next nucleotide on each strand. This would require chain growth in both the 5′-to-3′ direction (*bottom*) and the 3′-to-5′ direction (*top*). No enzyme that catalyzes 3′-to-5′ nucleotide polymerization has ever been found.

Sarepta Exhibit 1040, page 87

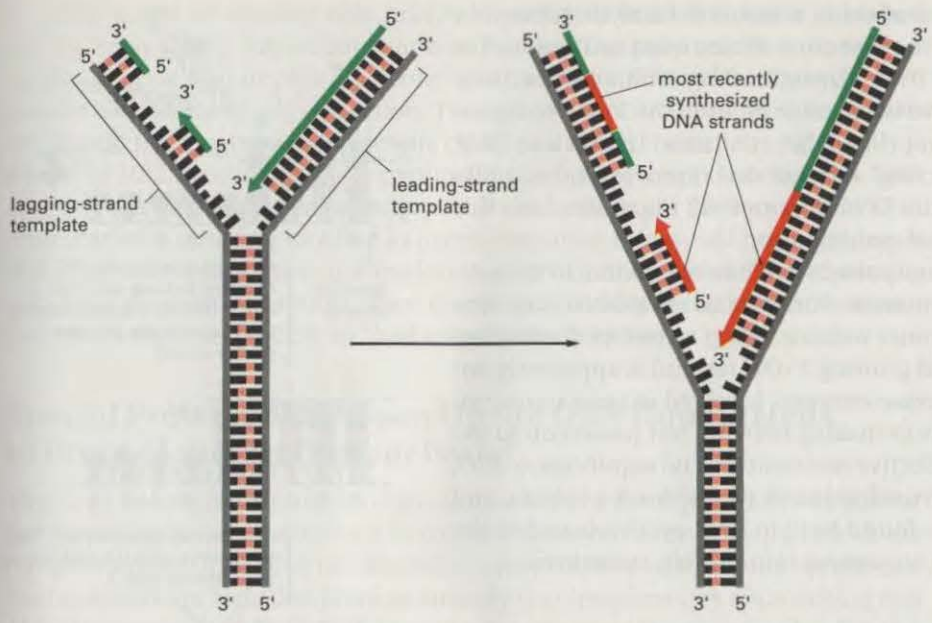


Figure 6–41 The structure of a DNA replication fork. Because both daughter DNA strands (*colored*) are synthesized in the 5′-to-3′ direction, the DNA synthesized on the lagging strand must be made initially as a series of short DNA molecules, called *Okazaki fragments.*

be joined together after their synthesis to create long DNA chains by the same *DNA ligase* enzyme that seals nicks during DNA repair (see Figure 6–37).

A replication fork has an asymmetric structure. The DNA daughter strand that is synthesized continuously is known as the **leading strand,** and its synthesis slightly precedes the synthesis of the daughter strand that is synthesized discontinuously, which is known as the **lagging strand.** The synthesis of the lagging strand is delayed because it must wait for the leading strand to expose the template strand on which each Okazaki fragment is synthesized (Figure 6–41). The synthesis of the lagging strand by a discontinuous, "backstitching" mechanism means that only the 5′-to-3′ type of DNA polymerase is needed for DNA replication.

The High Fidelity of DNA Replication Requires a Proofreading Mechanism [29]

The fidelity of copying that is observed after DNA replication has occurred is such that only about 1 error is made in every 10^9 base-pair replications, as required to maintain the mammalian genome of 3×10^9 DNA base pairs. This fidelity is much higher than expected, given that the standard complementary base pairs are not the only ones possible. With small changes in helix geometry, for example, two hydrogen bonds will form between G and T in DNA. In addition, rare tautomeric forms of the four DNA bases occur transiently in ratios of 1 part to 10^4 or 10^5. These forms will mispair without a change in helix geometry: the rare tautomeric form of C pairs with A instead of G, for example. If the DNA polymerase accepts a mispairing that occurs between an incoming deoxyribonucleoside triphosphate and the DNA template, the wrong nucleotide can be incorporated into the new DNA chain, producing a mutation. The high fidelity of DNA replication depends on several "proofreading" mechanisms that act sequentially to remove errors brought about in these ways.

One important proofreading process depends on special properties of the DNA polymerase enzyme. Unlike RNA polymerases, DNA polymerases do not begin a new polynucleotide chain by linking two nucleoside triphosphates together. They absolutely require the 3′-OH end of a base-paired *primer strand* on which to add further nucleotides (see Figure 6–36). Moreover, DNA molecules with a mismatched (not base-paired) nucleotide at the 3′-OH end of the primer strand are not effective as templates. DNA polymerase molecules are able to deal

Sarepta Exhibit 1040, page 88

with such mismatched DNAs by means of either a separate catalytic subunit or a covalently linked, separate catalytic site that clips off any unpaired residues at the primer terminus. Clipping by this *3′-to-5′ proofreading exonuclease* activity continues until enough nucleotides have been removed from the 3′ end to regenerate a base-paired terminus that can prime DNA synthesis. In this way DNA polymerase functions as a "self-correcting" enzyme that removes its own polymerization errors as it moves along the DNA. Figure 6–42 illustrates how this proofreading process can correct a base-pairing error.

The requirement for a perfectly base-paired terminus is essential to the self-correcting properties of the DNA polymerase. For such an enzyme to start synthesis in the complete absence of a primer without losing any of its discrimination between base-paired and unpaired growing 3′-OH termini is apparently not possible. By contrast, the RNA polymerase enzymes involved in gene transcription need not be self-correcting: errors in making RNA are not passed on to the next generation, and an occasional defective molecule has no significance. RNA polymerases are able to start new polynucleotide chains without a primer, and an error frequency of about 1 in 10^4 is found both in RNA synthesis and in the separate process of translating mRNA sequences into protein sequences.

Only DNA Replication in the 5′-to-3′ Direction Allows Efficient Error Correction

The need for accuracy probably explains why DNA replication occurs only in the 5′-to-3′ direction. If there were a DNA polymerase that added deoxyribonucleoside triphosphates in such a way as to cause chains to grow in the 3′-to-5′ chain direction, the growing 5′-chain end rather than the incoming mononucleotide would carry the activating triphosphate. In this case the mistakes in polymerization could not be simply hydrolyzed away, since the bare 5′-chain end thus created would immediately terminate DNA synthesis. It is much easier, therefore, to correct a mismatched base that has just been added to the 3′ end than one that has just been added to the 5′ end of a DNA chain. Although the type of mechanism for DNA replication shown in Figure 6–41 seems at first sight much more complex than the incorrect mechanism depicted in Figure 6–40, it is much more accurate because it involves DNA synthesis only in the 5′-to-3′ direction.

A Special Nucleotide Polymerizing Enzyme Synthesizes Short RNA Primer Molecules on the Lagging Strand [30]

For the leading strand a special primer is needed only at the start of replication; once a replication fork is established, the DNA polymerase is continuously presented with a base-paired chain end on which to add new nucleotides. But the DNA polymerase on the lagging side of the fork requires only about 4 seconds to complete each short DNA fragment, after which it must start synthesizing a completely new fragment at a site farther along the template strand (see Figure 6–41). A special mechanism is needed to produce the base-paired primer strand required by this DNA polymerase molecule. The mechanism involves an enzyme called *DNA primase*, which uses ribonucleoside triphosphates to synthesize short **RNA primers** (Figure 6–43). These primers are about 10 nucleotides long in eucaryotes, and they are made at intervals on the lagging strand, where they are elongated by the DNA polymerase to begin each Okazaki fragment. The synthesis of each Okazaki fragment ends when this DNA polymerase runs into the RNA primer attached to the 5′ end of the previous fragment. To produce a continuous DNA chain from the many DNA fragments made on the lagging strand, a special DNA repair system acts quickly to erase the old RNA primer and replace it with DNA. DNA ligase then joins the 3′ end of the new DNA fragment to the 5′ end of the previous one to complete the process (Figure 6–44).

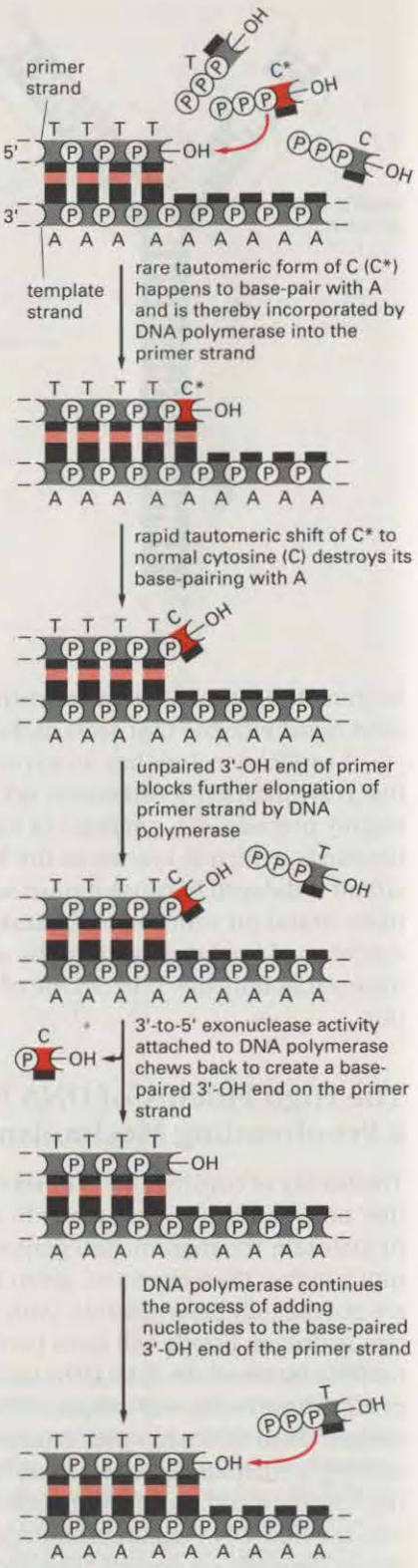


Figure 6–42 Proofreading during DNA replication.

Sarepta Exhibit 1040, page 89

Why might an erasable RNA primer be preferred to a DNA primer that need not be erased? The argument that a self-correcting polymerase cannot start chains *de novo* also implies its converse: an enzyme that starts chains *de novo* cannot be efficient at self-correction. Thus any enzyme that primes the synthesis of Okazaki fragments will of necessity make a relatively inaccurate copy (at least 1 error in 10^5). Even if the copies retained in the final product constituted as little as 5% of the total genome (for example, 10 nucleotides per 200-nucleotide DNA fragment), the resulting increase in overall mutation rate would be enormous. It therefore seems likely that the evolution of RNA rather than DNA for priming entailed a powerful advantage, since the ribonucleotides in the primer automatically mark these sequences as "bad copy" to be removed.

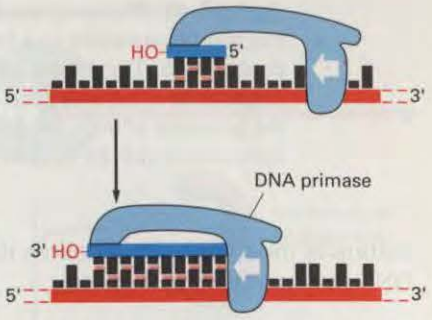


Figure 6–43 RNA primer synthesis. A schematic view of the reaction catalyzed by DNA primase, the enzyme that synthesizes the short RNA primers made on the lagging strand. Unlike DNA polymerase, this enzyme can start a new polynucleotide chain by joining two nucleoside triphosphates together. The primase stops after a short polynucleotide has been synthesized and makes the 3′ end of this primer available for the DNA polymerase.

Special Proteins Help Open Up the DNA Double Helix in Front of the Replication Fork [31]

The DNA double helix must be opened up ahead of the replication fork so that the incoming deoxyribonucleoside triphosphates can form base pairs with the template strand. The DNA double helix is very stable under normal conditions: the base pairs are locked in place so strongly that temperatures approaching that of boiling water are required to separate the two strands in a test tube. For this reason most DNA polymerases can copy DNA only when the template strand has already been separated from its complementary strand. Additional proteins are needed to help open the double helix and thus provide the appropriate exposed DNA template for the DNA polymerase to copy. Two types of replication proteins contribute to this process—DNA helicases and single-strand DNA-binding proteins.

DNA helicases were first isolated as proteins that hydrolyze ATP when they are bound to single strands of DNA. As described in Chapter 5, the hydrolysis of ATP can change the shape of a protein molecule in a cyclical manner that allows the protein to perform mechanical work. DNA helicases utilize this principle to move rapidly along a DNA single strand; where they encounter a region of double helix, they continue to move along their strand, thereby prying apart the helix (Figure 6–45). We have previously described how a special DNA repair helicase functions in nucleotide excision repair (see Figure 6–38B).

The unwinding of the template DNA helix at a replication fork could in principle be catalyzed by two DNA helicases acting in concert—one running along the leading strand and one along the lagging strand. These two helicases would need to move in opposite directions along a DNA single strand and therefore would have to be different enzymes. Both types of DNA helicase, in fact, do exist, although in bacteria the DNA helicase on the lagging strand plays the predominant role, for reasons that will become clear shortly.

Single-strand DNA-binding (SSB) proteins—also called *helix-destabilizing proteins*—bind to exposed DNA strands without covering the bases, which therefore remain available for templating. These proteins are unable to open a long DNA helix directly, but they aid helicases by stabilizing the unwound, single-stranded conformation. In addition, their cooperative binding completely coats the regions of single-stranded DNA on the lagging strand, thereby preventing for-

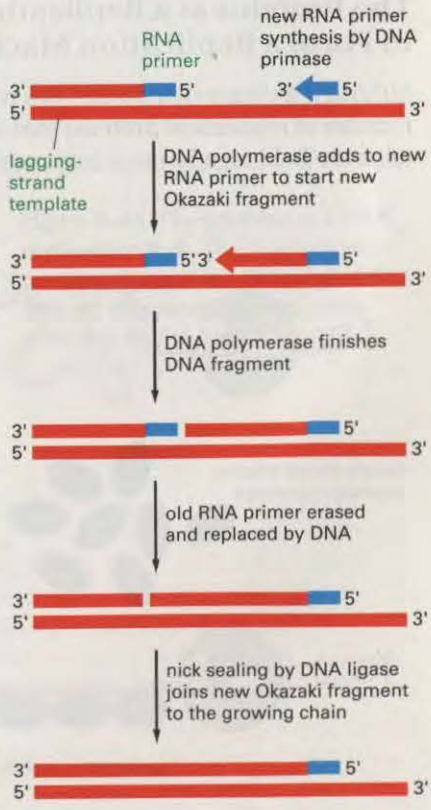


Figure 6–44 The synthesis of one of the many DNA fragments on the lagging strand. In eucaryotes the RNA primers are made at intervals spaced by about 200 nucleotides on the lagging strand, and each RNA primer is 10 nucleotides long. This primer is erased by a special DNA repair enzyme that recognizes an RNA strand in an RNA/DNA helix and excises it; this leaves a gap that is filled in by DNA polymerase and DNA ligase, as we saw for the DNA repair process (see Figure 6–35).

Sarepta Exhibit 1040, page 90

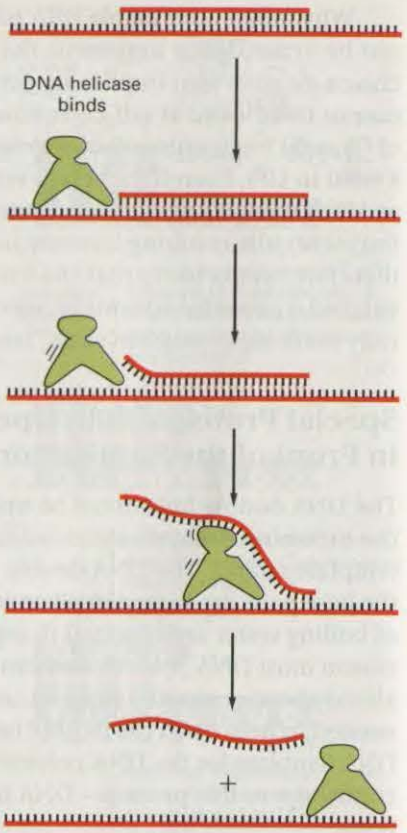


Figure 6–45 The assay used to test for DNA helicase enzymes. A short DNA fragment is annealed to a long DNA single strand to form a region of DNA double helix. The double helix is melted as the helicase runs along the DNA single strand, releasing the short DNA fragment in a reaction that requires the presence of both the helicase protein and ATP. The movement of the helicase is powered by its ATP hydrolysis (see Figure 5–22).

mation of the short hairpin helices that would otherwise impede synthesis by the DNA polymerase (Figure 6–46).

A Moving DNA Polymerase Molecule Is Kept Tethered to the DNA by a Sliding Ring [32]

On their own, most DNA polymerase molecules will synthesize only a short string of nucleotides before falling off a DNA template. This tendency to leave a DNA molecule quickly allows the DNA polymerase molecule that has just finished synthesizing one Okazaki fragment on the lagging strand to be recycled quickly to begin the synthesis of the next Okazaki fragment on the same strand. This rapid dissociation, however, would make it difficult for the polymerase to synthesize long DNA strands at a replication fork were it not for an accessory protein that functions as a regulated clamp. This clamp keeps the polymerase firmly on the DNA when it is moving, but releases it as soon as the polymerase stops.

How can a clamp prevent the polymerase from dissociating without at the same time impeding the polymerase's rapid movement along the DNA molecule? The three-dimensional structure of a clamp protein, determined by x-ray diffraction, indicates that it forms a large ring around the DNA helix. One side of the ring binds to the back of the DNA polymerase, and the whole ring slides freely as the polymerase moves along a DNA strand (Figure 6–47). The assembly of the clamp around DNA requires ATP hydrolysis by special accessory proteins that bind both to the clamp protein and to DNA; it is not known how the clamp is disassembled to remove it from the DNA.

The Proteins at a Replication Fork Cooperate to Form a Replication Machine [33]

Although we have discussed DNA replication as though it were carried out by a mixture of replication proteins that act independently, in reality most of the proteins are held together in a large multienzyme complex that moves rapidly along

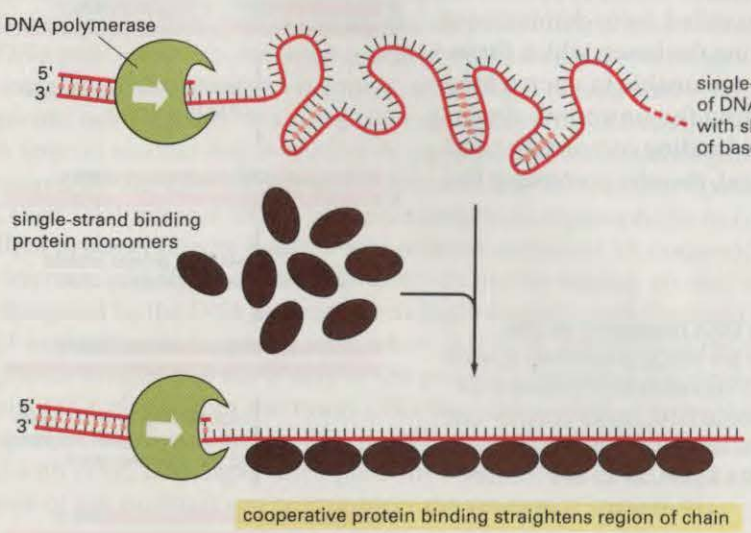
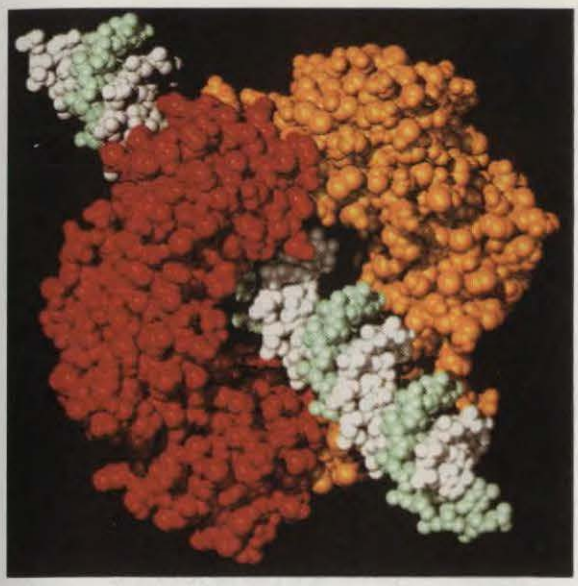
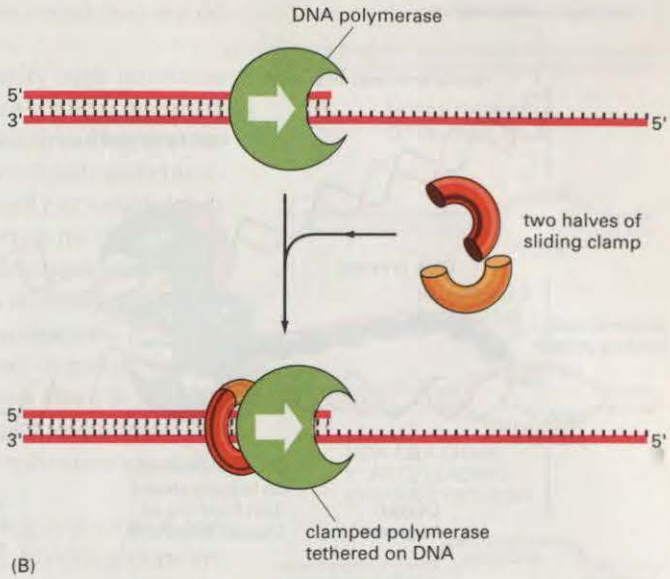


Figure 6–46 The effect of single-strand binding proteins on the structure of single-stranded DNA. Because each protein molecule prefers to bind next to a previously bound molecule (*cooperative binding*) long rows of this protein will form on a DNA single strand. This cooperative binding straightens out the DNA template and facilitates the DNA polymerization process. The "hairpin helices" shown in the bare single-stranded DNA result from a chance matching of short regions of complementary nucleotide sequence; they are similar to the short helices that typically form in RNA molecules.

Sarepta Exhibit 1040, page 91



(A)



(B)

Figure 6–47 The regulated sliding clamp that holds DNA polymerase on the DNA. (A) The structure of the sliding clamp from *E. coli*, with a DNA helix added to indicate how the protein fits around DNA. A similar protein is present in eucaryotic cells. (B) Schematic illustration of how the clamp is thought to hold a moving DNA polymerase molecule on the DNA. (A, from X.-P. Kong et al., *Cell* 69:425–437, 1992. © Cell Press.)

the DNA. This complex can be likened to a tiny sewing machine composed of protein parts and powered by nucleoside triphosphate hydrolyses. Although the replication complex has been best characterized in *E. coli* and several of its viruses, a very similar complex operates in eucaryotes (see p. 358).

The functions of the subunits of the replication machine are summarized in the two-dimensional diagram of the complete replication fork shown in Figure 6–48. Two identical DNA polymerase molecules work at the fork, one on the leading strand and one on the lagging strand. The DNA helix is opened by a DNA polymerase molecule clamped on the leading strand, acting in concert with a DNA helicase molecule running along the lagging strand; helix opening is aided by cooperatively bound molecules of single-strand DNA-binding protein. While the DNA polymerase molecule on the leading strand can operate in a continuous fashion, the DNA polymerase molecule on the lagging strand must restart at short intervals, using a short RNA primer made by a DNA primase molecule.

The efficiency of replication is greatly increased by the close association of all these protein components. The primase molecule is linked directly to the DNA

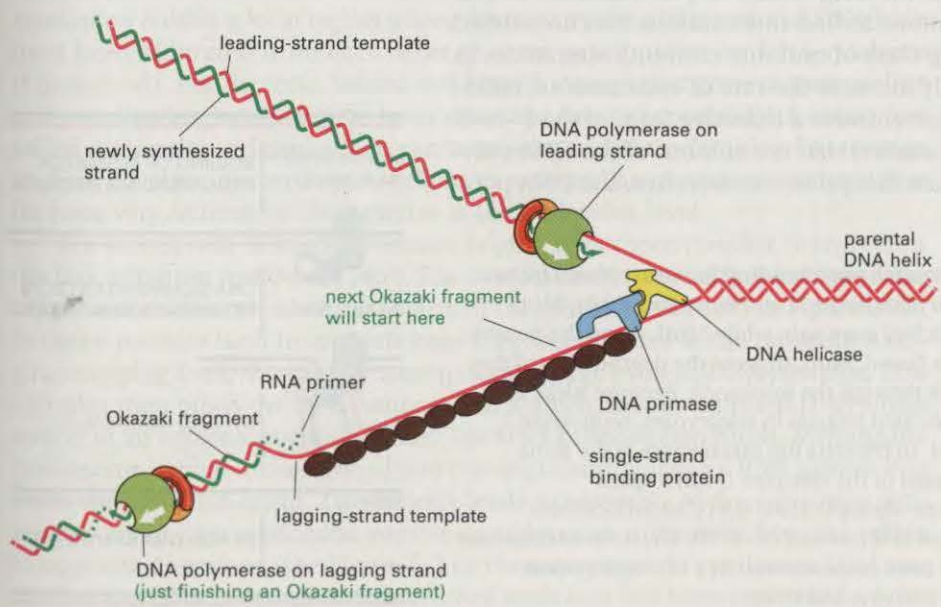


Figure 6–48 The proteins at a DNA replication fork. The major types of proteins that act at a DNA replication fork are illustrated, showing their positions on the DNA.

Sarepta Exhibit 1040, page 92

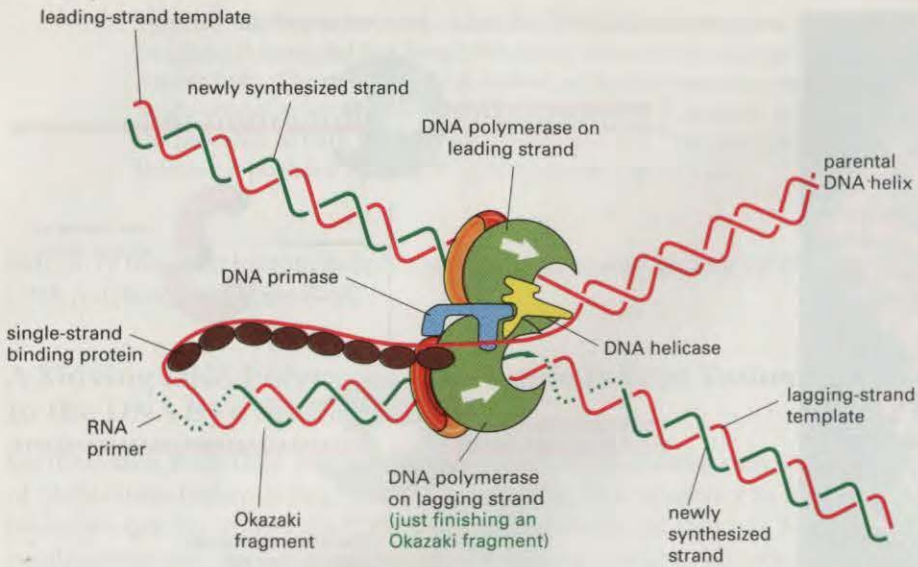


Figure 6–49 A replication fork in three dimensions. This diagram shows a current view of how the replication proteins are arranged at a replication fork when the fork is moving. The two-dimensional structure of Figure 6–48 has been altered by folding the DNA on the lagging strand to bring the lagging-strand DNA polymerase molecule into a complex with the leading-strand DNA polymerase molecule. This folding process also brings the 3′ end of each completed Okazaki fragment close to the start site for the next Okazaki fragment (compare with Figure 6–48). Because the lagging-strand DNA polymerase molecule is held to the rest of the replication proteins, it can be reused to synthesize successive Okazaki fragments; thus it is about to let go of its completed DNA fragment and move to the RNA primer that will be synthesized nearby, as required to start the next DNA fragment. Note that one daughter DNA helix extends toward the bottom right and the other toward the top left in this diagram.

helicase to form a unit on the lagging strand called a **primosome**. Powered by the DNA helicase, the primosome moves with the fork, synthesizing RNA primers as it goes. Similarly, the DNA polymerase molecule that synthesizes DNA on the lagging strand moves in concert with the rest of the proteins, synthesizing a succession of new Okazaki fragments. To accommodate this arrangement, its DNA template strand is thought to be folded back in the manner shown in Figure 6–49. The replication proteins are thus linked together into a single large unit (total mass > 10^6 daltons) that moves rapidly along the DNA, enabling DNA to be synthesized on both sides of the fork in a coordinated and efficient manner.

This DNA replication machine leaves behind on the lagging strand a series of unsealed Okazaki fragments, which still contain the RNA that primed their synthesis at their 5′ ends. This RNA must be removed and the fragments joined up by DNA repair enzymes that operate behind the replication fork (see Figure 6–44).

A Mismatch Proofreading System Removes Replication Errors That Escape from the Replication Machine [34]

Bacteria such as *E. coli* are capable of dividing once every 30 minutes, so it is relatively easy to screen large populations to find rare mutants that are altered in a specific process. One interesting class of mutants contains alterations in so-called *mutator genes*, which greatly increase the rate of spontaneous mutation. Not surprisingly, one such mutant encodes a defective form of the 3′-to-5′ proofreading exonuclease (discussed earlier) that is a subunit of the DNA polymerase enzyme (see Figure 6–42). When this protein is defective, the DNA poly-

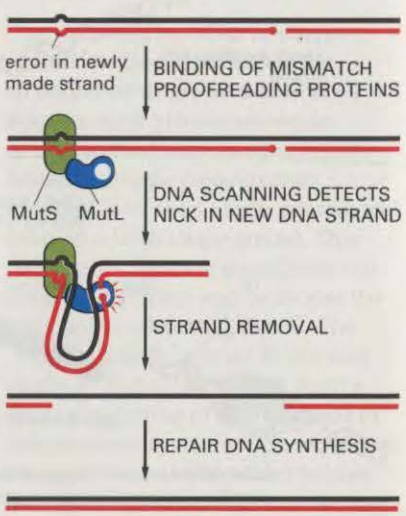


Figure 6–50 A model for mismatch proofreading in eucaryotes. The two proteins shown are present in both bacteria and eucaryotic cells: MutS binds specifically to a mismatched base pair, while MutL scans the nearby DNA for a nick. Once a nick is found, MutL triggers the degradation of the nicked strand all the way back through the mismatch. Because nicks are largely confined to newly replicated strands in eucaryotes, replication errors are selectively removed. In bacteria the mechanism is the same except that an additional protein in the complex (MutH) nicks unmethylated (and therefore newly replicated) GATC sequences and thereby begins the process that is illustrated here. We know the mechanism because these reactions have been reconstituted in a cell-free system containing purified bacterial proteins and DNA.

Sarepta Exhibit 1040, page 93

merase no longer proofreads effectively, and many replication errors that would otherwise have been removed accumulate in the DNA.

The study of other *E. coli* mutants that exhibit abnormally high mutation rates has uncovered another proofreading system that removes replication errors missed by the proofreading exonuclease. This **mismatch proofreading** system (also called a *mismatch repair* system) differs from most DNA repair systems in that it does not depend on the presence in the DNA of abnormal nucleotides that can be recognized and excised. Instead, it detects the distortion on the outside of the helix that results from the misfit between noncomplementary base pairs. But if the proofreading system simply recognized a mismatch in newly replicated DNA and randomly excised one of the two mismatched nucleotides, it would make the mistake of "correcting" the original template strand to match the error exactly half the time and would not therefore lower the overall error rate. To be effective, the proofreading system must be able to distinguish and remove the mismatched nucleotide only on the new strand, where the replication error occurred.

The recognition system used by the mismatch proofreading system in *E. coli* depends on the methylation of selected A residues in the DNA. Methyl groups are added to all A residues in the sequence GATC, but not until some time after the A has been incorporated into a newly synthesized DNA chain. Because only the new strands just behind a replication fork will contain GATC sequences that have not yet been methylated, these new DNA strands can be distinguished from old ones.

More recently, eucaryotic proteins have been discovered that are homologous in their amino acid sequence to several of the bacterial proteins that catalyze mismatch proofreading. As expected, when the genes that encode these proteins are deleted in a yeast cell, mutation rates can increase by 100-fold or more. There must, however, be some important differences between the bacterial and eucaryotic proofreading mechanisms, as the mechanism for distinguishing the newly synthesized strand from the parental template strand at the site of a mismatch cannot depend on DNA methylation as in bacteria, since some eucaryotes, such as yeasts and *Drosophila*, do not methylate any of their DNA. Newly synthesized DNA strands are known to be preferentially *nicked*, and it has been suggested that such nicks (single-strand breaks) provide the signal that directs mismatch proofreading to the appropriate strand in a eucaryotic cell (Figure 6–50).

Replication Forks Initiate at Replication Origins [35]

In both bacteria and mammals replication forks originate at a structure called a *replication bubble*, a local region where the two strands of the parental DNA helix have been separated from each other to serve as templates for DNA synthesis (Figure 6–51). For bacteria, yeasts, and several viruses that grow in mammalian cells, replication bubbles have been shown to form at special DNA sequences called **replication origins,** which can be as long as 300 nucleotides. For reasons that are not clear, the replication origins in mammalian chromosomes have thus far been very difficult to characterize at the molecular level.

For several well-defined replication origins, it has been possible to reproduce the fork initiation reaction *in vitro*. The *in vitro* studies reveal that fork initiation in bacteria and bacterial viruses starts in the manner indicated in Figure 6–52. *Initiator proteins* bind in multiple copies to specific sites at the replication origin, wrapping the DNA around them to form a large protein-DNA complex. This complex then binds the DNA helicase and loads it onto an exposed DNA single strand in an adjacent region of helix. The DNA primase also binds, forming the primosome, which moves away from the origin and makes an RNA primer that starts the first DNA chain. This quickly leads to assembly of the remaining proteins to create two replication protein complexes moving away from the origin in opposite directions (see Figure 6–51); these continue to synthesize DNA until all of the DNA template downstream of each fork has been replicated.

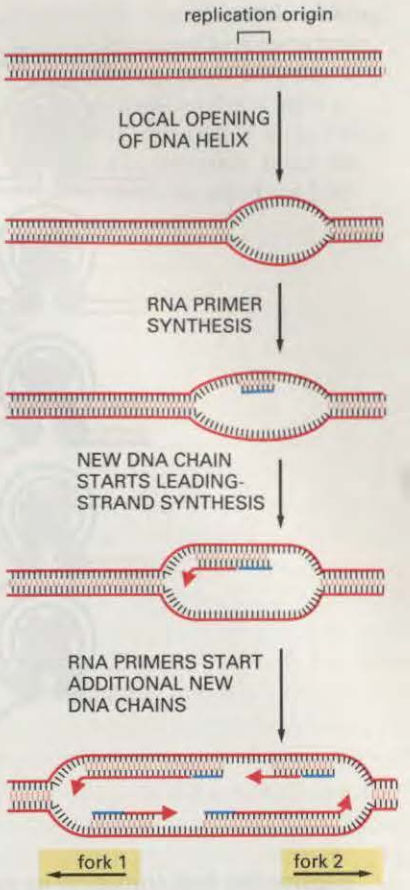


Figure 6–51 Replication fork initiation. The figure outlines the processes involved in the initiation of replication forks at replication origins. (See also Figure 6–52.)

Sarepta Exhibit 1040, page 94

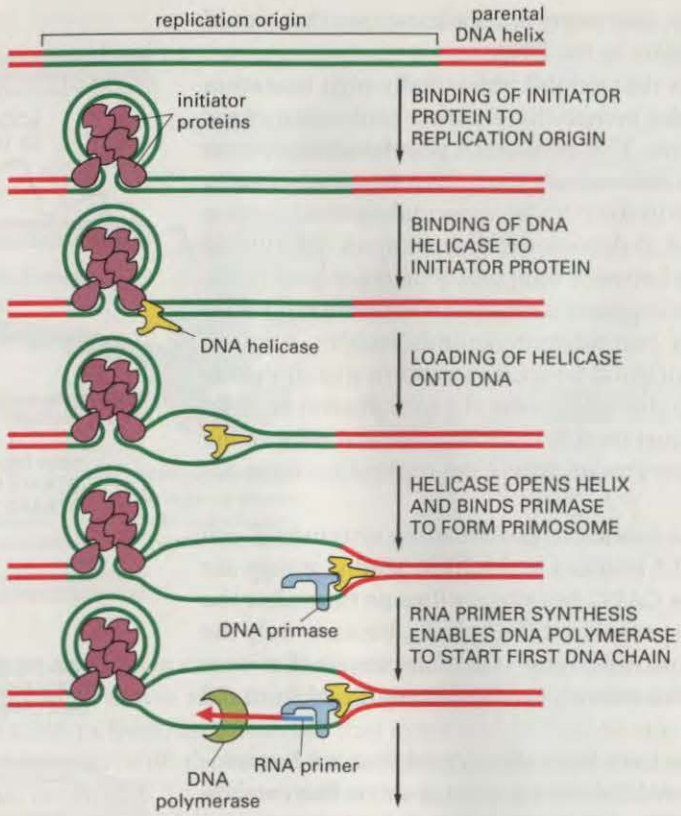


Figure 6–52 The proteins that initiate DNA replication. The major types of proteins involved in the formation of replication forks at the *E. coli* and bacteriophage lambda replication origins are indicated. The mechanism shown was established by *in vitro* studies utilizing a mixture of highly purified proteins. Subsequent steps result in the initiation of three more DNA chains (see Figure 6–51) by a pathway that is not yet clear. For *E. coli* DNA replication, the major initiator protein is the dnaA protein; for both lambda and *E. coli*, the primosome is composed of the dnaB (DNA helicase) and dnaG (DNA primase) proteins.

Replication fork initiation in eucaryotic chromosomes is discussed in detail in Chapter 8.

DNA Topoisomerases Prevent DNA Tangling During Replication [36]

When we draw the DNA helix (incorrectly) as a flat, ladderlike structure, we are ignoring the "winding problem" that arises during DNA replication. Every 10 base pairs replicated at the fork correspond to one complete turn about the axis of the parental double helix. Therefore, for a replication fork to move, the entire chromosome ahead of the fork would normally have to rotate rapidly (Figure 6–53), which would require large amounts of energy for long chromosomes. An alternative strategy is used during DNA replication: a swivel is formed in the DNA helix by proteins known as **DNA topoisomerases.**

A DNA topoisomerase can be viewed as a reversible nuclease that adds itself covalently to a DNA phosphate, thereby breaking a phosphodiester bond in a DNA strand. Because the covalent linkage that joins a topoisomerase to a DNA phosphate retains the energy of the cleaved phosphodiester bond, the cleavage reaction is reversible; resealing is rapid and does not require additional energy input. The rejoining mechanism is different in this respect from that of the enzyme DNA ligase, discussed previously (see Figure 6–37).

One type of topoisomerase (*topoisomerase I*) causes a *single-strand break* (or *nick*), which can allow the two sections of DNA helix on either side of the nick to rotate freely relative to each other, using the phosphodiester bond in the strand opposite the nick as a swivel point (Figure 6–54). Any tension in the DNA helix will drive this rotation in the direction that relieves the tension. As a result, DNA replication can occur with the rotation of only a short length of helix—the part just ahead of the fork. The analogous problem that arises during DNA transcription is solved in a similar way.

A second type of DNA topoisomerase (*topoisomerase II*) forms a covalent linkage to both strands of the DNA helix at the same time, making a transient

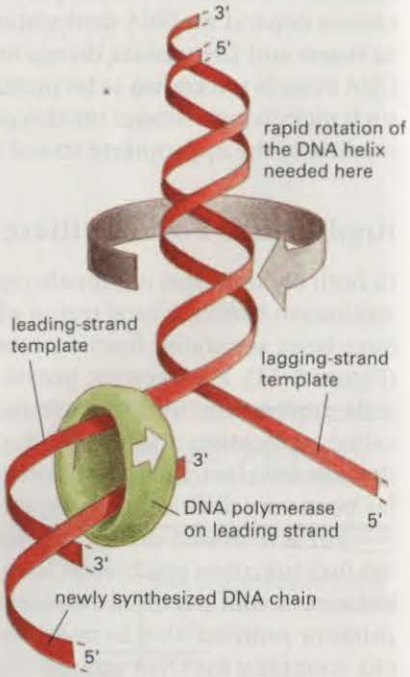


Figure 6–53 The "winding problem" that arises during DNA replication. For a bacterial replication fork moving at 500 nucleotides per second, the parental DNA helix ahead of the fork must rotate at 50 revolutions per second.

Sarepta Exhibit 1040, page 95

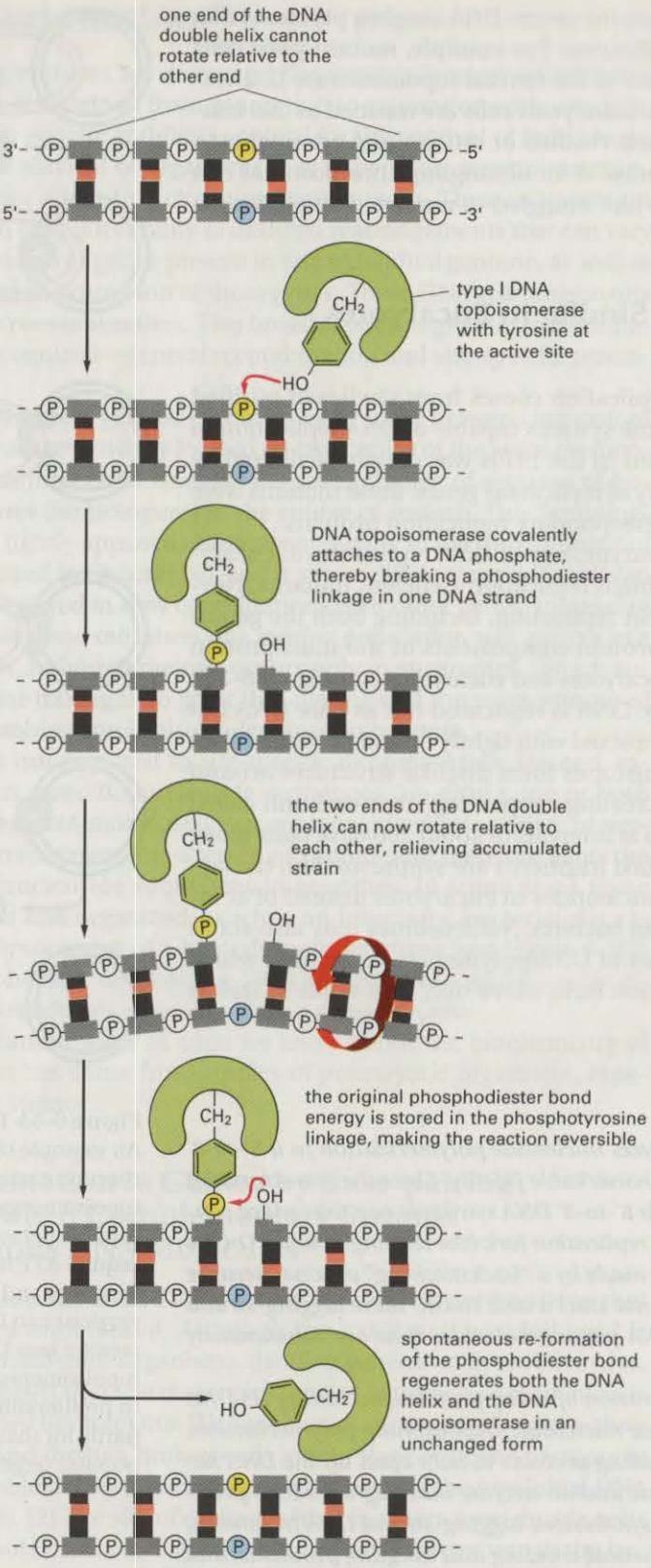


Figure 6–54 The reversible nicking reaction catalyzed by a eucaryotic DNA topoisomerase I enzyme. As indicated, these enzymes form a transient covalent bond with DNA so as to allow free rotation about the covalent bonds linked to the *blue* phosphate.

double-strand break in the helix. These enzymes are activated by sites on chromosomes where two double helices cross over each other. When the topoisomerase binds to such a crossing site, it (1) breaks one double helix reversibly to create a DNA "gate," (2) causes the second, nearby double helix to pass through this break, and (3) reseals the break and dissociates from the DNA. In this way type II DNA topoisomerases can efficiently separate two interlocked DNA circles

Sarepta Exhibit 1040, page 96

(Figure 6–55). The same reaction prevents the severe DNA tangling problems that would otherwise arise during DNA replication. For example, mutant yeast cells have been isolated that produce, in place of the normal topoisomerase II, a version that is inactive at 37°C. When the mutant yeast cells are warmed to this temperature, their chromosomes remain intertwined at mitosis and are unable to separate. The usefulness of topoisomerase II for untangling chromosomes can readily be appreciated by anyone who has struggled to remove a tangle from a fishing line without the aid of scissors.

DNA Replication Is Basically Similar in Eucaryotes and Procaryotes [37]

Much of what we know about DNA replication comes from studies of purified bacterial and bacteriophage multienzyme systems capable of DNA replication *in vitro*. The development of these systems in the 1970s was greatly facilitated by the prior isolation of mutants in a variety of replication genes; these mutants were exploited to identify and purify the corresponding replication proteins.

Less is known about the detailed enzymology of DNA replication in eucaryotes, largely because it is difficult to obtain replication-deficient mutants. Nevertheless, the basic mechanisms of DNA replication, including both the geometry of the replication fork and the protein components of the multiprotein replication machine, are similar for procaryotes and eucaryotes (see Figure 8–35). The major difference is that eucaryotic DNA is replicated not as bare DNA but as *chromatin*, in which the DNA is complexed with tightly bound proteins called *histones*. As described in Chapter 8, histones form disclike structures around which the eucaryotic DNA is wound, creating a repeating structural unit called a *nucleosome*. Nucleosomes are spaced at intervals of about 200 base pairs along the DNA, which may be why new Okazaki fragments are synthesized on the lagging strand at intervals of 100 to 200 nucleotides in eucaryotes instead of at intervals of 1000 to 2000 nucleotides as in bacteria. Nucleosomes may also act as barriers that slow down the movement of DNA polymerase molecules, which could explain why eucaryotic replication forks move only one-tenth as fast as bacterial replication forks.

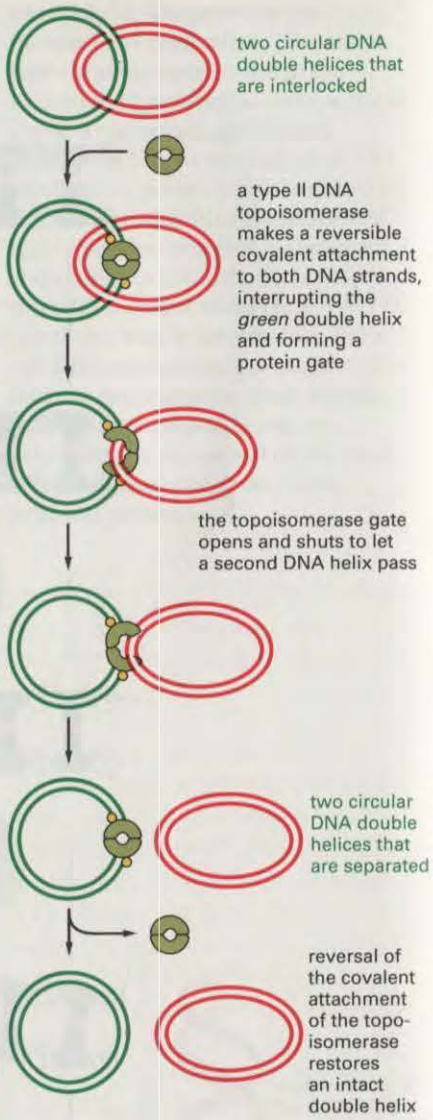


Figure 6–55 DNA topoisomerase II. An example of a DNA-helix-passing reaction catalyzed by a type II DNA topoisomerase. Unlike type I topoisomerases, these enzymes require ATP hydrolysis for their function, and some of the bacterial versions can introduce superhelical tension into DNA (see p. 438). Type II topoisomerases are largely confined to proliferating cells in eucaryotes; partly for that reason, they have been popular targets for anticancer drugs.

Summary

A self-correcting DNA polymerase catalyzes nucleotide polymerization in a 5′-to-3′ direction, copying a DNA template with remarkable fidelity. Since the two strands of a DNA double helix are antiparallel, this 5′-to-3′ DNA synthesis can take place continuously on only one of the strands at a replication fork (the leading strand). On the lagging strand short DNA fragments are made by a "backstitching" process. Because the self-correcting DNA polymerase cannot start a new chain, these lagging-strand DNA fragments are primed by short RNA primer molecules that are subsequently erased and replaced with DNA.

DNA replication requires the cooperation of many proteins, including (1) DNA polymerase and DNA primase to catalyze nucleoside triphosphate polymerization, (2) DNA helicases and single-strand binding proteins to help open up the DNA helix so that it can be copied, (3) DNA ligase and an enzyme that degrades RNA primers to seal together the discontinuously synthesized lagging-strand DNA fragments, (4) DNA topoisomerases to help relieve helical winding and tangling problems, and (5) initiator proteins that bind to specific DNA sequences at a replication origin and catalyze the formation of a replication fork at that site. At a replication origin a specialized protein-DNA structure is formed that subsequently loads a DNA helicase onto the DNA template; other proteins are then added to form the multienzyme "replication machine" that catalyzes DNA synthesis.

Sarepta Exhibit 1040, page 97

Genetic Recombination [38]

In the two preceding sections we discussed the mechanisms by which DNA sequences in cells are maintained from generation to generation with very little change. Although such genetic stability is crucial for the survival of individuals, in the longer term the survival of organisms may depend on genetic variation, through which they can adapt to a changing environment. Thus an important property of the DNA in cells is its ability to undergo rearrangements that can vary the particular combination of genes present in any individual genome, as well as the timing and the level of expression of these genes. These DNA rearrangements are caused by **genetic recombination.** Two broad classes of genetic recombination are commonly recognized—general recombination and site-specific recombination.

In *general recombination,* genetic exchange takes place between any pair of homologous DNA sequences, usually located on two copies of the same chromosome. One of the most important examples is the exchange of sections of homologous chromosomes (homologues) in the course of *meiosis.* This "crossing-over" occurs between tightly apposed chromosomes early in the development of eggs and sperm (discussed in Chapter 20), and it allows different versions (*alleles*) of the same gene to be tested in new combinations with other genes, increasing the chance that at least some members of a mating population will survive in a changing environment. Although meiosis occurs only in eucaryotes, the advantage of this type of gene mixing is so great that mating and the reassortment of genes by general recombination is also widespread in bacteria.

DNA homology is not required in *site-specific recombination.* Instead, exchange occurs at short, specific nucleotide sequences (on either one or both of the two participating DNA molecules) that are recognized by a variety of site-specific recombination enzymes. Site-specific recombination therefore alters the relative positions of nucleotide sequences in genomes. In some cases these changes are scheduled and organized, as when an integrated bacterial virus is induced to leave a chromosome of a bacterium under stress (see Figure 6–80); in others they are haphazard, as when the DNA sequence of a transposable element is inserted at a randomly selected site in a chromosome.

As for DNA replication, most of what we know about the biochemistry of genetic recombination has come from studies of procaryotic organisms, especially of *E. coli* and its viruses.

General Recombination Is Guided by Base-pairing Interactions Between Complementary Strands of Two Homologous DNA Molecules [39]

General recombination involves DNA strand-exchange intermediates that require some effort to understand. Although the exact pathway followed is likely to be different in different organisms, detailed genetic analyses of viruses, bacteria, and fungi suggest that the major outcome of general recombination is always the same. (1) Two homologous DNA molecules "cross over"; that is, their double helices break and the two broken ends join to their opposite partners to re-form two intact double helices, each composed of parts of the two initial DNA molecules (Figure 6–56). (2) The site of exchange (that is, where a *red* double helix is joined to a *green* double helix in Figure 6–56) can occur anywhere in the homologous nucleotide sequences of the two participating DNA molecules. (3) At the site of exchange, a strand of one DNA molecule becomes base-paired to a strand of the second DNA molecule to create a *staggered joint* (usually called a *heteroduplex joint*) between the two double helices (Figure 6–57). The heteroduplex region can be thousands of base pairs long; we shall explain later how it

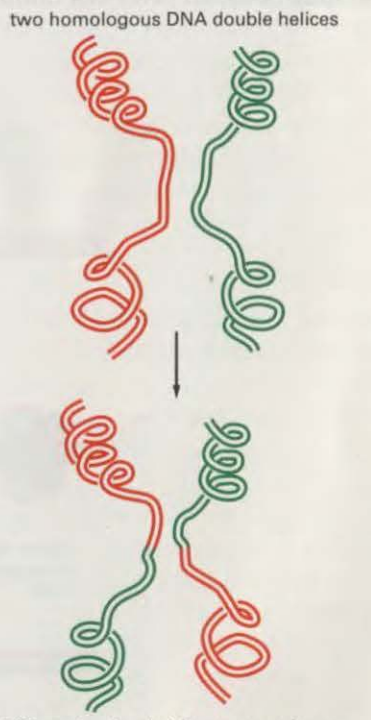


Figure 6–56 General recombination. The breaking and rejoining of two homologous DNA double helices creates two DNA molecules that have "crossed over."

Sarepta Exhibit 1040, page 98

forms. (4) No nucleotide sequences are altered at the site of exchange; the cleavage and rejoining events occur so precisely that not a single nucleotide is lost or gained. Despite this precision, general recombination creates DNA molecules of novel sequence: the heteroduplex joint can contain a small number of mismatched base pairs, and, more important, the two DNAs that cross over are usually not exactly the same on either side of the joint.

The mechanism of general recombination ensures that two regions of DNA double helix undergo an exchange reaction only if they have extensive sequence homology. The formation of a heteroduplex joint requires that such homology be present because it involves a long region of complementary base-pairing between a strand from one of the two original double helices and a complementary strand from the other. But how does this heteroduplex joint arise, and how do the two homologous regions of DNA at the site of crossing-over recognize each other? As we shall see, recognition takes place by means of a direct base-pairing interaction. The formation of base pairs between complementary strands from the two DNA molecules then guides the general recombination process, allowing it to occur only between long regions of matching DNA sequence.

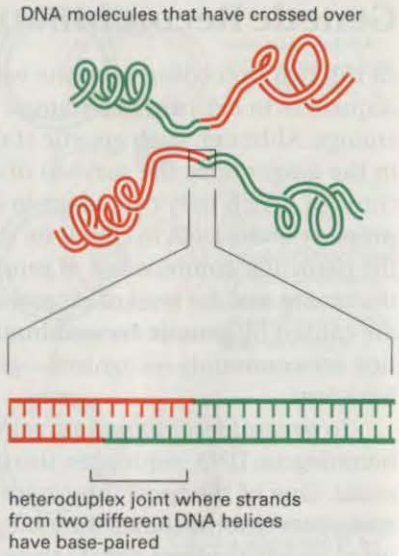


Figure 6–57 A heteroduplex joint. This structure unites two DNA molecules where they have crossed over. Such a joint is often thousands of nucleotides long.

General Recombination Can Be Initiated at a Nick in One Strand of a DNA Double Helix [40]

Each of the two strands in a DNA molecule is helically wound around the other. As a result, extensive base-pair interactions can occur between two homologous DNA double helices only if a nick is first made in a strand of one of them, freeing that strand for the unwinding and rewinding events required to form a heteroduplex with another DNA molecule. For the same reason, any *exchange* of strands between two DNA double helices requires at least two nicks, one in a

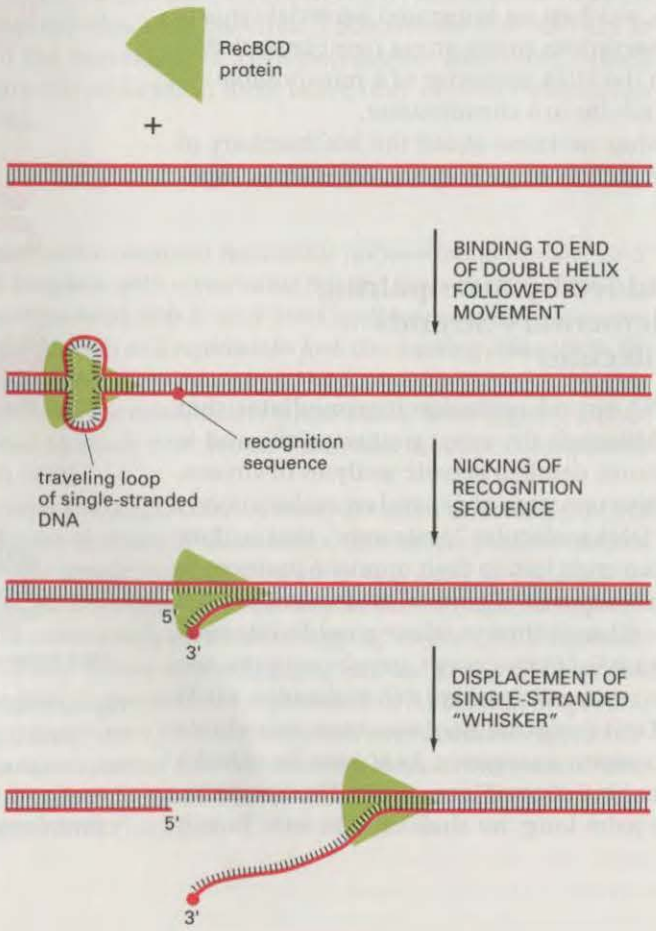


Figure 6–58 One way to start a recombination event. The RecBCD protein is an enzyme required for general genetic recombination in *E. coli*. The protein enters the DNA from one end of the double helix and then uses energy derived from the hydrolysis of bound ATP molecules to propel itself in one direction along the DNA at a rate of about 300 nucleotides per second. A special recognition site (a DNA sequence of eight nucleotides scattered throughout the *E. coli* chromosome) is cut in the traveling loop of DNA created by the RecBCD protein, and thereafter a single-stranded whisker is displaced from the helix, as shown. This whisker is thought to initiate genetic recombination by pairing with a homologous helix, as in Figure 6–59.

Sarepta Exhibit 1040, page 99

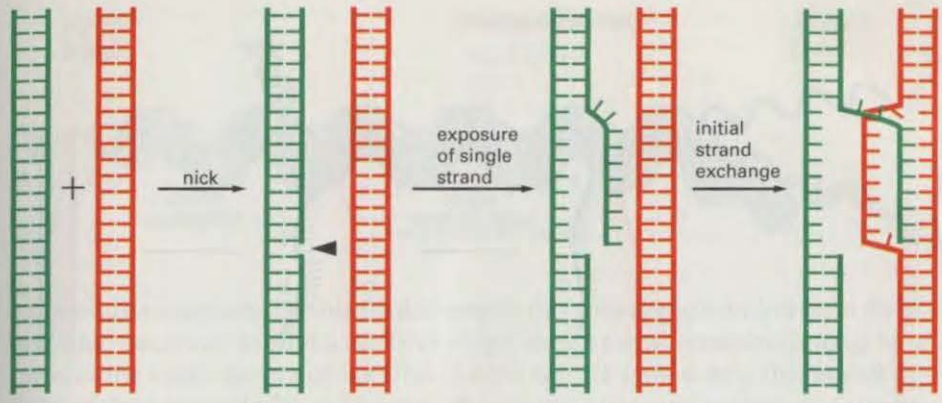


Figure 6–59 The initial strand exchange in general recombination. A nick in a single DNA strand frees the strand, which then invades a homologous DNA double helix to form a short pairing region with one of the strands in the second helix. Only two DNA molecules that are complementary in nucleotide sequence can base-pair in this way and thereby initiate a general recombination event. All of the steps shown here can be catalyzed by known enzymes (see Figures 6–58 and 6–62).

strand of each interacting double helix. Finally, to produce the heteroduplex joint illustrated in Figure 6–57, each of the four strands present must be cut to allow each to be joined to a different partner. In general recombination, these nicking and resealing events are coordinated so that they occur only when two DNA helices share an extensive region of matching DNA sequence.

There is evidence from a number of sources that a single nick in only one strand of a DNA molecule is sufficient to initiate general recombination. Chemical agents or types of irradiation that introduce single strand nicks, for example, will trigger a genetic recombination event. Moreover, one of the special proteins required for general recombination in *E. coli*—the *RecBCD protein*—has been shown to make single strand nicks in DNA molecules. The RecBCD protein is also a DNA helicase, hydrolyzing ATP and traveling along a DNA helix transiently exposing its strands. By combining its nuclease and helicase activities, the RecBCD protein will create a single-stranded "whisker" on the DNA double helix (Figure 6–58). Figure 6–59 shows how such a whisker could initiate a base-pairing interaction between two complementary stretches of DNA double helix.

DNA Hybridization Reactions Provide a Simple Model for the Base-pairing Step in General Recombination [41]

In its simplest form, the type of base-pairing interaction central to general recombination can be mimicked in a test tube by allowing a DNA double helix to reform from its separated single strands. This process, called **DNA renaturation** or **hybridization,** occurs when a rare random collision juxtaposes complementary nucleotide sequences on two matching DNA single strands, allowing the formation of a short stretch of double helix between them. This relatively slow *helix nucleation* step is followed by a very rapid "zippering" step as the region of double helix is extended to maximize the number of base-pairing interactions (Figure 6–60).

Formation of a new double helix in this way requires that the annealing strands be in an open, unfolded conformation. For this reason *in vitro* hybridization reactions are carried out at high temperature or in the presence of an organic solvent such as formamide; these conditions "melt out" the short hairpin helices formed where base-pairing interactions occur within a single strand that folds back on itself. Bacterial cells could not survive such harsh conditions and instead use a single-strand binding protein, the *SSB protein,* to open their helices. This protein is essential for DNA replication as well as for general recombination in *E. coli;* it binds tightly and cooperatively to the sugar-phosphate backbone of all single-stranded regions of DNA, holding them in an extended conformation with their bases exposed (see Figure 6–46). In this extended conformation a DNA single strand can base-pair efficiently with either a nucleoside triphosphate molecule (in DNA replication) or a complementary section of another DNA

Sarepta Exhibit 1040, page 100

forms. (4) No nucleotide sequences are altered at the site of exchange; the cleavage and rejoining events occur so precisely that not a single nucleotide is lost or gained. Despite this precision, general recombination creates DNA molecules of novel sequence: the heteroduplex joint can contain a small number of mismatched base pairs, and, more important, the two DNAs that cross over are usually not exactly the same on either side of the joint.

The mechanism of general recombination ensures that two regions of DNA double helix undergo an exchange reaction only if they have extensive sequence homology. The formation of a heteroduplex joint requires that such homology be present because it involves a long region of complementary base-pairing between a strand from one of the two original double helices and a complementary strand from the other. But how does this heteroduplex joint arise, and how do the two homologous regions of DNA at the site of crossing-over recognize each other? As we shall see, recognition takes place by means of a direct base-pairing interaction. The formation of base pairs between complementary strands from the two DNA molecules then guides the general recombination process, allowing it to occur only between long regions of matching DNA sequence.

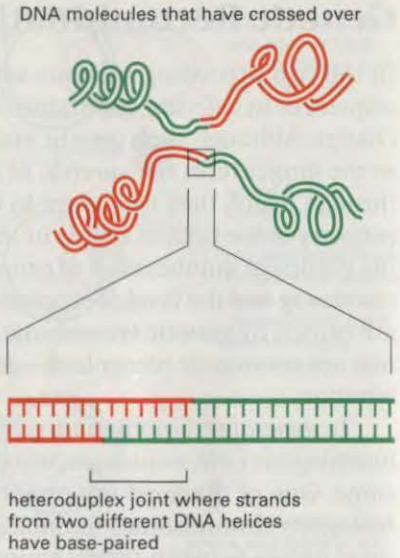


Figure 6–57 A heteroduplex joint. This structure unites two DNA molecules where they have crossed over. Such a joint is often thousands of nucleotides long.

General Recombination Can Be Initiated at a Nick in One Strand of a DNA Double Helix [40]

Each of the two strands in a DNA molecule is helically wound around the other. As a result, extensive base-pair interactions can occur between two homologous DNA double helices only if a nick is first made in a strand of one of them, freeing that strand for the unwinding and rewinding events required to form a heteroduplex with another DNA molecule. For the same reason, any *exchange* of strands between two DNA double helices requires at least two nicks, one in a

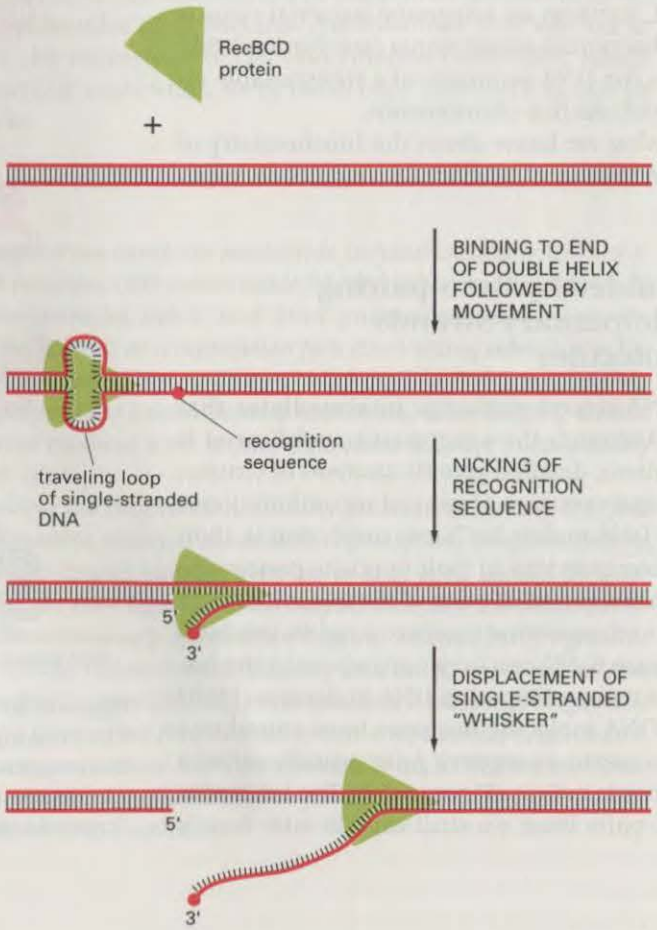


Figure 6–58 One way to start a recombination event. The RecBCD protein is an enzyme required for general genetic recombination in *E. coli*. The protein enters the DNA from one end of the double helix and then uses energy derived from the hydrolysis of bound ATP molecules to propel itself in one direction along the DNA at a rate of about 300 nucleotides per second. A special recognition site (a DNA sequence of eight nucleotides scattered throughout the *E. coli* chromosome) is cut in the traveling loop of DNA created by the RecBCD protein, and thereafter a single-stranded whisker is displaced from the helix, as shown. This whisker is thought to initiate genetic recombination by pairing with a homologous helix, as in Figure 6–59.

Sarepta Exhibit 1040, page 101

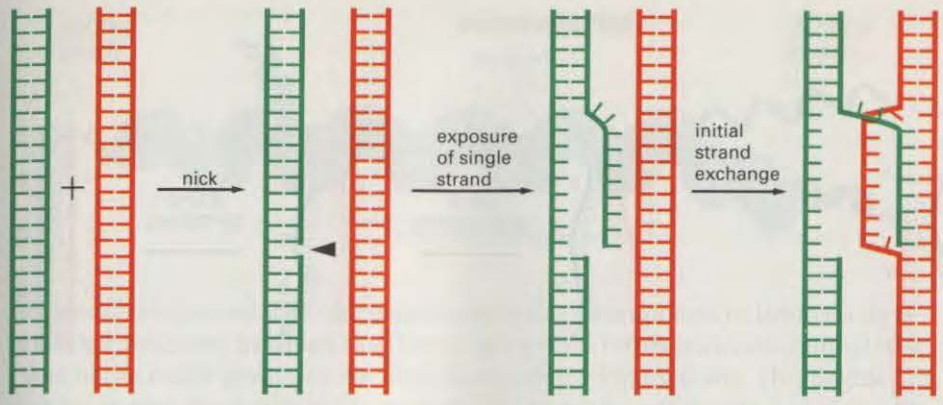


Figure 6–59 The initial strand exchange in general recombination. A nick in a single DNA strand frees the strand, which then invades a homologous DNA double helix to form a short pairing region with one of the strands in the second helix. Only two DNA molecules that are complementary in nucleotide sequence can base-pair in this way and thereby initiate a general recombination event. All of the steps shown here can be catalyzed by known enzymes (see Figures 6–58 and 6–62).

strand of each interacting double helix. Finally, to produce the heteroduplex joint illustrated in Figure 6–57, each of the four strands present must be cut to allow each to be joined to a different partner. In general recombination, these nicking and resealing events are coordinated so that they occur only when two DNA helices share an extensive region of matching DNA sequence.

There is evidence from a number of sources that a single nick in only one strand of a DNA molecule is sufficient to initiate general recombination. Chemical agents or types of irradiation that introduce single strand nicks, for example, will trigger a genetic recombination event. Moreover, one of the special proteins required for general recombination in *E. coli*—the *RecBCD protein*—has been shown to make single strand nicks in DNA molecules. The RecBCD protein is also a DNA helicase, hydrolyzing ATP and traveling along a DNA helix transiently exposing its strands. By combining its nuclease and helicase activities, the RecBCD protein will create a single-stranded "whisker" on the DNA double helix (Figure 6–58). Figure 6–59 shows how such a whisker could initiate a base-pairing interaction between two complementary stretches of DNA double helix.

DNA Hybridization Reactions Provide a Simple Model for the Base-pairing Step in General Recombination [41]

In its simplest form, the type of base-pairing interaction central to general recombination can be mimicked in a test tube by allowing a DNA double helix to reform from its separated single strands. This process, called **DNA renaturation** or **hybridization,** occurs when a rare random collision juxtaposes complementary nucleotide sequences on two matching DNA single strands, allowing the formation of a short stretch of double helix between them. This relatively slow *helix nucleation* step is followed by a very rapid "zippering" step as the region of double helix is extended to maximize the number of base-pairing interactions (Figure 6–60).

Formation of a new double helix in this way requires that the annealing strands be in an open, unfolded conformation. For this reason *in vitro* hybridization reactions are carried out at high temperature or in the presence of an organic solvent such as formamide; these conditions "melt out" the short hairpin helices formed where base-pairing interactions occur within a single strand that folds back on itself. Bacterial cells could not survive such harsh conditions and instead use a single-strand binding protein, the *SSB protein,* to open their helices. This protein is essential for DNA replication as well as for general recombination in *E. coli;* it binds tightly and cooperatively to the sugar-phosphate backbone of all single-stranded regions of DNA, holding them in an extended conformation with their bases exposed (see Figure 6–46). In this extended conformation a DNA single strand can base-pair efficiently with either a nucleoside triphosphate molecule (in DNA replication) or a complementary section of another DNA

Sarepta Exhibit 1040, page 102

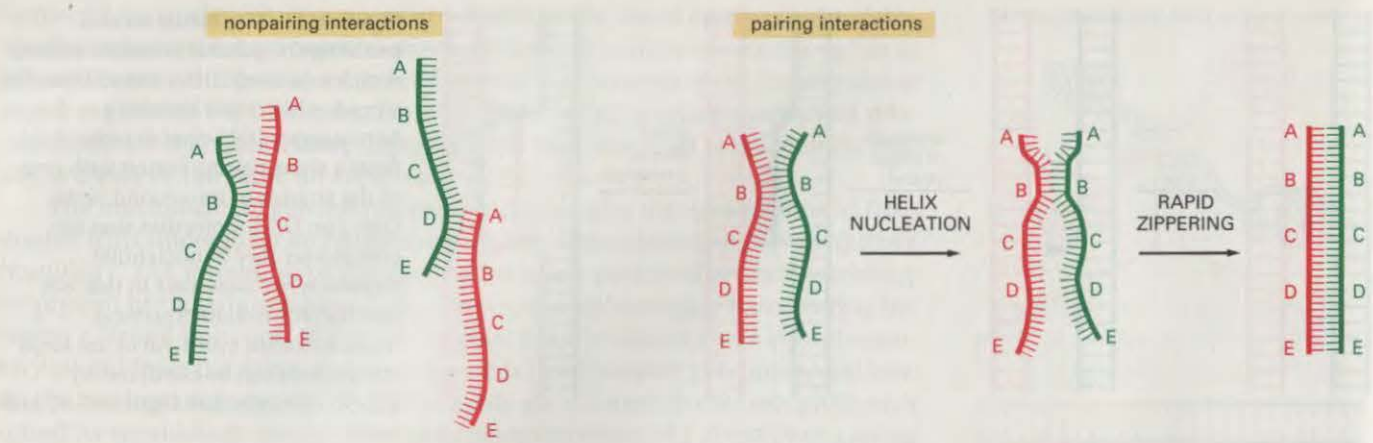


Figure 6–60 DNA hybridization. DNA double helices re-form from their separated strands in a reaction that depends on the random collision of two complementary strands (see p. 300). Most such collisions are not productive, as shown at the left, but a few result in a short region where complementary base pairs have formed (helix nucleation). A rapid zippering then leads to the formation of a complete double helix. A DNA strand can use this trial-and-error process to find its complementary partner in the midst of millions of nonmatching DNA strands. Trial-and-error recognition of a complementary partner DNA sequence appears to initiate all general recombination events.

single strand (in genetic recombination). When hybridization reactions are carried out *in vitro* under conditions that mimic those inside a cell, the SSB protein speeds up the rate of DNA helix nucleation and thereby the overall rate of strand annealing by a factor of more than 1000.

The RecA Protein Enables a DNA Single Strand to Pair with a Homologous Region of DNA Double Helix in *E. coli* [42]

General recombination is more complex than the simple hybridization reactions just described. In the course of general recombination, a single DNA strand from one DNA double helix must invade another double helix (see Figure 6–59). In *E. coli* this requires the **RecA protein,** produced by the *recA* gene, which was identified in 1965 as being essential for recombination between chromosomes. Long sought by biochemists, this elusive gene product was finally purified to homogeneity in 1976, a feat that allowed its detailed characterization (Figure 6–61). Like a single-strand binding (SSB) protein, the RecA protein binds tightly and in large cooperative clusters to single-stranded DNA to form a nucleoprotein filament. This filament has several distinctive properties. The RecA protein has more than one DNA-binding site, for example, and it can therefore hold a single strand and a double helix together. These sites allow the RecA protein to catalyze a multistep reaction (called **synapsis**) between a DNA double helix and a homologous region of single-stranded DNA. The crucial step in synapsis occurs when a region of homology is identified by an initial base-pairing between complementary



Figure 6–61 The structure of the RecA protein. A string of three RecA monomers is shown, with the position of each ATP in *red*. The *white spheres* show the putative position of the single-strand DNA in the filament, with three nucleotides (each shown as a sphere) bound per monomer. (From R.M. Story, I.T. Weber, and T.A. Steitz, *Nature* 256:318–325, 1992. © 1992 Macmillan Magazines Ltd.)

Sarepta Exhibit 1040, page 103

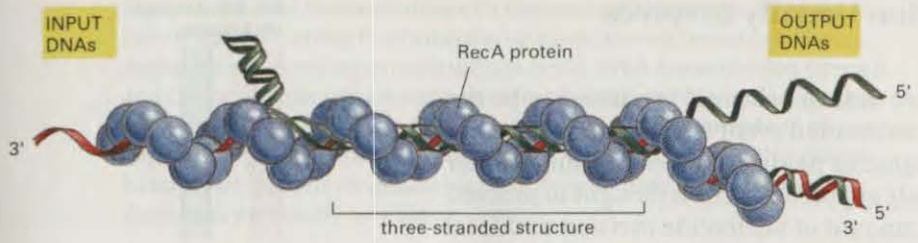


Figure 6–62 DNA synapsis catalyzed by the RecA protein. *In vitro* experiments show that several types of complexes are formed between a DNA single strand covered with RecA protein (*red*) and a DNA double helix (*green*). First a non-base-paired complex is formed, which is converted to a three-stranded structure as soon as a region of homologous sequence is found. This complex is presumably unstable because it involves an unusual form of DNA, and it spins out a DNA heteroduplex (one strand *green* and the other strand *red*) plus a displaced single strand from the original helix (*green*); thus the structure shown in this diagram migrates to the left, reeling in the "input DNAs" while producing the "output DNAs." The net result is a DNA strand exchange identical to that diagrammed earlier in Figure 6–59. (Adapted from S.C. West, *Annu. Rev. Biochem.* 61:603–640, 1992. © Annual Reviews Inc.)

nucleotide sequences. The nucleation step in this case appears to involve a three-stranded structure, in which the DNA single strand forms nonconventional base pairs in the major groove of the DNA double helix (Figure 6–62). This begins the pairing shown previously in Figure 6–59 and so initiates the exchange of strands between two recombining DNA double helices. Studies *in vitro* suggest that the *E. coli* SSB protein cooperates with the RecA protein to facilitate these reactions.

Once synapsis has occurred, a short heteroduplex region where the strands from two different DNA molecules have begun to pair is enlarged through *protein-directed branch migration*, which can also be catalyzed by the RecA protein. **Branch migration** can take place at any point where two single DNA strands with the same sequence are attempting to pair with the same complementary strand; an unpaired region of one of the single strands will displace a paired region of the other single strand, moving the branch point without changing the total number of DNA base pairs. Spontaneous branch migration proceeds equally in both directions, and so it makes little progress and is unlikely to complete recombination efficiently (Figure 6–63A). Because the RecA protein catalyzes unidirectional branch migration, it readily produces a region of heteroduplex that is thousands of base pairs long (Figure 6–63B).

The catalysis of branch migration depends on a further property of the RecA protein. In addition to having two DNA-binding sites, the RecA protein is a DNA-dependent ATPase, with an additional site for binding and hydrolyzing ATP. The protein associates much more tightly with DNA when it has ATP bound than when it has ADP bound. Moreover, new RecA molecules with ATP bound are preferentially added at one end of the RecA protein filament, and the ATP is then hydrolyzed to ADP. The RecA protein filaments that form on DNA may therefore share many of the dynamic assembly properties displayed by the cytoskeletal filaments formed from actin or tubulin (discussed in Chapter 16); an ability of the protein to "treadmill" unidirectionally along a DNA strand, for example, could drive the branch migration reaction shown in Figure 6–63B.

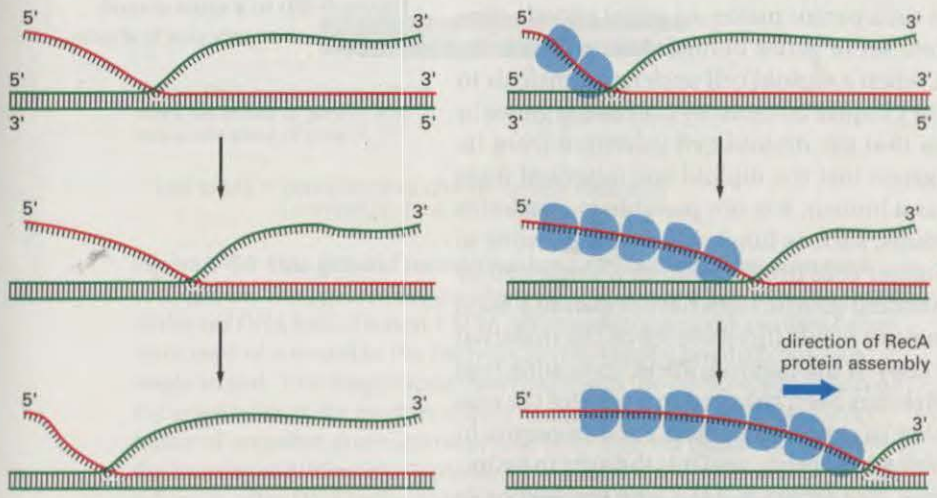


Figure 6–63 Two types of DNA branch migration observed in experiments *in vitro*. (A) Spontaneous branch migration is a back-and-forth, random-walk type of process, and it therefore makes little progress over long distances. (B) RecA-protein-directed branch migration proceeds at a uniform rate in one direction, and it may be driven by the polarized assembly of the RecA protein filament on a DNA single strand, which occurs in the direction indicated. In addition, special DNA helicases that catalyze protein-directed branch migration even more efficiently are involved in recombination.

Sarepta Exhibit 1040, page 104

General Genetic Recombination Usually Involves a Cross-Strand Exchange [43]

Exchanging a single strand between two double helices is presumed to be the slow and difficult step in a general recombination event (see Figure 6–59). After this initial exchange, extending the region of pairing and establishing further strand exchanges between the two closely apposed helices is thought to proceed rapidly. During these events a limited amount of nucleotide excision and local DNA resynthesis often occurs, resembling some of the events in DNA repair. Because of the large number of possibilities, different organisms are likely to follow different pathways at this stage. In most cases, however, an important intermediate structure, the **cross-strand exchange,** will be formed by the two participating DNA helices. One of the simplest ways in which this structure can form is shown in Figure 6–64.

In the cross-strand exchange (also called a *Holliday junction*) the two homologous DNA helices that initially paired are held together by mutual exchange of two of the four strands present, one originating from each of the helices. No disruption of base-pairing is necessary to maintain this structure, which has two important properties—(1) the point of exchange between the two homologous DNA double helices (where the two strands cross in Figure 6–64) can migrate rapidly back and forth along the helices by a double branch migration; (2) the cross-strand exchange contains two pairs of strands: one pair of crossing strands and one pair of noncrossing strands. The structure can *isomerize,* however, by undergoing a series of rotational movements, so that the two original noncrossing strands become crossing strands and vice versa (Figure 6–65).

In order to regenerate two separate DNA helices and thus terminate the pairing process, the two crossing strands must be cut. If the crossing strands are cut *before* isomerization, the two original DNA helices separate from each other nearly unaltered, with only a very short piece of single-stranded DNA exchanged. If the crossing strands are cut *after* isomerization, however, one section of each original DNA helix is joined to a section of the other DNA helix; in other words, the two DNA helices have crossed over (see Figure 6–65).

The isomerization of the cross-strand exchange should occur spontaneously at some rate, but it may also be enzymatically driven or otherwise regulated by cells. Some kind of control probably operates during meiosis, when the two DNA double helices that pair are constrained in an elaborate structure called the *synaptonemal complex* (discussed in Chapter 20).

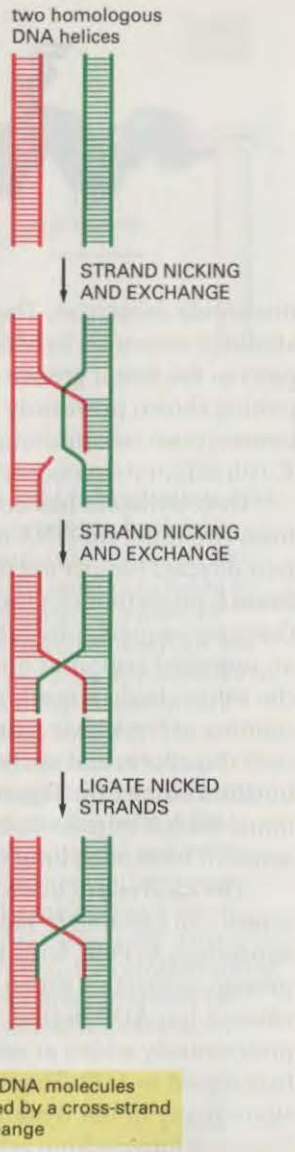


Figure 6–64 The formation of a cross-strand exchange. There are many possible pathways that can lead from a single-strand exchange (see Figure 6–59) to a cross-strand exchange, but only one is shown.

Gene Conversion Results from Combining General Recombination and Limited DNA Synthesis [44]

It is a fundamental law of genetics that each parent makes an equal genetic contribution to the offspring, one complete set of genes being inherited from the father and one from the mother. Thus, when a diploid cell undergoes meiosis to produce four haploid cells (discussed in Chapter 20), exactly half of the genes in these cells should be maternal (genes that the diploid cell inherited from its mother) and the other half paternal (genes that the diploid cell inherited from its father). In a complex animal, such as a human, it is not possible to check this prediction directly. But in other organisms, such as fungi, where it is possible to recover and analyze all four of the daughter cells produced from a single cell by meiosis, one finds cases in which the standard genetic rules have apparently been violated. Occasionally, for example, meiosis yields three copies of the maternal version of a gene (allele) and only one copy of the paternal allele, indicating that one of the two copies of the paternal allele has been changed to a copy of the maternal allele. This phenomenon is known as **gene conversion.** It often occurs in association with general genetic recombination events, and it is thought to be important in the evolution of certain genes (see Figure 8–74). Gene conversion is believed to be a straightforward consequence of the mechanisms of general recombination and DNA repair.

Sarepta Exhibit 1040, page 105

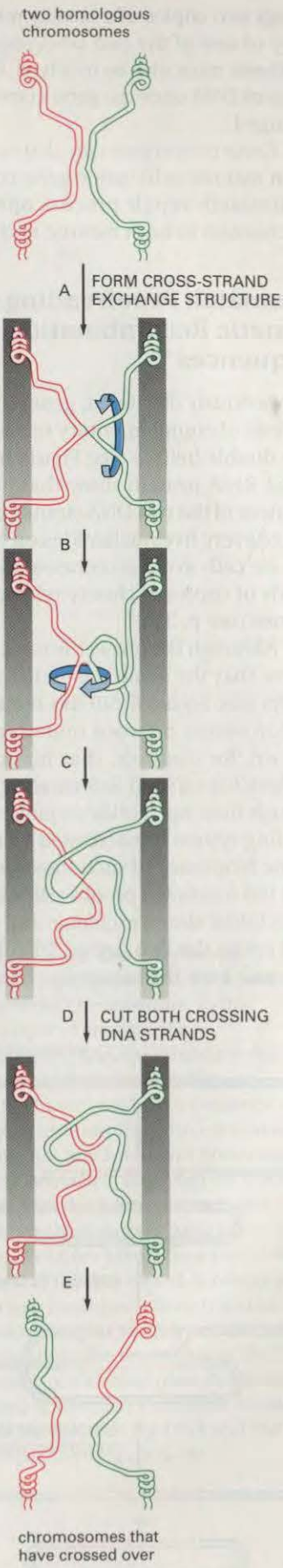


Figure 6–65 The isomerization of a cross-strand exchange. Without isomerization, cutting the two crossing strands would terminate the exchange and crossing over would not occur. With isomerization (steps B and C), cutting the two crossing strands creates two DNA molecules that have crossed over (*bottom*). Isomerization is therefore thought to be required for the breaking and rejoining of two homologous DNA double helices that result from general genetic recombination. Step A was illustrated previously (see Figure 6–64).

During meiosis heteroduplex joints are formed at the sites of crossing-over between homologous maternal and paternal chromosomes. If the maternal and paternal DNA sequences are slightly different, the heteroduplex joint may include some mismatched base pairs. The resulting mismatch in the double helix may then be corrected by the DNA repair machinery, which either can erase nucleotides on the paternal strand and replace them with nucleotides that match the maternal strand or vice versa. The consequence of this mismatch repair will be a gene conversion. Gene conversion can also take place by a number of other mechanisms, but they all require some type of general recombination event that

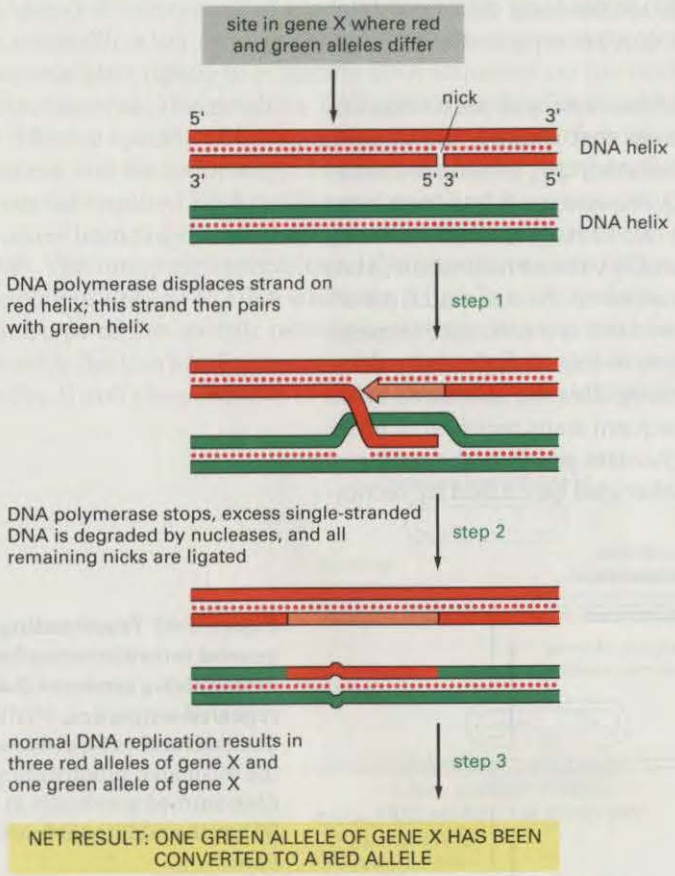


Figure 6–66 One general recombination pathway that can cause gene conversion. The process begins when a nick is formed in one of the strands in the *red* DNA helix. In step 1 DNA polymerase begins the synthesis of an extra copy of a strand in the *red* helix, displacing the original copy as a single strand. This single strand then pairs with the homologous region of the *green* helix in the manner shown in Figure 6–59. In step 2 the short region of unpaired *green* strand produced in step 1 is degraded, completing the transfer of nucleotide sequences. The result is normally seen in the next cell cycle, after DNA replication has separated the two nonmatching strands (step 3). As described in the text, the repair of mismatched base pairs in a heteroduplex joint also causes gene conversion.

Sarepta Exhibit 1040, page 106

brings two copies of a closely related DNA sequence together. Because an extra copy of one of the two DNA sequences is generated, a limited amount of DNA synthesis must also be involved. Genetic studies show that usually only small sections of DNA undergo gene conversion, and in many cases only part of a gene is changed.

Gene conversion can also occur in mitotic cells, but it does so more rarely. As in meiotic cells, some gene conversions in mitotic cells probably result from a mismatch repair process operating on heteroduplex DNA. Another likely mechanism in both meiotic and mitotic cells is illustrated in Figure 6–66.

Mismatch Proofreading Can Prevent Promiscuous Genetic Recombination Between Two Poorly Matched DNA Sequences [45]

As previously discussed, general recombination is triggered whenever two DNA strands of complementary sequence pair to form a heteroduplex joint between two double helices (see Figure 6–64). Experiments carried out *in vitro* with purified RecA protein show that pairing can occur efficiently even when the sequences of the two DNA strands do not match well—when, for example, only four out of every five nucleotides on average can form base pairs. How, then, do vertebrate cells avoid promiscuous general recombination between the many thousands of copies of closely related DNA sequences that are repeated in their genomes (see p. 395)?

Although the answer is not known, studies with bacteria and yeasts demonstrate that the same mismatch proofreading system that removes replication errors (see Figure 6–50) has the additional role of interrupting genetic recombination events between imperfectly matched DNA sequences. It has long been known, for example, that homologous genes in two closely related bacteria, *Escherichia coli* and *Salmonella typhimurium*, generally will not recombine, even though their nucleotide sequences are 80% identical; when the mismatch proofreading system is inactivated by mutation, however, there is a 1000-fold increase in the frequency of such interspecies recombination events. It is thought, then, that the mismatch proofreading system normally recognizes the mispaired bases in an initial strand exchange and prevents the subsequent steps required to break and rejoin the two paired DNA helices. This mechanism protects the bacterial genome from the sequence changes that would otherwise be caused by recom-

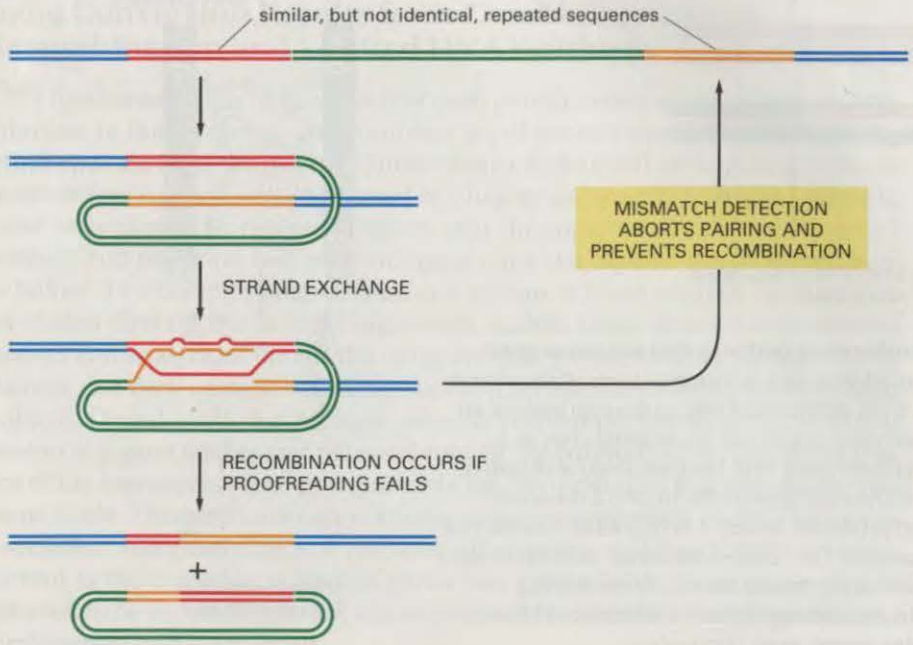


Figure 6–67 Proofreading prevents general recombination from destabilizing genomes that contain repeated sequences. Studies with bacterial and yeast cells suggest that the mismatch proofreading system diagrammed previously in Figure 6–50 has the additional function shown here.

Sarepta Exhibit 1040, page 107

bination with foreign DNA molecules that occasionally enter the cell. In vertebrate cells, which contain many closely related DNA sequences, the same type of proofreading is thought to help prevent promiscuous recombination events that would otherwise scramble the genome (Figure 6–67).

Site-specific Recombination Enzymes Move Special DNA Sequences into and out of Genomes [46]

Site-specific genetic recombination, unlike general recombination, is guided by a recombination enzyme that recognizes specific nucleotide sequences present on one or both of the recombining DNA molecules. Base-pairing between the recombining DNA molecules need not be involved, and even when it is, the heteroduplex joint that is formed is only a few base pairs long. By separating and joining double-stranded DNA molecules at specific sites, this type of recombination enables various types of mobile DNA sequences to move about within and between chromosomes.

Site-specific recombination was first discovered as the means by which a bacterial virus, bacteriophage *lambda,* moves its genome into and out of the *E. coli* chromosome. In its integrated state the virus is hidden in the bacterial chromosome and replicated as part of the host's DNA. When the virus enters a cell, a virus-encoded enzyme called *lambda integrase* is synthesized. This enzyme catalyzes a recombination process that begins when several molecules of the integrase protein bind tightly to a specific DNA sequence on the circular bacteriophage chromosome. The resulting DNA-protein complex can now bind to a related but different specific DNA sequence on the bacterial chromosome, bringing the bacterial and bacteriophage chromosomes close together. The integrase then catalyzes the required DNA cutting and resealing reactions, using a short region of sequence homology to form a tiny heteroduplex joint at the point of union (Figure 6–68). The integrase resembles a DNA topoisomerase in that it forms a reversible covalent linkage to DNA wherever it breaks a DNA chain.

The same type of site-specific recombination mechanism can also be carried out in *reverse* by the lambda bacteriophage, enabling it to exit from its integration site in the *E. coli* chromosome in order to multiply rapidly within the bac-

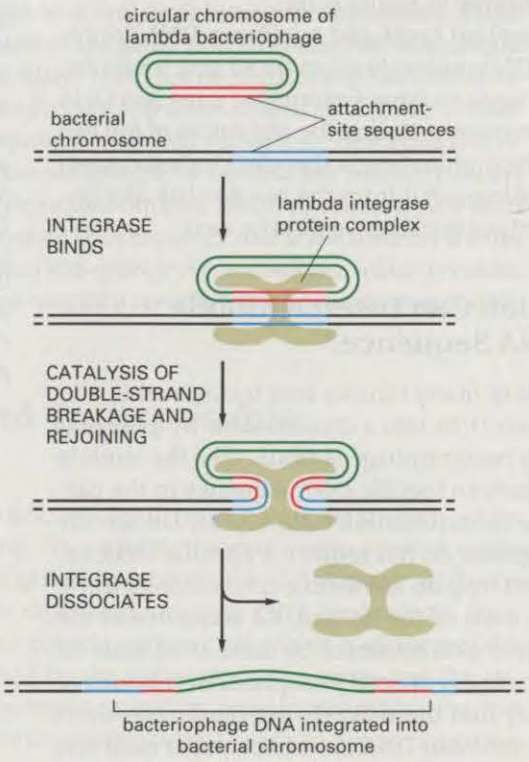


Figure 6–68 The insertion of bacteriophage lambda DNA into the bacterial chromosome. In this example of site-specific recombination, the lambda integrase enzyme binds to a specific "attachment site" DNA sequence on each chromosome, where it makes cuts that bracket a short homologous DNA sequence; the integrase thereby switches the partner strands and reseals them so as to form a heteroduplex joint 7 base pairs long. Each of the four strand-breaking and strand-joining reactions required resembles that made by a DNA topoisomerase, inasmuch as the energy of a cleaved phosphodiester bond is stored in a transient covalent linkage between the DNA and the enzyme (see Figure 6–64).

Sarepta Exhibit 1040, page 108

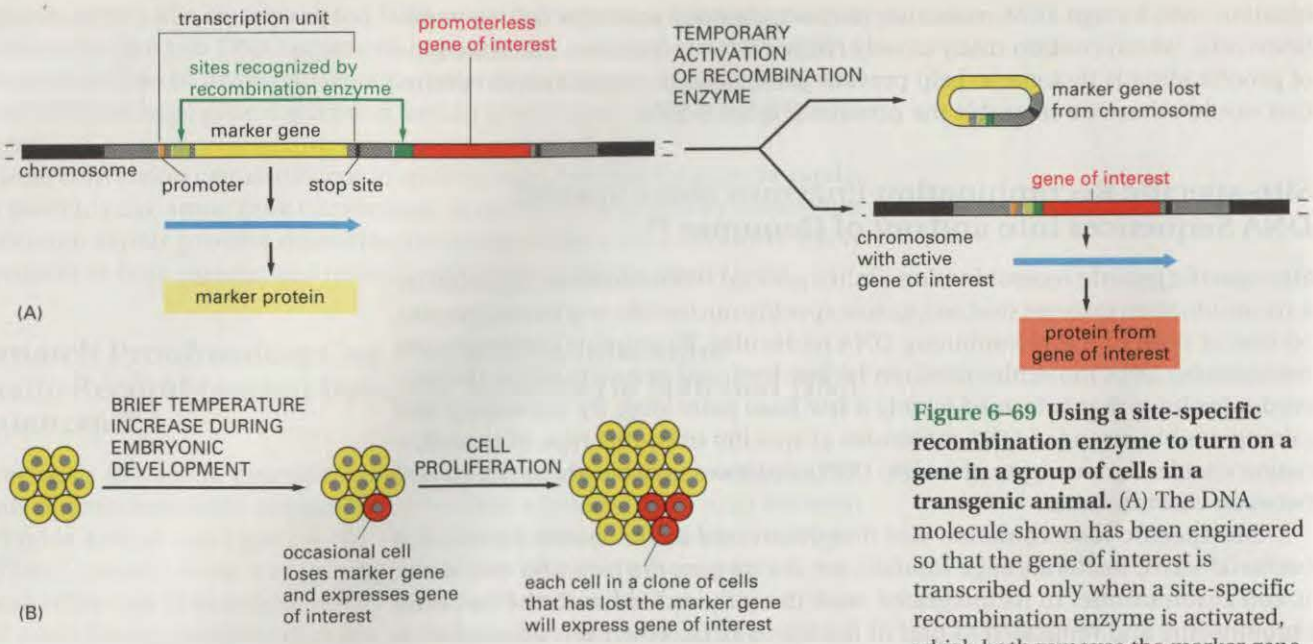


Figure 6–69 Using a site-specific recombination enzyme to turn on a gene in a group of cells in a transgenic animal. (A) The DNA molecule shown has been engineered so that the gene of interest is transcribed only when a site-specific recombination enzyme is activated, which both removes the marker gene and brings the promoter next to the gene of interest. The recombination enzyme is encoded by a second DNA molecule (not shown) that is engineered so that the enzyme is made only when the temperature is increased. Both DNA molecules are introduced into the chromosomes of the same transgenic animal. When the temperature of this animal is transiently increased, there is a brief burst of synthesis of the recombination enzyme, which causes a DNA rearrangement in an occasional cell such that the marker gene is removed and the gene of interest is simultaneously activated. (B) The strategy can be used to turn on a gene of interest permanently in small clones of cells in a developing animal. The clones can be identified by their loss of the marker gene product, which, for example, could cause a change in the pigmentation of the cells. This technique therefore allows one to study the effect of expressing any gene of interest in a group of cells in an intact animal.

terial cell. This excision reaction is catalyzed by a complex of the integrase enzyme with a second bacteriophage protein, which is produced by the virus only when its host cell is stressed. If the sites recognized by such a recombination enzyme are flipped, the DNA between them will be inverted rather than excised (see Figure 9–57).

Many other enzymes that catalyze site-specific recombination resemble lambda integrase in requiring a short region of identical DNA sequence on the two regions of DNA helix to be joined. Because of this requirement, each enzyme in this class is fastidious with respect to the DNA sequences that it recombines, and it can be expected to catalyze one particular DNA joining event that is useful to the virus, plasmid, transposable element, or cell that contains it. These enzymes can be exploited as tools in transgenic animals to study the influence of specific genes on cell behavior, as illustrated in Figure 6–69.

Site-specific recombination enzymes that break and rejoin two DNA double helices at specific sequences on each DNA molecule often do so in a reversible way: as for lambda bacteriophage, the same enzyme system that joins two DNA molecules can take them apart again, precisely restoring the sequences of the two original DNA molecules. This type of recombination is therefore called *conservative site-specific recombination* to distinguish it from the mechanistically distinct *transpositional site-specific recombination* that we discuss next.

Transpositional Recombination Can Insert a Mobile Genetic Element into Any DNA Sequence [47]

Many mobile DNA sequences, including many viruses and transposable elements, encode integrases that insert their DNA into a chromosome by a mechanism that is different from that used by bacteriophage lambda. Like the lambda integrase, each of these enzymes recognizes a specific DNA sequence in the particular mobile genetic element whose recombination it catalyzes. Unlike the lambda enzyme, however, these integrases do not require a specific DNA sequence in the "target" chromosome and they do not form a heteroduplex joint. Instead, they introduce cuts into both ends of the linear DNA sequence of the mobile genetic element and then catalyze a direct attack by these DNA ends on the target DNA molecule, breaking two closely spaced phosphodiester bonds in the target molecule. Because of the way that these breaks are made, two short single-stranded gaps are left in the recombinant DNA molecule, one at each end

Sarepta Exhibit 1040, page 109

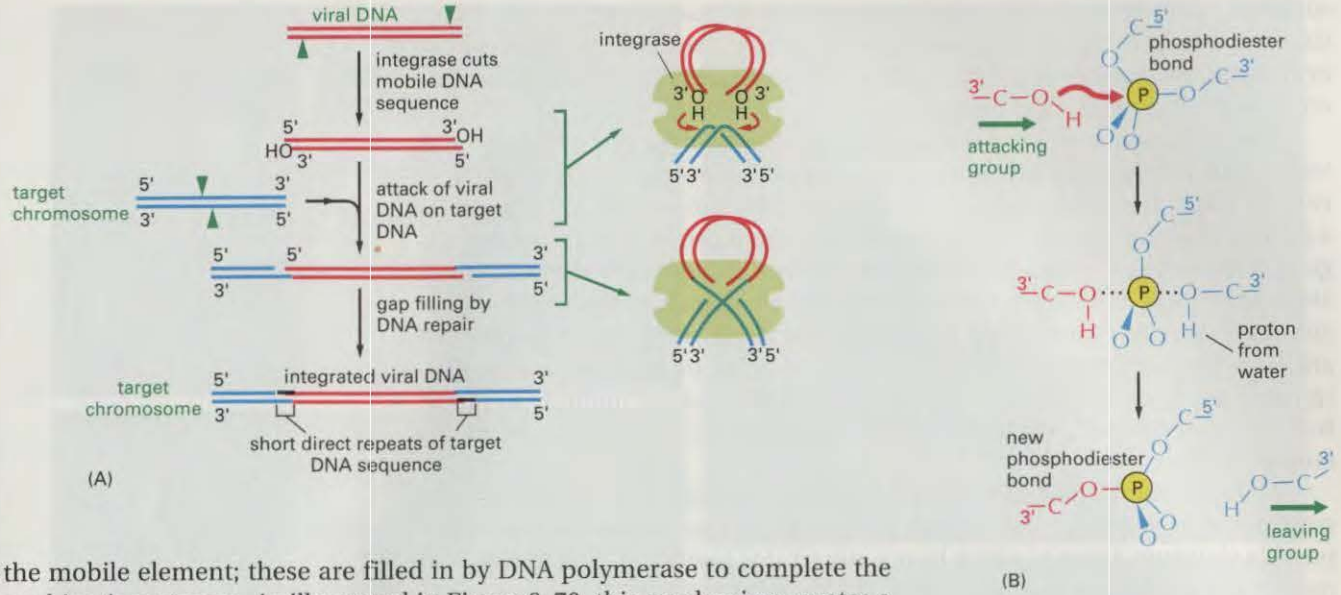


Figure 6–70 Transpositional site-specific recombination. (A) Outline of the strand-breaking and -rejoining events that lead to integration of the linear double-stranded DNA of a retrovirus (*red*) into an animal cell chromosome (*blue*). In an initial endonuclease step the integrase enzyme makes a cut in one strand at each end of the viral DNA sequence, exposing a protruding 3′-OH group. Each of these 3′-OH ends then directly attacks a phosphodiester bond on opposite strands of a randomly selected site on a target chromosome. This inserts the viral DNA sequence into the target chromosome, leaving short gaps on each side that are filled in by DNA repair processes. Because of the gap filling, this type of mechanism leaves short repeats of target DNA sequence [3 to 12 nucleotides in length (*black*), depending on the integrase enzyme] on either side of the integrated DNA segment. (B) An atomic-level view of the attack by one DNA chain end in (A) on a phosphodiester bond of the target DNA (*blue*). This mechanism resembles that used in RNA splicing, and is distinctly different from the topoisomerase-like activity of lambda integrase. (Adapted from K. Mizuuchi, *J. Biol. Chem.* 267:21273–21276, 1992.)

of the mobile element; these are filled in by DNA polymerase to complete the recombination process. As illustrated in Figure 6–70, this mechanism creates a short duplication of the adjacent target DNA sequence; such flanking duplications are the hallmark of a transpositional site-specific recombination event.

An integrase enzyme of this type was first purified in active form from bacteriophage Mu. Like the bacteriophage lambda integrase, it carries out all of its cutting and rejoining reactions without requiring an energy source (such as ATP). Very similar enzymes are present in organisms as diverse as bacteria, fruit flies, and humans—all of which contain mobile genetic elements, as we discuss next.

Summary

Genetic recombination mechanisms allow large sections of DNA double helix to move from one chromosome to another. There are two broad classes of recombination events. In general recombination the initial reactions rely on extensive base-pairing interactions between strands of the two DNA double helices that will recombine. As a result, general recombination occurs only between two homologous DNA molecules, and although it moves sections of DNA back and forth between chromosomes, it does not normally change the arrangement of the genes in a chromosome. Site-specific recombination, on the other hand, alters the relative positions of nucleotide sequences in chromosomes because the pairing reactions depend on a protein-mediated recognition of the two DNA sequences that will recombine, and extensive sequence homology is not required. Two site-specific recombination mechanisms are common: (1) conservative site-specific recombination, which produces a very short heteroduplex and therefore requires some DNA sequence that is the same on the two DNA molecules, and (2) transpositional site-specific recombination, which produces no heteroduplex and usually does not require a specific sequence on the target DNA.

Viruses, Plasmids, and Transposable Genetic Elements [48]

In our description of the basic genetic mechanisms, we have so far focused on their selective advantage for the cell. We saw that the short-term survival of the cell depends on the maintenance of genetic information by DNA repair, while the multiplication of the cell requires rapid and accurate DNA replication. On a longer time scale the appearance of genetic variants, on which evolution of the species depends, is greatly facilitated by the reassortment of genes and the occasional rearrangement of DNA sequences caused by genetic recombination. We shall now examine a group of genetic elements that seem to act as parasites,

Sarepta Exhibit 1040, page 110

subverting the genetic mechanisms of the cell for their own benefit. These genetic elements are interesting in their own right. In addition, because they must heavily exploit the metabolism of the host cell in order to multiply, they serve as powerful tools for investigating the normal cell machinery.

Many DNA sequences can replicate independently of the rest of the genome. Such sequences have widely different degrees of independence from their host cells. Of these, virus chromosomes are the most independent because they have a protein coat that allows them to move freely from cell to cell. To varying degrees, the *viruses* are closely related to *plasmids* and *transposable elements*, which are DNA sequences that lack a coat and are therefore more host-cell-dependent and confined to replicate within a single cell and its progeny. More primitive still are some DNA sequences that are suspected of being mobile because they are repeated many times in a cell's chromosome. They move or multiply so rarely, however, that it is not clear if they should be considered as separate genetic elements at all.

We begin our discussion with viruses, which are the best understood of the mobile genetic elements. Then we describe the properties of plasmids and transposable elements, some of which bear a remarkable resemblance to viruses and may in fact have been their ancestors. The many repetitive DNA sequences in vertebrate chromosomes are discussed in Chapter 8.

Viruses Are Mobile Genetic Elements [49]

Viruses were first described as disease-causing agents that can multiply only in cells and that by virtue of their tiny size pass through ultrafine filters that hold back even the smallest bacteria. Before the advent of the electron microscope, their nature was obscure, although it was suspected that they might be naked genes that had somehow acquired the ability to move from one cell to another. The use of ultracentrifuges in the 1930s made it possible to separate viruses from host cell components, and by the early 1940s the generalization emerged that all viruses contain nucleic acids. The idea that viruses and genes carry out similar functions was confirmed by studies on **bacteriophages,** which are bacterial viruses. In 1952 it was shown for the bacteriophage T4 that only the phage DNA, and not the phage protein, enters the bacterial host cell and initiates the replication events that lead to the production of several hundred progeny viruses in every infected cell.

These observations led to the notion of viruses as genetic elements enclosed by a protective coat that enables them to move from one cell to another. Virus multiplication per se is often lethal to the cells in which it occurs; in many cases the infected cell breaks open (*lyses*) and thereby allows the progeny viruses access to nearby cells. Many of the clinical manifestations of viral infection reflect this cytolytic effect of the virus. Both the cold sores formed by herpes simplex virus and the lesions caused by smallpox, for example, reflect the killing of the epithelial cells in a local area of the skin.

As we shall see, the type of nucleic acid in a virus, the structure of its coat, its mode of entry into the host cell, and its mechanism of replication once inside all vary from one type of virus to another.

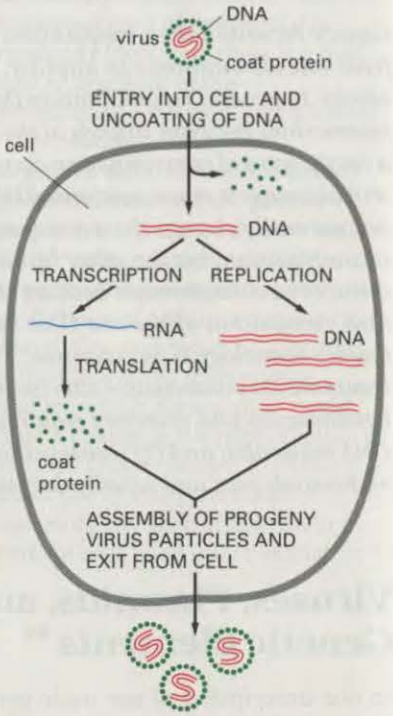
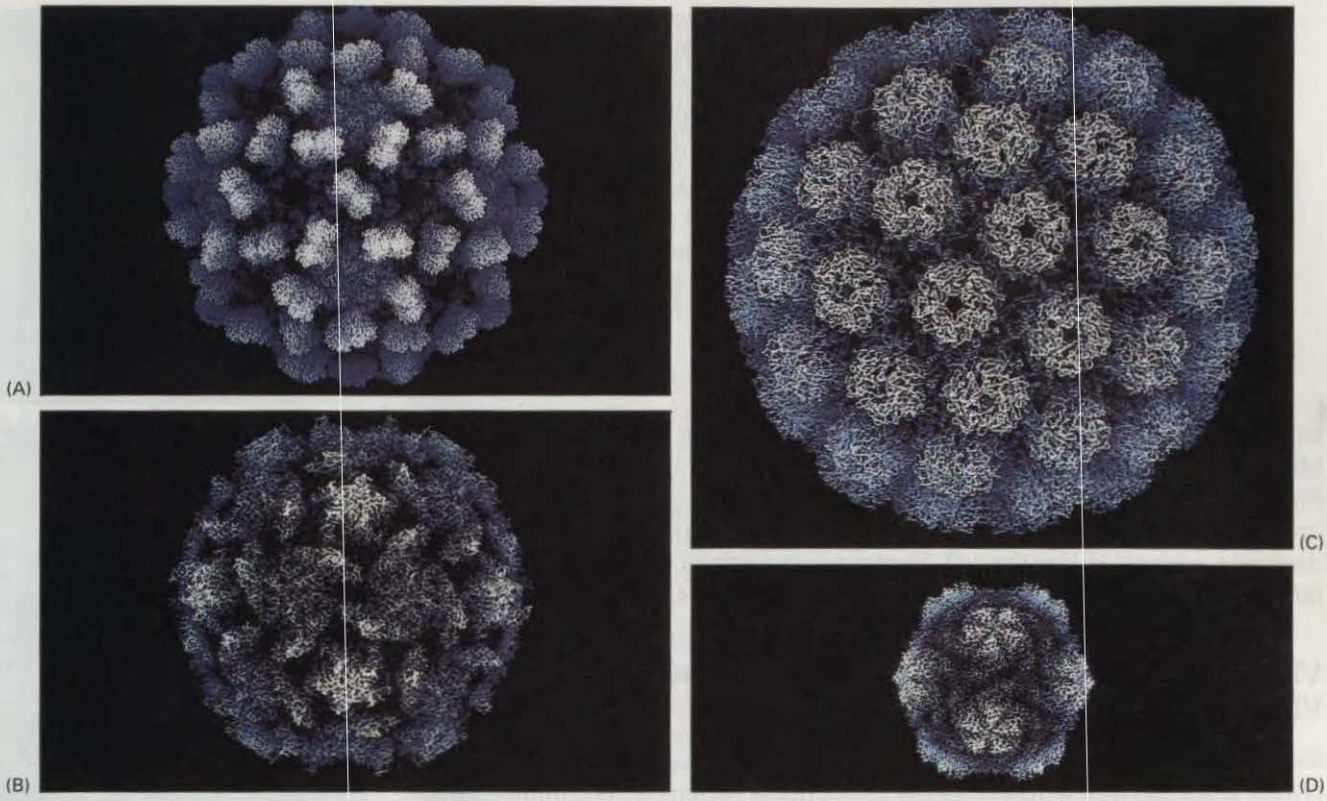


Figure 6–71 The simplest of all viral life cycles. The hypothetical virus shown consists of a small double-stranded DNA molecule that codes for only a single viral capsid protein. No known virus is this simple.

The Outer Coat of a Virus May Be a Protein Capsid or a Membrane Envelope [50]

Initially, it was thought that the outer coat of a virus might be constructed from a single type of protein molecule. Viral infections were believed to start with the dissociation of the viral chromosome (its nucleic acid) from its protein coat, followed by replication of the chromosome inside the host cell, to form many identical copies. After the synthesis of new copies of the virus-specific coat protein from virally encoded messenger RNA molecules, formation of the progeny virus

Sarepta Exhibit 1040, page 111



particles would occur by the spontaneous assembly of these coat protein molecules around the progeny viral chromosomes (Figure 6–71).

It is now known that these ideas vastly oversimplify the diversity of virus life cycles. The protein shell that surrounds the nucleic acid of most viruses (the **capsid**), for example, contains more than one type of polypeptide chain, often arranged in several layers (Figure 6–72). In many viruses, moreover, the protein

Figure 6–72 The capsids of some viruses, all shown at the same scale. (A) Tomato bushy stunt virus; (B) poliovirus; (C) simian virus 40 (SV40); (D) satellite tobacco necrosis virus. The structures of all of these capsids have been determined by x-ray crystallography and are known in atomic detail. (Courtesy of Robert Grant, Stephan Crainic, and James M. Hogle.)

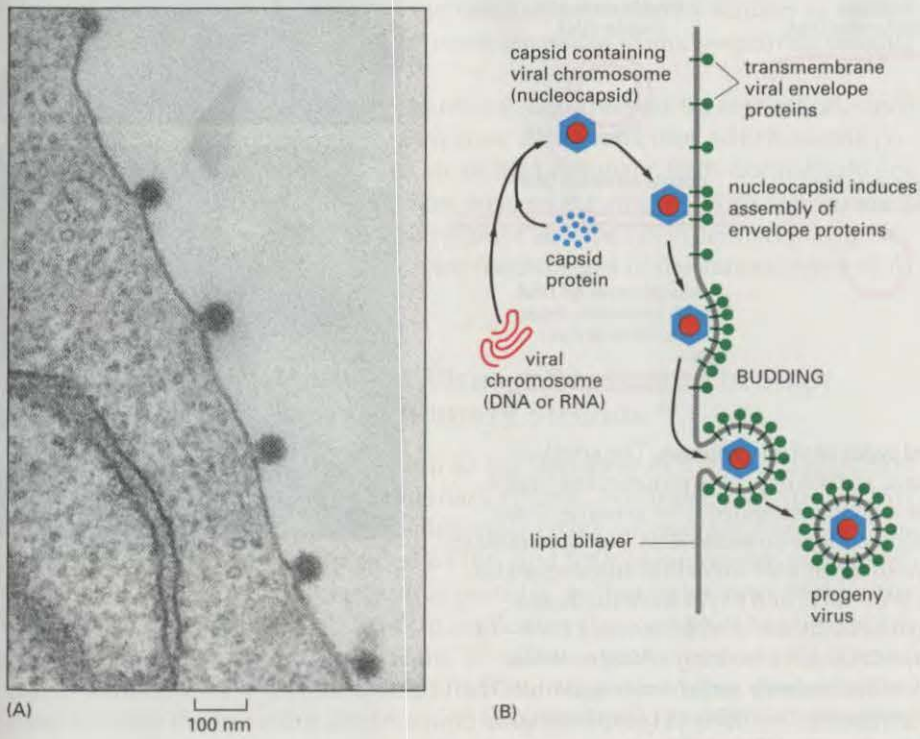
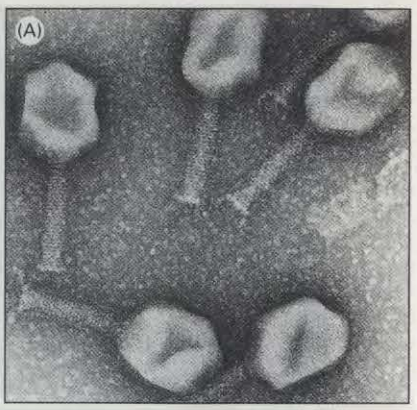
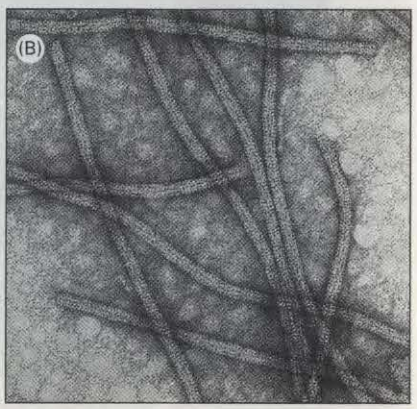
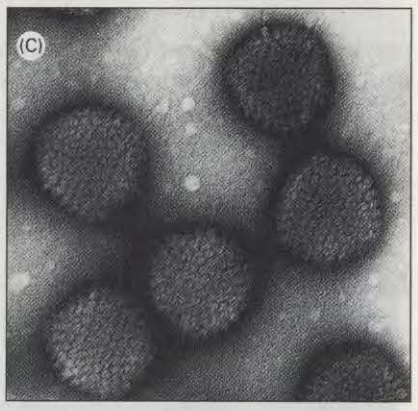


Figure 6–73 Acquisition of a viral envelope. (A) Electron micrograph of a thin section of an animal cell from which several copies of an enveloped virus (Semliki forest virus) are budding. (B) Schematic view of the envelope assembly and budding process. Whereas the lipid bilayer that surrounds the capsid is parasitized directly from the plasma membrane of the host cell, the only proteins in this lipid bilayer are those encoded by the viral genome. (A, courtesy of M. Olsen and G. Griffiths.)

Sarepta Exhibit 1040, page 112







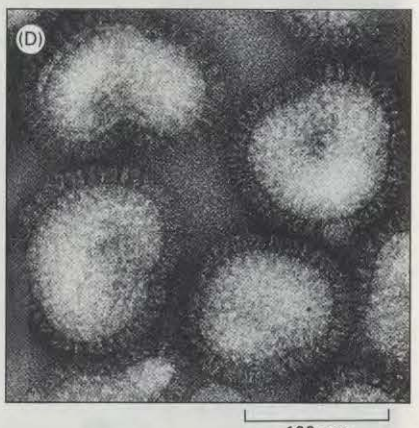


Figure 6–74 The coats of viruses. These electron micrographs of negatively stained virus particles are all at the same scale. (A) *Bacteriophage T4*, a large DNA-containing virus that infects *E. coli*. The DNA is stored in the bacteriophage head and injected into the bacterium through the cylindrical tail. (B) *Potato virus X*, a filamentous plant virus that contains an RNA genome. (C) *Adenovirus*, a DNA-containing virus that can infect human cells. The protein capsid forms the outer surface of this virus. (D) *Influenza virus*, a large RNA-containing animal virus whose protein capsid is further enclosed in a lipid-bilayer-based envelope containing protruding spikes of viral glycoprotein. (A, courtesy of James Paulson; B, courtesy of Graham Hills; C, courtesy of Mei Lie Wong; D, courtesy of R.C. Williams and H.W. Fisher.)

capsid is further enclosed by a lipid bilayer membrane that contains proteins. Many of these *enveloped viruses* acquire their envelope in the process of budding from the plasma membrane (Figure 6–73). This budding process allows the virus particles to leave the cell without disrupting the plasma membrane and, therefore, without killing the cell. Electron micrographs that emphasize the differences among viral coats are presented in Figure 6–74.

Viral Genomes Come in a Variety of Forms and Can Be Either RNA or DNA [51]

As discussed earlier, the DNA double helix has the advantages of stability and easy repair. If one polynucleotide chain is accidentally damaged, its complementary chain permits the damage to be readily corrected. This concern with repair, however, need not bother small viral chromosomes that contain only several thousand nucleotides. The chance of accidental damage is very small compared with the risk to a cell genome containing millions of nucleotides.

The genetic information of a virus can, therefore, be carried in a variety of unusual forms, including RNA instead of DNA. A viral chromosome may be a single-stranded RNA chain, a double-stranded RNA helix, a circular single-

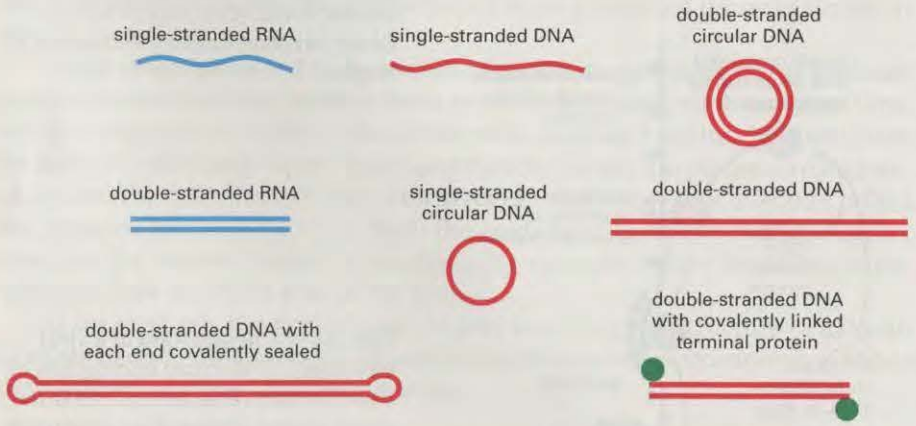


Figure 6–75 Schematic drawings of several types of viral genomes. The smallest viruses contain only a few genes and can have an RNA or a DNA genome; the largest viruses contain hundreds of genes and have a double-stranded DNA genome. Some examples of these types of viruses are as follows: *single-stranded RNA*—tobacco mosaic virus, bacteriophage R17, poliovirus; *double-stranded RNA*—reovirus; *single-stranded DNA*—parvovirus; *single-stranded circular DNA*—M13 and ϕX174 bacteriophages; *double-stranded circular DNA*—SV40 and polyomaviruses; *double-stranded DNA*—T4 bacteriophage, herpes virus; *double-stranded DNA with covalently linked terminal protein*—adenovirus; *double-stranded DNA with covalently sealed ends*—poxvirus. The peculiar ends (as well as the circular forms) overcome the difficulty of replicating the last few nucleotides at the end of a DNA chain (see pp. 338 and 364).

Sarepta Exhibit 1040, page 113

stranded DNA chain, or a linear single-stranded DNA chain. Moreover, although some viral chromosomes are simple linear DNA double helices, circular DNA double helices and more complex linear DNA double helices are also common. Several viruses have protein molecules covalently attached to the 5′ ends of their DNA strands, for example, and the DNA double helices from the very large poxviruses have their opposite strands at each end covalently joined through phosphodiester linkages (Figure 6–75).

A Viral Chromosome Codes for Enzymes Involved in the Replication of Its Nucleic Acid [52]

Each type of viral genome requires unique enzymatic tricks for its replication and thus must encode not only the viral coat protein but also one or more of the enzymes needed to replicate the viral nucleic acid. The amount of information that a virus brings into a cell to ensure its own selective replication varies greatly. The DNA of the relatively large bacteriophage T4, for example, contains about 300 genes, including at least 30 genes that ensure the rapid replication of the T4 chromosome in its *E. coli* host cell (Figure 6–76). T4 DNA replication has the unusual feature that 5-hydroxymethyl-C is incorporated in place of C in its DNA. The unusual base composition of the T4 DNA makes it readily distinguishable from host DNA and selectively protects it from nucleases encoded in the T4 genome that thus degrade only the *E. coli* chromosome. Still other proteins alter host cell RNA polymerase molecules so that they are unable to transcribe *E. coli* DNA and instead transcribe different sets of bacteriophage genes at different stages of infection, according to the needs of the phage.

Smaller DNA viruses, such as the monkey virus SV40 and the tiny bacteriophage M13, carry much less genetic information. They rely heavily on host-cell enzymes to carry out their DNA synthesis, parasitizing most of the host-cell DNA replication proteins. Most DNA viruses, however, code for proteins that selectively initiate the synthesis of their own DNA, recognizing a particular nucleotide sequence in the virus that serves as a *replication origin*. This is important because a virus must override the cellular control signals that would otherwise cause the viral DNA to replicate in pace with the host cell DNA, doubling only once in each cell cycle. We do not yet understand very much about how eucaryotic cells regulate their own DNA synthesis, and the mechanisms used by viruses to escape from this regulation—which are much more accessible to study—provide insights into the host mechanisms.

RNA viruses have particularly specialized requirements for replication, since to reproduce their genomes they must copy RNA molecules, which means polymerizing nucleoside triphosphates on an RNA template. Cells normally do not have enzymes to carry out this reaction, so even the smallest RNA viruses must encode their own RNA-dependent polymerase enzymes in order to replicate. We now look in more detail at the replication mechanisms of the various types of viruses.

Both RNA Viruses and DNA Viruses Replicate Through the Formation of Complementary Strands [53]

Like DNA replication, the replication of the genomes of RNA viruses occurs through the formation of complementary strands. For most RNA viruses this process is catalyzed by specific RNA-dependent RNA polymerase enzymes (*replicases*). These enzymes are encoded by the viral RNA chromosome and are often incorporated into the progeny virus particles, so that upon entry of the virus into a cell, they can immediately begin replicating the viral RNA. Replicases are always packaged into the capsid of the so-called *negative-strand RNA viruses*, such as influenza or vesicular stomatitis virus. Negative-strand viruses are so called because the infecting single strand does not code for protein; instead its complementary strand carries the coding sequences. Thus the infecting strand

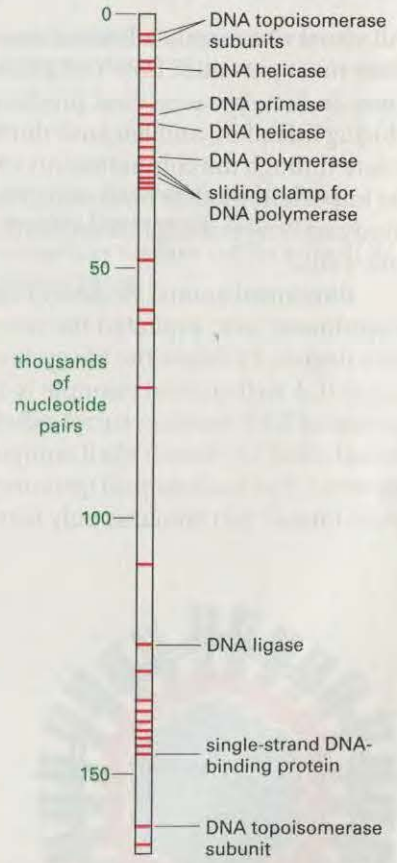


Figure 6–76 The T4 bacteriophage chromosome, showing the positions of the more than 30 genes involved in T4 DNA replication. The genome of bacteriophage T4 consists of 169,000 nucleotide pairs and encodes about 300 different proteins.

Sarepta Exhibit 1040, page 114

remains impotent without a preformed replicase. In contrast, the viral RNA of *positive-strand RNA viruses,* such as poliovirus, can serve as mRNA and produce a replicase once it enters the cell; therefore the naked genome itself is infectious.

The synthesis of viral RNA always begins at the 3′ end of the RNA template, starting with the synthesis of the 5′ end of the new viral RNA molecule and progressing in the 5′-to-3′ direction until the 5′ end of the template is reached. There are no error-correcting mechanisms for viral RNA synthesis, and error rates are similar to those in DNA transcription (about 1 error in 10^4 nucleotides synthesized). This is not a serious deficiency as long as the RNA chromosome is relatively short; for this reason the genomes of all RNA viruses are small relative to those of the large DNA viruses.

All DNA viruses begin their replication at a replication origin, where special initiator proteins bind and then attract the replication enzymes of the host cell (see Figure 8–34). There are many different replication pathways, however. The complexity of these diverse replication schemes reflects, in part, the problem of replicating the ends of a simple linear DNA molecule, given a DNA polymerase enzyme that cannot begin synthesis without a primer (see pp. 253–254). DNA viruses have solved this problem in a variety of ways: some have circular DNA genomes and thus no ends; others have linear DNA genomes that repeat their terminal sequences or end in loops; while still others have special terminal proteins that serve to prime the DNA polymerase directly (see Figure 6–75).

Viruses Exploit the Intracellular Traffic Machinery of their Host Cells [54]

All viruses have only a limited amount of nucleic acid in their genome, and so they must parasitize host-cell pathways for most of the steps in their reproduction. In fact, because viral products are usually synthesized in large amounts during infection, and because during its life cycle the virus follows a sequential route through the compartments of the host cell, virus-infected cells have served as important models for tracing the pathways of intracellular transport and for studying how essential biosynthetic reactions are compartmentalized in eucaryotic cells.

Enveloped animal viruses, in which the genome is enclosed in a lipid-bilayer membrane, have exploited the compartmentalization of the cell to an especially fine degree. To follow the life cycle of an enveloped virus is to take a tour through the cell. A well-studied example is *Semliki forest virus,* which consists of a single-stranded RNA genome surrounded by a *capsid* formed by a regularly arranged icosahedral (20-faced) shell composed of many copies of one protein (called C protein). The *nucleocapsid* (genome + capsid) is surrounded by a closely apposed lipid bilayer that contains only three types of polypeptide chains, each encoded

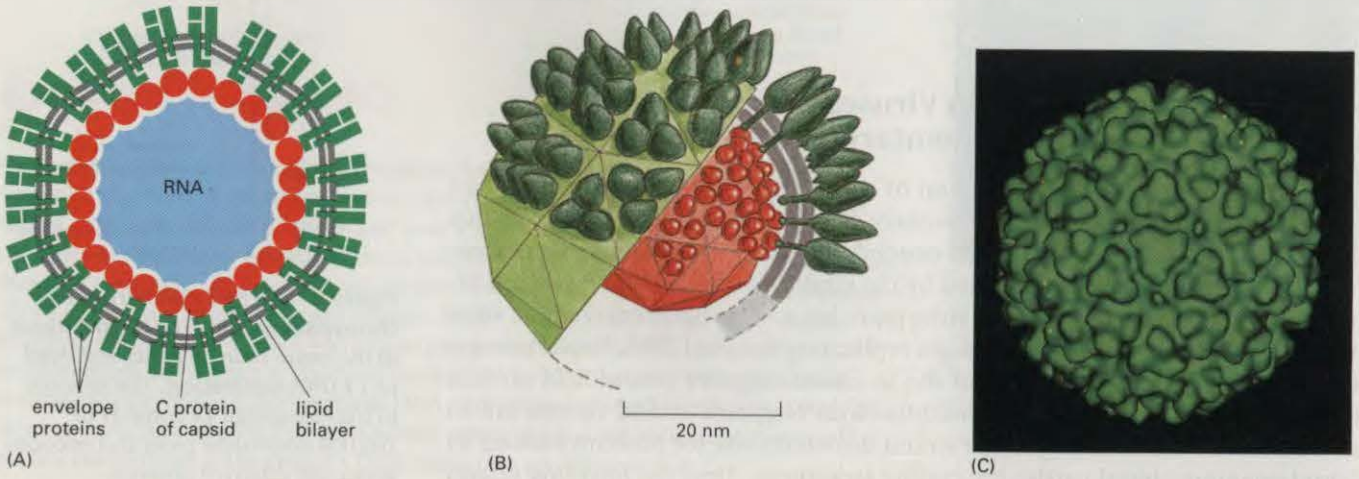


Figure 6–77 The structure of Semliki forest virus. Schematic drawings of a cross-section (A) and an exploded three-dimensional view (B) of the virus. (C) A three-dimensional reconstruction of the surface of the virus derived from cryoelectron micrographs of unstained specimens. The virus has a total mass of 46 million daltons. (B, adapted from S.C. Harrison, *Curr. Opin. Struct. Biol.* 2:293–299, 1992. Current Science; C, courtesy of Stephen Fuller.)

Sarepta Exhibit 1040, page 115

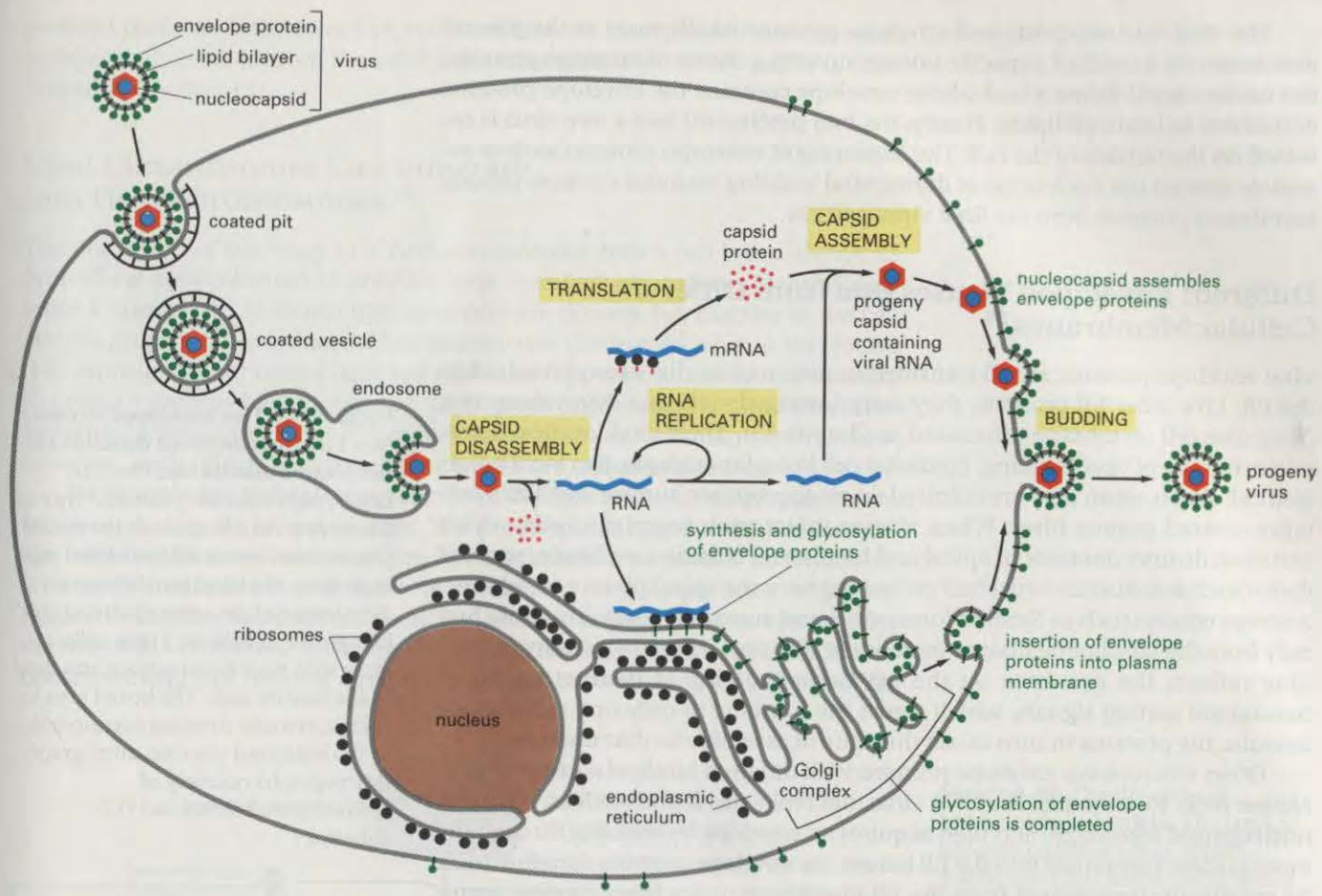


Figure 6–78 The life cycle of the Semliki forest virus. The virus parasitizes the host cell for most of its biosyntheses.

by the viral RNA. These **envelope proteins** form heterotrimers that span the lipid bilayer and interact with the C protein of the nucleocapsid, linking the membrane and nucleocapsid together (Figure 6–77). The glycosylated portions of the envelope proteins are always on the outside of the lipid bilayer, and each trimer forms a "spike" that can be seen in electron micrographs projecting outward from the surface of the virus (Figure 6–77C).

Infection is initiated when an envelope protein on the virus binds to a normal cell protein that serves as its receptor on the host-cell plasma membrane. The virus then uses the cell's normal endocytic pathway to enter the cell by receptor-mediated endocytosis and is delivered to early endosomes (discussed in Chapter 13). But instead of being transferred from endosomes to lysosomes, the virus escapes from the endosome by virtue of the special properties of one of its envelope proteins. At the acidic pH of the endosome, this protein causes the viral envelope to fuse with the endosome membrane, releasing the bare nucleocapsid into the cytosol. The nucleocapsid is "uncoated" in the cytosol, releasing the viral RNA, which is then translated by host-cell ribosomes to produce a virus-encoded RNA polymerase. This in turn makes many copies of viral RNA, some of which serve as mRNA molecules to direct the synthesis of the structural proteins of the virus—the capsid C protein and the three envelope proteins.

The newly synthesized capsid and envelope proteins follow separate pathways through the cytoplasm. The envelope proteins, like the plasma membrane proteins of the host cell, are synthesized by ribosomes that are bound to the rough ER; in contrast, the capsid protein, like the cytosolic proteins of the cell, is synthesized by ribosomes that are not membrane bound. The newly synthesized capsid protein binds to the recently replicated viral RNA to form new nucleocapsids. The envelope proteins, in contrast, are inserted into the membrane of the ER, where they are glycosylated, transported to the Golgi apparatus, and then delivered to the plasma membrane (Figure 6–78).

Sarepta Exhibit 1040, page 116

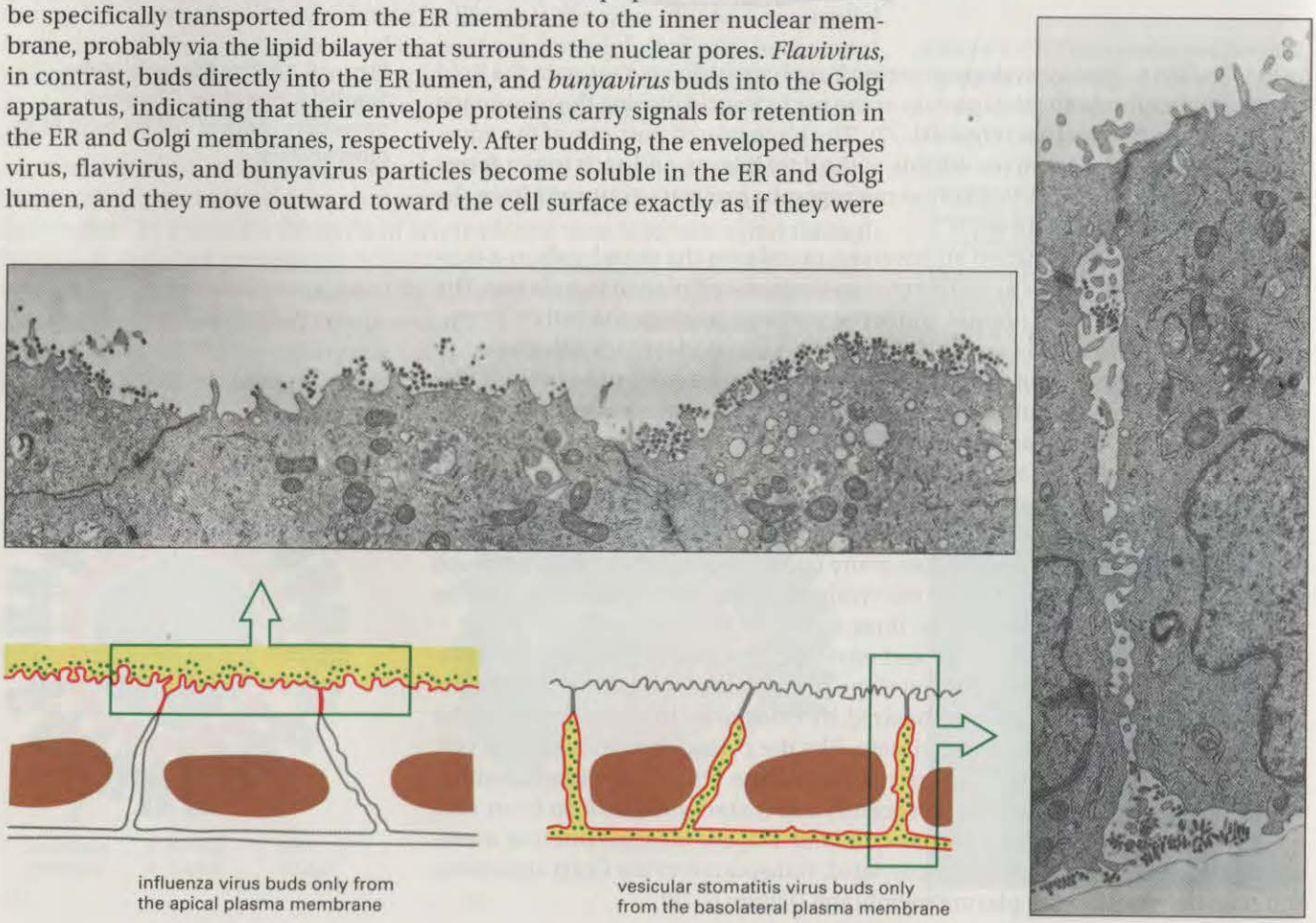
The viral nucleocapsids and envelope proteins finally meet at the plasma membrane. As a result of a specific interaction with a cluster of envelope proteins, the nucleocapsid forms a bud whose envelope contains the envelope proteins embedded in host-cell lipids. Finally, the bud pinches off and a free virus is released on the outside of the cell. The clustering of envelope proteins as they assemble around the nucleocapsid during viral budding excludes the host plasma membrane proteins from the final virus particle.

Different Enveloped Viruses Bud from Different Cellular Membranes [55]

Viral envelope proteins are all transmembrane proteins that are synthesized in the ER. Like other ER proteins, they carry sorting signals that direct them to a particular cell membrane (discussed in Chapter 13). Their final location determines the site of viral budding. Epithelial cell lines, for example, can form polarized cell sheets when they are cultured on an appropriate surface, such as a collagen-coated porous filter. When viruses infect such polarized cells, which maintain distinct domains of apical and basolateral plasma membrane, some of them (such as influenza virus) bud exclusively from the apical plasma membrane, whereas others (such as Semliki forest virus and vesicular stomatitis virus) bud only from the basolateral plasma membrane (Figure 6–79). This polarity of budding reflects the presence on the envelope proteins of distinct apical or basolateral sorting signals, which direct the proteins to only one cell-surface domain; the proteins in turn cause the virus to assemble in that domain.

Other viruses have envelope proteins with different kinds of sorting signals. *Herpes virus,* for example, is a DNA virus that replicates in the nucleus, where its nucleocapsid assembles, and then acquires an envelope by budding through the inner nuclear membrane into the ER lumen; the envelope proteins therefore must be specifically transported from the ER membrane to the inner nuclear membrane, probably via the lipid bilayer that surrounds the nuclear pores. *Flavivirus,* in contrast, buds directly into the ER lumen, and *bunyavirus* buds into the Golgi apparatus, indicating that their envelope proteins carry signals for retention in the ER and Golgi membranes, respectively. After budding, the enveloped herpes virus, flavivirus, and bunyavirus particles become soluble in the ER and Golgi lumen, and they move outward toward the cell surface exactly as if they were

Figure 6–79 Two enveloped viruses that bud from different domains of the plasma membrane. Electron micrographs showing that one type of enveloped virus buds from the apical plasma membrane while another type buds from the basolateral plasma membrane of the same epithelial cell line grown in culture. These cells grow with their basal surface attached to the culture dish. The boxed area in each schematic drawing corresponds to the indicated electron micrograph. (Micrographs courtesy of E. Rodriguez-Boulan and D.D. Sabatini.)



Sarepta Exhibit 1040, page 117

secreted proteins; in the *trans* Golgi network they are incorporated into transport vesicles and secreted from the cell by the constitutive secretory pathway (discussed in Chapter 13).

Viral Chromosomes Can Integrate into Host Chromosomes [56]

The end result of the entry of a viral chromosome into a cell is not always its immediate multiplication to produce large numbers of progeny. Many viruses enter a *latent* state, in which their genomes are present but inactive in the cell and no progeny are produced. Viral latency was discovered when it was found that exposure to ultraviolet light induced many apparently uninfected bacteria to produce progeny bacteriophages. Subsequent experiments showed that these *lysogenic bacteria* carry in their chromosomes a dormant but complete viral chromosome. Such integrated viral chromosomes are called **proviruses.**

Bacteriophages that can integrate their DNA into bacterial chromosomes are known as *temperate bacteriophages.* The prototypic example is the bacteriophage lambda, discussed earlier. When lambda infects a suitable *E. coli* host cell, it normally multiplies to produce several hundred progeny particles, which are released when the bacterial cell lyses; this is called a *lytic infection.* More rarely, the free ends of the linear infecting DNA molecules join to form a DNA circle that becomes integrated into the circular host *E. coli* chromosome by a site-specific

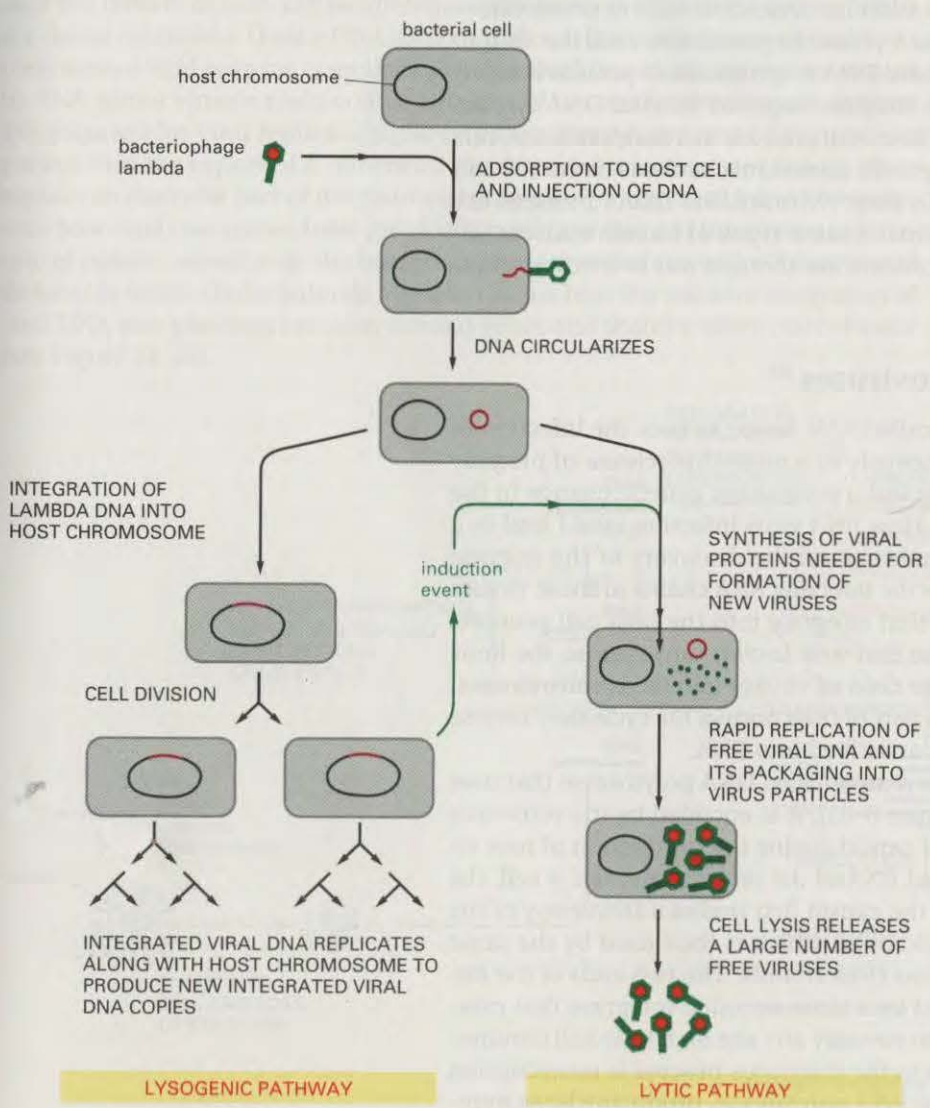


Figure 6–80 The life cycle of bacteriophage lambda. The lambda genome contains about 50,000 nucleotide pairs and encodes about 50 proteins. Its double-stranded DNA can exist in either linear or circular forms. As shown, the bacteriophage can multiply by either a lytic or a lysogenic pathway in the *E. coli* bacterium. When the bacteriophage is growing in the lysogenic state, damage to the cell causes the integrated viral DNA (provirus) to exit from the host chromosome and shift to lytic growth. The entrance and exit of the DNA from the chromosome are site-specific genetic recombination events catalyzed by the lambda *integrase* protein (see Figure 6–68).

Sarepta Exhibit 1040, page 118

recombination event. The resulting lysogenic bacterium, carrying the proviral lambda chromosome, multiples normally until it is subjected to an environmental insult, such as exposure to ultraviolet light or ionizing radiation. The resulting cell debilitation induces the integrated provirus to leave the host chromosome and begin a normal cycle of viral replication. In this way the integrated provirus need not perish with its damaged host cell but has a chance to escape to other *E. coli* cells (Figure 6–80).

The Continuous Synthesis of Some Viral Proteins Can Make Cells Cancerous [57]

Animal cells, like bacteria, can offer viruses an alternative to lytic growth. *Permissive* cells permit DNA viruses to multiply lytically and kill the cell. *Nonpermissive* cells may allow the DNA virus to enter but not to replicate lytically; in a small percentage of such cells the viral chromosome either becomes integrated into the host cell genome, where it is replicated along with the host chromosomes, or forms a plasmid—a circular DNA molecule—that replicates in a controlled fashion without killing the cell. Such nonpermissive infections sometimes result in a genetic change in the host cell, causing it to proliferate in an ill-controlled way and thus transforming it into its cancerous equivalent. In this case the DNA virus is called a *DNA tumor virus* and the process is called virus-mediated *neoplastic transformation.* The most extensively studied DNA tumor viruses are two papovaviruses, SV40 and polyoma. Their transforming ability has been traced to several viral proteins that cooperate to stimulate quiescent cells to proliferate—that is, they drive the cells from G_0 into S phase. In permissive cells the shift to S phase (the phase of the cell cycle where DNA is synthesized) provides the virus with all of the host-cell replication enzymes required for viral DNA synthesis. When a provirus happens to make these viral proteins in a nonpermissive cell, they can override some of the normal growth control mechanisms in the cell and its progeny. By this means some DNA tumor viruses that infect humans are known to contribute to the development of some types of human cancers (although the great majority of human cancers are thought not to involve tumor viruses).

RNA Tumor Viruses Are Retroviruses [58]

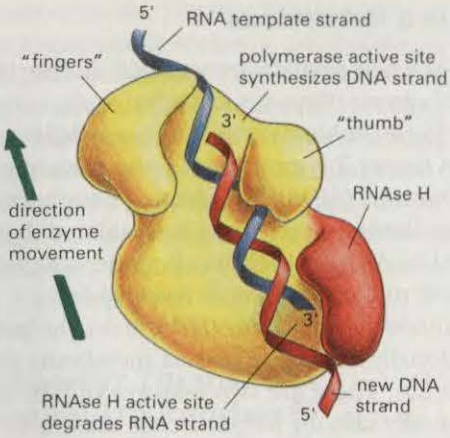
For one group of RNA viruses, the so-called *RNA tumor viruses,* the infection of a permissive cell often leads simultaneously to a nonlethal release of progeny virus from the cell surface by budding and a permanent genetic change in the infected cell that makes it cancerous. How RNA virus infection could lead to a permanent genetic alteration was unclear until the discovery of the enzyme *reverse transcriptase,* which transcribes the infecting RNA chains of these viruses into complementary DNA molecules that integrate into the host cell genome. RNA tumor viruses—which include the first well-known tumor virus, the Rous sarcoma virus—are members of a large class of viruses known as **retroviruses.** These viruses are so named because as part of their normal life cycle they reverse the normal process in which DNA is transcribed into RNA.

The enzyme **reverse transcriptase** is an unusual DNA polymerase that uses either RNA or DNA as a template (Figure 6–81); it is encoded by the retrovirus RNA and is packaged inside each viral capsid during the production of new virus particles. When the single-stranded RNA of the retrovirus enters a cell, the reverse transcriptase brought in with the capsid first makes a DNA copy of the RNA strand to form a DNA-RNA hybrid helix, which is then used by the same enzyme to make a double helix with two DNA strands. The two ends of the linear viral DNA molecule are recognized by a virus-encoded integrase that catalyzes the insertion of the viral DNA into virtually any site on a host-cell chromosome (see Figure 6–70). The next step in the infectious process is transcription of the integrated viral DNA by host-cell RNA polymerase, producing large num-

Sarepta Exhibit 1040, page 119



(A)



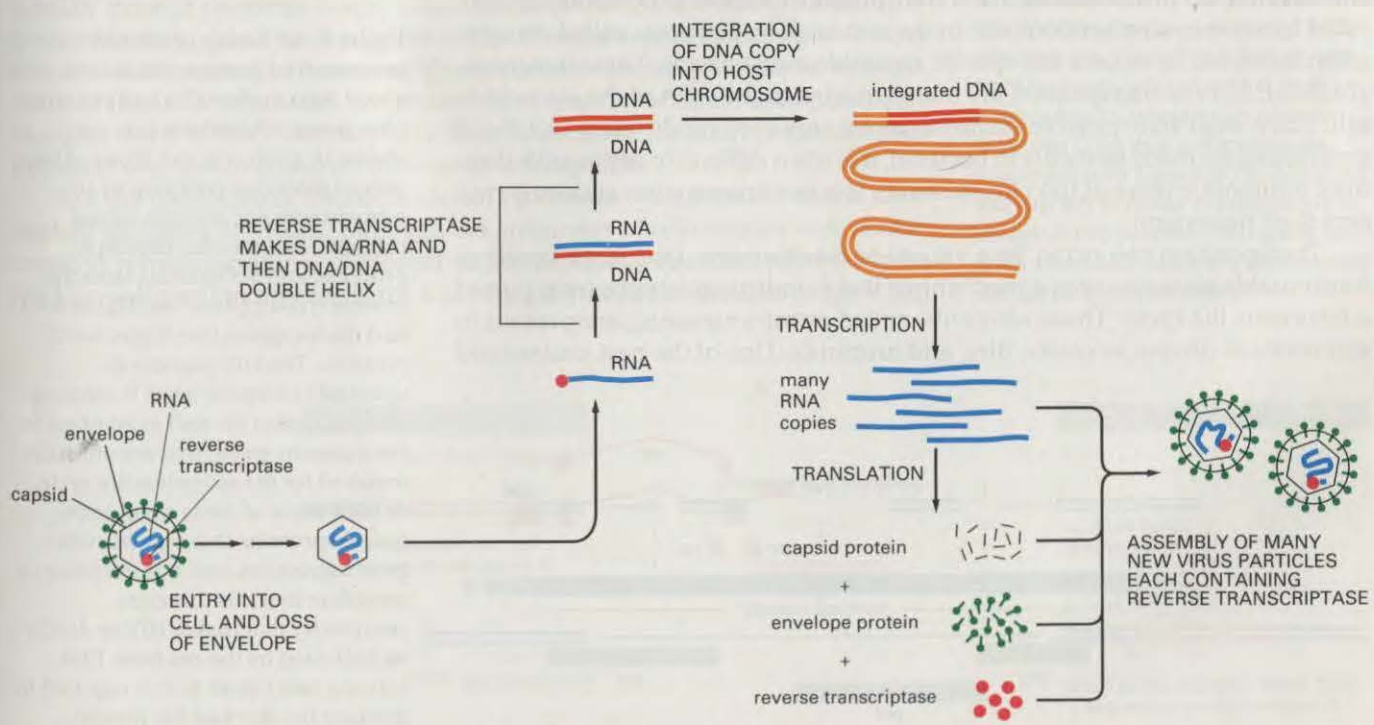
(B)

Figure 6–81 Reverse transcriptase. (A) The three-dimensional structure of the enzyme from HIV-1 (the AIDS virus), determined by x-ray crystallography; (B) a schematic view of a model for its activity on an RNA template. Note that the polymerase domain (*yellow*) has a covalently attached RNAse domain (*red*) that degrades an RNA strand in an RNA/DNA helix. This activity helps the polymerase convert the initial hybrid helix into a DNA double helix. (A, courtesy of Tom Steitz; B, adapted from L.A. Kohlstaedt et al., *Science* 256:1783–1790, 1992. © 1992 the AAAS.)

bers of viral RNA molecules identical to the original infecting genome. Finally, these RNA molecules are translated to produce the capsid, envelope, and reverse transcriptase proteins that are assembled with the RNA into new enveloped virus particles, which bud from the plasma membrane (Figure 6–82).

Both RNA and DNA tumor viruses transform cells because the permanent presence of the viral DNA in the cell causes the synthesis of new proteins that alter the control of host-cell proliferation. The genes that code for such proteins are called *oncogenes.* Unlike DNA tumor viruses, whose oncogenes typically encode normal viral proteins essential for viral multiplication, the oncogenes carried by RNA tumor viruses are modified versions of normal host-cell genes that are not required for viral replication. Since only a limited amount of RNA can be packed into the capsid of a retrovirus, the acquired oncogene sequences often replace an essential part of the retroviral genome. In Chapters 15 and 24 we discuss how viral oncogenes have provided important clues to the causes and nature of cancer, as well as to the normal mechanisms that control cell growth and division in multicellular animals. We also discuss how the random integration of viral DNA into genomes can alter normal genes and thereby affect cell behavior (see Figure 24–24).

Figure 6–82 The life cycle of a retrovirus. The retrovirus genome consists of an RNA molecule of about 8500 nucleotides; two such molecules are packaged into each viral particle. The enzyme *reverse transcriptase* first makes a DNA copy of the viral RNA molecule and then a second DNA strand, generating a double-stranded DNA copy of the RNA genome. The integration of this DNA double helix into the host chromosome, catalyzed by the viral integrase, is required for the synthesis of new viral RNA molecules by the host-cell RNA polymerase.



Sarepta Exhibit 1040, page 120

The Virus That Causes AIDS Is a Retrovirus [59]

In 1982 physicians first became aware of a new sexually transmitted disease that was associated with an unusual form of cancer (Kaposi's sarcoma) and a variety of unusual infections. Because both of these problems reflect a severe deficiency in the immune system—specifically in helper T lymphocytes—the disease was named *acquired immune deficiency syndrome (AIDS)*. By culturing lymphocytes from patients with an early stage of the disease, a retrovirus was isolated that is now known to be the causative agent of AIDS, which has become a rapidly spreading epidemic that threatens to kill millions of people worldwide.

The retrovirus, called *human immunodeficiency virus (HIV)*, enters helper T lymphocytes by first binding to a functionally important plasma membrane protein called CD4 (discussed in Chapter 23). There are two features of HIV that make it especially deadly. First, it eventually kills the helper T cells that it infects rather than living in symbiosis with them, as do most other retroviruses, and helper T cells are vitally important in defending us against infection. Second, the provirus tends to persist in a latent state in the chromosomes of an infected cell without producing virus until it is activated by an unknown rare event; this ability to hide greatly complicates any attempt to treat the infection with antiviral drugs.

Much current research on AIDS is aimed at understanding the life cycle of HIV. The complete nucleotide sequence of the viral RNA has been determined. This has made it possible to identify and study each of the proteins that it encodes. The three-dimensional structure of its reverse transcriptase (see Figure 6–81) is being used to help design new drugs that inhibit the enzyme. The nine genes of this retrovirus are displayed on the HIV genetic map in Figure 6–83.

Some Transposable Elements Are Close Relatives of Retroviruses [60]

Because many viruses can move into and out of their host chromosomes, any large genome is likely to contain a number of different proviruses. Most genomes are also likely to house a variety of mobile DNA sequences that do not form viral particles and cannot leave the cell. Such **transposable elements** range in length from a few hundred to tens of thousands of base pairs, and they are usually present in multiple copies per cell. One can consider these elements as tiny parasites hidden in chromosomes. Each transposable element is occasionally activated to move to another DNA site in the same cell by a process called *transposition*, catalyzed by its own site-specific recombination enzyme. These integrases, also referred to as *transposases*, are often encoded in the DNA of the element itself. Since most transposable elements move only very rarely (once in 10^5 cell generations for many elements in bacteria), it is often difficult to distinguish them from nonmobile parts of the chromosome. It is not known what suddenly triggers their movement.

Transposition can occur by a variety of mechanisms. One large family of transposable elements uses a mechanism that is indistinguishable from part of a retrovirus life cycle. These elements, called *retrotransposons*, are present in organisms as diverse as yeasts, flies, and mammals. One of the best-understood

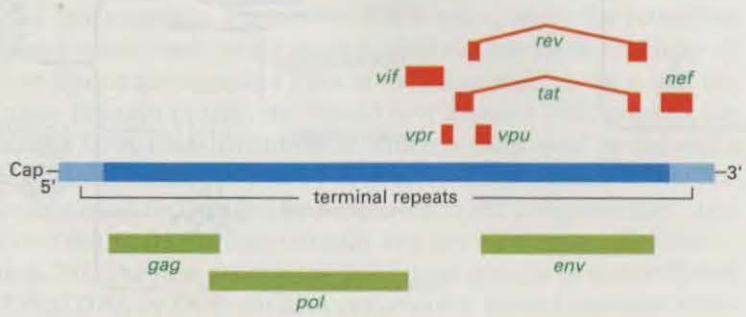


Figure 6–83 A map of the HIV genome. The genome consists of about 9000 nucleotides and contains nine genes, whose locations are shown in *green* and *red*. Three of the genes (*green*) are common to all retroviruses: *gag* encodes capsid proteins, *env* encodes envelope proteins, and *pol* encodes both the reverse transcriptase (see Figure 6–81) and the integrase (see Figure 6–70) proteins. The HIV genome is unusually complex, since it contains six small genes (in *red*) in addition to the three (in *green*) that are normally required for the retrovirus life cycle. At least some of these small genes encode proteins that regulate viral gene expression, and it is tempting to speculate that it is this extra complexity that makes HIV so deadly. As indicated by the *red lines*, RNA splicing (see Figure 8–7) is required to produce the Rev and Tat proteins.

Sarepta Exhibit 1040, page 121

retrotransposons is the so-called Ty1 element of yeasts. The first step in its transposition is the transcription of the entire transposable element, producing an RNA copy of the element that is more than 5000 nucleotides long. This transcript encodes a reverse transcriptase enzyme that makes a double-stranded DNA copy of the RNA molecule via a RNA/DNA hybrid intermediate, precisely mimicking the early stages of infection by a retrovirus (see Figure 6–82). The analogy continues as the linear DNA molecule uses an integrase to integrate into a randomly selected site on the chromosome. Although the resemblance to a retrovirus is striking, unlike a retrovirus, the Ty1 element does not have a functional protein coat and therefore can only move within a single cell and its progeny.

Other Transposable Elements Transfer Themselves Directly from One Site in the Genome to Another [61]

Unlike retrotransposons, many transposable elements rarely exist free of the host chromosome; the transposases that catalyze their movement can act on the DNA of the element while it is still integrated in the host genome. The transposase binds to a short sequence that is repeated in reverse orientation at each end of the element, thereby holding these two ends close together while catalyzing the subsequent recombination event. The mechanism is closely related to that used by the retrovirus integrase (see Figure 6–70). For some transposable elements the transposition mechanism differs only in that the linear DNA molecule to be integrated must be cut out of a much longer DNA molecule, leaving a break in the vacated chromosome (Figure 6–84). This break is subsequently resealed, but in the process the DNA sequence is often altered, resulting in a mutation at the old chromosomal site.

Other transposable elements replicate when they move. In the best-studied example, a covalent connection is first made between the transposable element and a randomly selected target site; this connection then triggers a localized synthesis of DNA that results in one copy of the replicated transposable element being inserted at a new chromosomal site, while the other copy remains at the old one (Figure 6–85). The mechanism is closely related to the nonreplicative mechanism just described, and it starts in nearly the same way; indeed, some transposable elements can move by either pathway.

In addition to moving themselves, all types of transposable elements occasionally move or rearrange neighboring DNA sequences of the host genome. They frequently cause deletions of adjacent nucleotide sequences, for example, or carry them to another site. The presence of transposable elements makes the arrangement of the DNA sequences in chromosomes much less stable than previously thought, and it is likely that they have been responsible for many important evolutionary changes in genomes (discussed in Chapter 8).

Are the transposable elements also of evolutionary importance as the most ancient ancestors of viruses? Although the precursors of retroviruses were almost certainly retrotransposons, all present-day transposable elements rely heavily on DNA-based reaction mechanisms. But very early cells are thought to have had

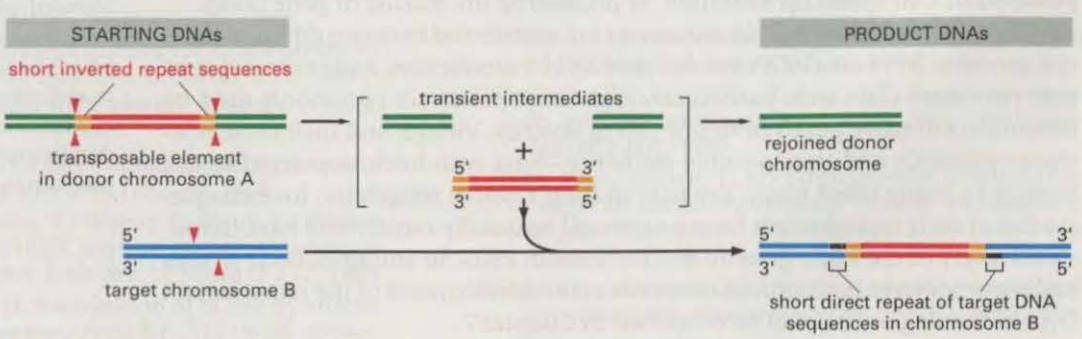


Figure 6–84 The direct movement of a transposable element from one chromosomal site to another. Transposable elements of this type can be recognized by the "inverted repeat DNA sequences" (*orange*) at their ends. Experiments show that these sequences, which can be as short as 20 nucleotides, are all that is necessary for the DNA between them to be transposed by the particular transposase enzyme associated with the element. The mechanism shown here is closely related to that used by a retrovirus to integrate its double-stranded DNA into a chromosome (compare with Figure 6–70). Although the gap left in donor chromosome is resealed, the process often alters the DNA sequence, causing a mutation at the donor site (not shown).

Sarepta Exhibit 1040, page 122

RNA rather than DNA genomes, so we must look to RNA-based mechanisms for the ultimate origin of viruses.

Most Viruses Probably Evolved from Plasmids [62]

Even the largest viruses depend heavily on their host cells for biosynthesis; no known virus makes its own ribosomes or generates the ATP it requires, for example. Clearly, therefore, cells must have evolved before viruses. The precursors of the first viruses were probably small nucleic acid fragments that developed the ability to multiply independently of the chromosomes of their host cells. Such independently replicating elements, called **plasmids,** can replicate indefinitely outside the host chromosome. Plasmids occur in both DNA and RNA forms, and, like viruses, they contain a special nucleotide sequence that serves as an origin of replication. Unlike viruses, however, they cannot make a protein coat and therefore cannot move from cell to cell in this way.

The first RNA plasmids may have resembled the *viroids* found in some plant cells. These small RNA circles, only 300 to 400 nucleotides long, are replicated despite the fact that they do not code for any protein. Having no protein coat, viroids exist as naked RNA molecules and pass from plant to plant only when the surfaces of both donor and recipient cells are damaged so that there is no membrane barrier for the viroid to pass. Under the pressure of natural selection, such independently replicating elements could be expected to acquire nucleotide sequences from the host cell that would facilitate their own multiplication, including sequences that code for proteins. Some present-day plasmids are indeed quite complex, encoding proteins and RNA molecules that regulate their replication, as well as proteins that control their partitioning into daughter cells. The largest known plasmids are double-stranded DNA circles more than 100,000 base pairs long.

The first virus probably appeared when an RNA plasmid acquired a gene coding for a capsid protein. But a capsid can enclose only a limited amount of nucleic acid; therefore a virus is limited in the number of genes it can contain. Forced to make optimal use of their limited genomes, some small viruses evolved *overlapping genes,* in which part of the nucleotide sequence encoding one protein is used (in the same or a different reading frame) to encode a second protein. Other viruses evolved larger capsids and consequently could accommodate more genes.

With their unique ability to transfer nucleic acid sequences across species barriers, viruses have almost certainly played an important part in the evolution of the organisms they infect. Many recombine frequently with their host-cell genome and with one another. In this way they can pick up small pieces of host chromosome at random and carry them to different cells or organisms. Moreover, integrated copies of viral DNA (proviruses) have become a normal part of the genome of most organisms. Examples of such proviruses include the lambda family of bacteriophages and the so-called endogenous retroviruses found in numerous copies in vertebrate genomes. The integrated viral DNA can become altered so that it cannot produce a complete virus but can still encode proteins, some of which may be useful to the host cell. Therefore, viruses, like sexual reproduction, can speed up evolution by promoting the mixing of gene pools.

The process in which DNA sequences are transferred between different host-cell genomes by means of a virus is called *DNA transduction,* and several viruses that transduce DNA with particularly high frequencies are commonly used by researchers to move genes from one cell to another. Viruses and their close relatives—plasmids and transposable elements—have also been important to cell biology in many other ways. Because of their relative simplicity, for example, studies of their reproduction have progressed unusually rapidly and have illuminated many of the basic genetic mechanisms in cells. In addition, both viruses and plasmids have been crucial elements in the development of the recombinant DNA technologies that will be described in Chapter 7.

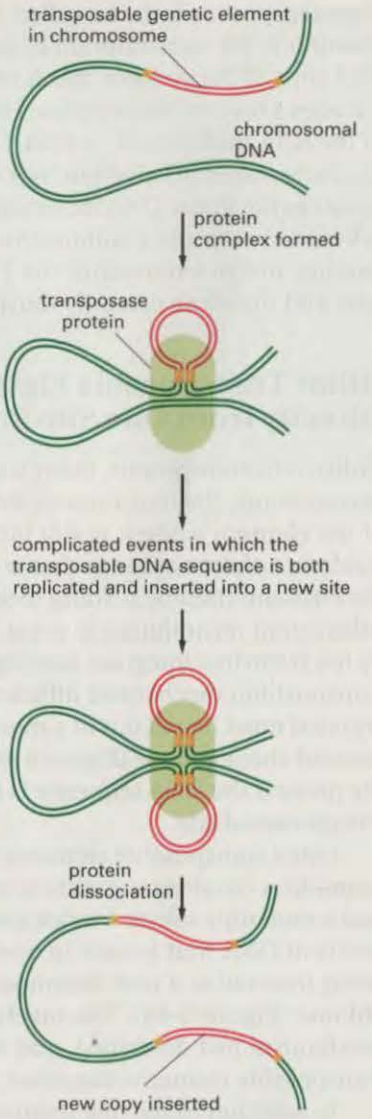


Figure 6–85 The replicative movement of a transposable element within a chromosome. The element shown replicates during transposition, its movement occurring without it being excised from its original site. The two inverted repeat DNA sequences that commonly flank the two ends of transposable elements are shown in *orange.* At the start of transposition the transposase cuts one of the two DNA strands at each end of the element, and the element then serves as a template for DNA synthesis, which begins by the addition of nucleotides to the 3′ ends of chromosomal DNA sequences. Many details are known, but the process is too complex to be illustrated here.

Sarepta Exhibit 1040, page 123

Summary

Viruses are infectious particles that consist of a DNA or an RNA molecule (the viral genome) packaged in a protein capsid, which in the enveloped viruses is surrounded by a lipid-bilayer-based membrane. Both the structure of the viral genome and its mode of replication vary widely among viruses. A virus can multiply only inside a host cell, whose genetic mechanisms it subverts for its own reproduction. A common outcome of a viral infection is the lysis of the infected cell and release of infectious viral particles. In some cases, however, the viral chromosome instead integrates into a host-cell chromosome, where it is replicated as a provirus along with the host genome. Many viruses are thought to have evolved from plasmids, which are self-replicating DNA or RNA molecules that lack the ability to wrap themselves in a protein coat.

Transposable elements are DNA sequences that differ from viruses in being able to multiply only in their host cell and its progeny; like plasmids, they cannot exist stably outside of cells. Unlike plasmids, they normally replicate only as an integral part of a chromosome. Some transposable elements, however, are closely related to retroviruses and can move from place to place in the genome by the reverse transcription of an RNA intermediate. Although both viruses and transposable elements can be viewed as parasites, many of the DNA sequence rearrangements they cause are important for the evolution of cells and organisms.

References

General

Judson, H.F. The Eighth Day of Creation: Makers of the Revolution in Biology. New York: Simon & Schuster, 1979.

Lewin, B. Genes IV. Oxford, NY: Oxford University Press, 1990.

Stent, G.S.; Calendar, R. Genetics: An Introductory Narrative, 2nd ed. San Francisco: W.H. Freeman, 1978.

Watson, J.D.; Hopkins, N.H.; Roberts, J.W.; Steitz, J.A.; Weiner, A.M. Molecular Biology of the Gene, 4th ed. Menlo Park, CA: Benjamin-Cummings, 1987.

Cited

1. Chamberlin, M. Bacterial DNA-dependent RNA polymerases. In The Enzymes, 3rd ed., Vol. 15B (P. Boyer, ed.), pp. 61–108. New York: Academic Press, 1982.

   Gralla, J.D. Promoter recognition and mRNA initiation by *Escherichia coli* $E\sigma^{70}$. *Methods Enzymol.* 185:37–54, 1990.

   Kerppola, T.K.; Kane, C.M. RNA polymerase: regulation of transcript elongation and termination. *FASEB J.* 5(13):2833–2842, 1991.

   Sentenac, A. Eukaryotic RNA polymerases. *Crc Crit. Rev. Biochem.* 18(1):31–90, 1985.

2. Collado-Vides, J.; Magasanik, B.; Gralla, J.D. Control site location and transcriptional regulation in *Escherichia coli*. *Microbiol. Rev.* 55(3):371–394, 1991.

   Murphy, S.; Moorefield, B.; Pieler, T. Common mechanisms of promoter recognition by RNA polymerases II and III. *Trends Genet.* 5(4):122–126, 1989.

   Travers, A.A. Structure and function of *E. coli* promoter DNA. *Crc Crit. Rev. Biochem.* 22(3):181–219, 1987.

3. McClain, W.H. Transfer RNA identity. *FASEB J.* 7(1):72–78, 1993.

   Rich, A.; Kim, S.H. The three-dimensional structure of transfer RNA. *Sci. Am.* 238(1):52–62, 1978.

   Sprinzl, M.; Hartmann, T.; Weber, J.; Blank, J.; Zeidler, R. Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res.* 17 Suppl.:1–172, 1989.

4. Cavarelli, J.; Moras, D. Recognition of tRNAs by aminoacyl-tRNA synthetases. *FASEB J.* 7(1):79–86, 1993.

   Freist, W. Mechanisms of aminoacyl-tRNA synthetases: a critical consideration of recent results. *Biochemistry* 28(17):6787–6795, 1989.

   Schimmel, P. Aminoacyl tRNA synthetases: general scheme of structure-function relationships in the polypeptides and recognition of transfer RNAs. *Annu. Rev. Biochem.* 56:125–158, 1987.

5. Crick, F.H.C. The genetic code. III. *Sci. Am.* 215(4):55–62, 1966.

   The Genetic Code. *Cold Spring Harbor Symp. Quant. Biol.*, Vol. 31, 1966.

   Wong, J.T. Evolution of the genetic code. *Microbiol. Sci.* 5(6):174–181, 1988.

6. Brimacombe, R. RNA-protein interactions in the *Escherichia coli* ribosome. *Biochimie* 73(7–8):927–936, 1991.

   Stern, S.; Powers, T.; Changchien, L.-M.; Noller, H.F. RNA-protein interactions in 30S ribosomal subunits: folding and function of 16S rRNA. *Science* 244(4906):783–790, 1989.

   Yonath, A.; Wittman, H.G. Challenging the three-dimensional structure of ribosomes. *Trends Biochem. Sci.* 14(8):329–335, 1989.

7. Nierhaus, K.H. The allosteric three-site model for the ribosomal elongation cycle: features and future. *Biochemistry* 29(21):4997–5008, 1990.

   Noller, H.F. Peptidyl transferase: protein, ribonucleoprotein, or RNA? *J. Bacteriol.* 175(17):5297–5300, 1993.

   Watson, J.D. The involvement of RNA in the synthesis of proteins. *Science* 140:17–26, 1963.

8. Craigen, W.J.; Lee, C.C.; Caskey, C.T. Recent advances in peptide chain termination. *Mol. Microbiol.* 4(6):861–865, 1990.

   Tate, W.P.; Brown, C.M. Translational termination: "stop" for protein synthesis or "pause" for regulation of gene expression. *Biochemistry* 31(9):2443–2450, 1992.

9. Gualerzi, C.O.; Pon, C.L. Initiation of mRNA translation in prokaryotes. *Biochemistry* 29(25):5881–5889, 1990.

Sarepta Exhibit 1040, page 124

Hunt, T. The initiation of protein synthesis. *Trends Biochem. Sci.* 5:178–181, 1980.

Merrick, W.C. Overview: mechanism of translation initiation in eukaryotes. *Enzyme* 44(1–4):7–16, 1990.

10. Jacques, N.; Dreyfus, M. Translation initiation in *Escherichia coli*: old and new questions. *Mol. Microbiol.* 4(7):1063–1067, 1990.

Kozak, M. Comparison of initiation of protein synthesis in procaryotes, eucaryotes, and organelles. *Microbiol. Rev.* 47:1–45, 1983.

Kozak, M. Structural features in eukaryotic mRNAs that modulate the initiation of translation. *J. Biol. Chem.* 266(30):19867–19870, 1991.

Yoon, H.; Donohue, T.F. Control of translation initiation in *Saccharomyces cerevisiae*. *Mol. Microbiol.* 6(11):1413–1419, 1992.

11. Hesketh, J.E.; Pryme, I.F. Interaction between mRNA, ribosomes and the cytoskeleton. *Biochem. J.* 277(Pt 1):1–10, 1991.

Rich, A. Polyribosomes. *Sci. Am.* 209(6):44–53, 1963.

Ryazanov, A.G.; Ovchinnikov, L.P.; Spirin, A.S. Development of structural organization of protein-synthesizing machinery from prokaryotes to eukaryotes. *Biosystems* 20(3):275–288, 1987.

12. Hershey, J.W.B. Overview: phosphorylation and translation control. *Enzyme* 44(1–4):17–27, 1990.

Lindahl, L.; Hinnebusch, A. Diversity of mechanisms in the regulation of translation in prokaryotes and lower eukaryotes. *Curr. Opin. Genet. Dev.* 2(5):720–726, 1992.

Merrick, W.C. Mechanism and regulation of eukaryotic protein synthesis. *Microbiol. Rev.* 56(2):291–315, 1992.

13. Riis, B.; Rattan, S.I.; Clark, B.F.; Merrick, W.C. Eukaryotic protein elongation factors. *Trends Biochem. Sci.* 15(11):420–424, 1990.

Soll, D. The accuracy of aminoacylation—ensuring the fidelity of the genetic code. *Experientia* 46(11–12):1089–1096, 1990.

Thompson, R.C. EFTu provides an internal kinetic standard for translational accuracy. *Trends Biochem. Sci.* 13:91–93, 1988.

14. Jiménez, A. Inhibitors of translation. *Trends Biochem. Sci.* 1:28–30, 1988.

Lord, J.M.; Hartley, M.R.; Roberts, L.M. Ribosome inactivating proteins of plants. *Semin. Cell Biol.* 2(1):15–22, 1991.

Perentesis, J.P.; Miller, S.P.; Bodley, J.W. Protein toxin inhibitors of protein synthesis. *Biofactors* 3(3):173–184, 1992.

15. Campbell, J.H. An RNA replisome as the ancestor of the ribosome. *J. Mol. Evol.* 32(1):3–5, 1991.

Lake, J.A. Evolving ribosome structure: domains in archaebacteria, eubacteria, eocytes and eukaryotes. *Annu. Rev. Biochem.* 54:507–530, 1985.

Orgel, L.E. RNA catalysis and the origin of life. *J. Theor. Biol.* 123:127–149, 1983.

16. Barnes, D.E.; Lindahl, T.; Sedgwick, B. DNA repair. *Curr. Opin. Cell Biol.* 5(3):424–433, 1993.

Friedberg, E.C. DNA Repair. San Francisco: W.H. Freeman, 1985.

Wevrick, R.; Buchwald, M. Mammalian DNA-repair genes. *Curr. Opin. Genet. Dev.* 3(3):470–474, 1993.

17. Behe, M.J. Histone deletion mutants challenge the molecular clock hypothesis. *Trends Biochem. Sci.* 15(10):374–376, 1990.

Wilson, A.C.; Carlson, S.S.; White, T.J. Biochemical Evolution. *Annu. Rev. Biochem.* 46:573–639, 1977.

Wilson, A.C.; Ochman, H.; Prager, E.M. Molecular time scale for evolution. *Trends Genet.* 3:241–247, 1987.

18. Drake, J.W. Comparative rates of spontaneous mutation. *Nature* 221:1132, 1969.

Drake, J.W. Spontaneous mutation. *Annu. Rev. Genet.* 25:125–146, 1991.

19. Ohta, T.; Kimura, M. Functional organization of genetic material as a product of molecular evolution. *Nature* 233:118–119, 1971. (The mutation rate limits the maximum number of genes in an organism.)

Smith, K.C. Spontaneous mutagenesis: experimental, genetic and other factors. *Mutat. Res.* 277(2):139–162, 1992.

Vulliamy, T.; Mason, P.; Luzzatto, L. The molecular basis of glucose-6-phosphate dehydrogenase deficiency. *Trends Genet.* 8(4):138–143, 1992.

20. Loomis, W.F. Similarities in eukaryotic genomes. *Comp. Biochem. Physiol.* 95B(1):21–27, 1990.

21. Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* 362(6422):709–715, 1993.

Schrodinger, E. What Is Life? Cambridge, UK: Cambridge University Press, 1945.

22. Kornberg, A.; Baker, T.A. DNA Replication, 2nd ed. New York: W.H. Freeman, 1992. (Chapters 4 and 6 describe DNA polymerases; Chapter 9 describes DNA ligase; Chapter 21 covers the enzymology of DNA repair.)

23. Hoeijmakers, J.H. Nucleotide excision repair I: from *E. coli* to yeast. *Trends Genet.* 9(5):173–177, 1993.

Hoeijmakers, J.H. Nucleotide excision repair II: from yeast to mammals. *Trends Genet.* 9(6):211–217, 1993.

Sancar, A.Z.; Sancar, G.B. DNA repair enzymes. *Annu. Rev. Biochem.* 57:29–67, 1988.

24. Elledge, S.J.; Zhou, Z.; Allen, J.B. Ribonucleotide reductase: regulation, regulation, regulation. *Trends Biochem. Sci.* 17(3):119–123, 1992.

Walker, G.C. Inducible DNA repair systems. *Annu. Rev. Biochem.* 54:425–457, 1985.

Witkin, E.M. RecA protein in the SOS response: milestones and mysteries. *Biochimie* 73(2–3):133–141, 1991.

25. Perutz, M.F. Frequency of abnormal human haemoglobins caused by C→T transitions in CpG dinucleotides. *J. Mol. Biol.* 213(2):203– 206, 1990.

26. Baker, T.A.; Wickner, S.H. Genetics and enzymology of DNA replication in *Escherichia coli*. *Annu. Rev. Genet.* 26:447–477, 1992.

Kornberg, A.; Baker, T.A. DNA Replication, 2nd ed. New York: W.H. Freeman, 1992.

So, A.G.; Downey, K.M. Eukaryotic DNA replication. *Crit. Rev. Biochem. Mol. Biol.* 27(1–2):129–155, 1992.

27. Linn, S. How many pols does it take to replicate nuclear DNA? *Cell* 66(2):185–187, 1991.

Meselsohn, M.; Stahl, F.W. The replication of DNA in *E. coli*. *Proc. Natl. Acad. Sci. USA* 44:671–682, 1958.

Young, M.C.; Reddy, M.K.; von Hippel, P.H. Structure and function of the bacteriophage T4 DNA polymerase holoenzyme. *Biochemistry* 31(37):8675–8690, 1992.

28. Inman, R.B.; Schnos, M. Structure of branch points in replicating DNA: presence of single-stranded connections in lambda DNA branch points. *J. Mol. Biol.* 56:319–325, 1971.

Sarepta Exhibit 1040, page 125

Ogawa, T.; Okazaki, T. Discontinuous DNA replication. *Annu. Rev. Biochem.* 49:421–457, 1980.

Thommes, P.; Hubscher, U. Eukaryotic DNA replication. Enzymes and proteins acting at the fork. *Eur. J. Biochem.* 194(3):699–712, 1990.

29. Echols, H.; Goodman, M.F. Fidelity mechanisms in DNA replication. *Annu. Rev. Biochem.* 60:477–511, 1991.

Fersht, A.R. Enzymatic editing mechanisms in protein synthesis and DNA replication. *Trends Biochem. Sci.* 5:262–265, 1980.

Goodman, M.F.; Creighton, S.; Bloom, L.B.; Petruska, J. Biochemical basis of DNA replication fidelity. *Crit. Rev. Biochem. Mol. Biol.* 28(2):83–126, 1993.

30. Crouch, R.J. Ribonuclease H: from discovery to 3D structure. *New Biol.* 2(9):771–777, 1990.

Kaguni, L.S; Lehman, I.R. Eukaryotic DNA polymerase-primase: structure, mechanism and function. *Biochim. Biophys. Acta* 950(2):87–101, 1988.

Rowen, L.; Kornberg, A. Primase, the dnaG protein of *Escherichia coli*: an enzyme which starts DNA chains. *J. Biol. Chem.* 253:758–764, 1978.

31. Lohman, T.M. Helicase-catalyzed DNA unwinding. *J. Biol. Chem.* 268(4):2269–2272, 1993.

Meyer, R.R.; Laine, P.S. The single-stranded DNA-binding protein of *Escherichia coli*. *Microbiol. Rev.* 54(4):342–380, 1990.

Thommes, P.; Hubscher, U. Eukaryotic DNA helicases: essential enzymes for DNA transactions. *Chromosoma* 101(8):467–473, 1992.

32. Kong, X.-P.; Onrust, R.; O'Donnell, M.; Kuriyan, J. Three-dimensional structure of the beta subunit of *E. coli* DNA polymerase III holoenzyme: a sliding DNA clamp. *Cell* 69(3):425–437, 1992.

33. Alberts, B.M. Protein machines mediate the basic genetic processes. *Trends Genet.* 1:26–30, 1985.

Nossal, N.G. Protein-protein interactions at a DNA replication fork: bacteriophage T4 as a model. *FASEB J.* 6(3):871–878, 1992.

Stukenberg, P.T; Studwell-Vaughan, P.S.; O'Donnell, M. Mechanism of the sliding beta-clamp of DNA polymerase III holoenzyme. *J. Biol. Chem.* 266(17):11328–11334, 1991.

34. Barras, F.; Marinus, M.G. The great GATC: DNA methylation in *E. coli*. *Trends Genet.* 5(5):139–143, 1989.

Heywood, L.A.; Burke, J.F. Mismatch repair in mammalian cells. *Bioessays* 12(10):473–477, 1990.

Modrich, P. Methyl-directed DNA mismatch correction. *J. Biol. Chem.* 264(12):6597–6600, 1989.

35. Echols, H. Nucleoprotein structures initiating DNA replication, transcription, and site-specific recombination. *J. Biol. Chem.* 265(25):14697–14700, 1990.

Georgopoulos, C. The *E. coli* dnaA initiation protein, a protein for all seasons. *Trends Genet.* 5(10):319–321, 1989.

Salas, M. Protein-priming of DNA replication. *Annu. Rev. Biochem.* 60:39–71, 1991.

36. Drlica, K. Bacterial topoisomerases and the control of DNA supercoiling. *Trends Genet.* 6(12):433–437, 1990.

Sternglanz, R. DNA topoisomerases. *Curr. Opin. Cell Biol.* 1(3):533–535, 1989.

Wang, J.C. DNA topoisomerases: why so many? *J. Biol. Chem.* 266(11):6659–6662, 1991.

37. Gruss, C.; Sogo, J.M. Chromatin replication. *Bioessays* 14(1):1–8, 1992.

Wang, T.S.-F. Eukaryotic DNA polymerases. *Annu. Rev. Biochem.* 60:513–552, 1991.

38. Roeder, G.S. Chromosome synapsis and genetic recombination: their roles in meiotic chromosome segregation. *Trends Genet.* 6(12):385–389, 1990.

Sadowski, P.D. Site-specific genetic recombination: hops, flips and flops. *FASEB J.* 7(9):760–767, 1993.

Whitehouse, H.L.K. Genetic Recombination: Understanding the Mechanisms. New York: Wiley, 1982.

39. Camerini-Otero, R.D.; Hsieh, P. Parallel DNA triplexes, homologous recombination, and other homology-dependent DNA interactions. *Cell* 73(2):217–223, 1993.

Smith, G.R. Homologous recombination in *E. coli*: multiple pathways for multiple reasons. *Cell* 58(5):807–809, 1989.

West, S.C. Enzymes and molecular mechanisms of genetic recombination. *Annu. Rev. Biochem.* 61:603–640, 1992.

40. Lloyd, R.G.; Sharples, G.J. Genetic analysis of recombination in prokaryotes. *Curr. Opin. Genet. Dev.* 2(5):683–690, 1992.

Lohman, T.M. *Escherichia coli* DNA helicases: mechanisms of DNA unwinding. *Mol. Microbiol.* 6(1):5–14, 1992.

Weinstock, G.M. General recombination in *Escherichia coli*. In *Escherichia coli* and *Salmonella Typhimurium:* Cellular and Molecular Biology (F.C. Neidhardt, ed.), pp. 1034–1043. Washington, DC: American Society for Microbiology, 1987.

41. Bradley, S.G. DNA reassociation and base composition. *Society for Applied Bacteriology Symposium Series* 8:11–26, 1980.

Gotoh, O. Prediction of melting profiles and local helix stability for sequenced DNA. *Adv. Biophys.* 16:1–52, 1983.

Wetmur, J.G.; Davidson, N. Kinetics of renaturation of DNA. *J. Mol. Biol.* 31:349–370, 1968.

42. Eggleston, A.K.; Kowalczykowski, S.C. An overview of homologous pairing and DNA strand exchange proteins. *Biochimie* 73(2–3):163–176, 1991.

Kowalczykowski, S.C. Biochemical and biological function of *Escherichia coli* RecA protein: behavior of mutant RecA proteins. *Biochimie* 73(2–3):289–304, 1991.

Roca, A.I., Cox, M.M. The RecA protein: structure and function. *Crit. Rev. Biochem. Mol. Biol.* 25(6):415–456, 1990.

43. Holliday, R. The history of the DNA heteroduplex. *Bioessays* 12(3):133–142, 1990.

Kowalczykowski, S.C. Biochemistry of genetic recombination: energetics and mechanism of DNA strand exchange. *Annu. Rev. Biophys. Biophys. Chem.* 20:539–575, 1991.

Messelsohn, M.S.; Radding, C.M. A general model for genetic recombination. *Proc. Natl. Acad. Sci. USA* 72:358–361, 1975.

44. Fogel, S.; Mortimer, R.K.; Lusnak, K. Mechanisms of meiotic gene conversion or "wanderings on a foreign strand." In The Molecular Biology of the Yeast *Saccharomyces Cerevisiae*: Life Cycle and Inheritance (J.N. Strathern, ed.), pp. 289–340. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 1981.

Kobayashi, I. Mechanisms for gene conversion and homologous recombination: the double-strand break repair model and the successive half crossing-over model. *Adv. Biophys.* 28:81–133, 1992.

Sarepta Exhibit 1040, page 126

Kourilsky, P. Molecular mechanisms of gene conversion in higher cells. *Trends Genet.* 2:60–62, 1986.

45. Radman, M. Mismatch repair and the fidelity of genetic recombination. *Genome* 31(1):68–73, 1989.

Rayssiguier, C.; Thaler, D.S.; Radman, M. The barrier to recombination between *Escherichia coli* and *Salmonella typhimurium* is disrupted in mismatch repair mutants. *Nature* 342(6248):396–401, 1989.

46. Landy, A. Dynamic, structural, and regulatory aspects of lambda site-specific recombination. *Annu. Rev. Biochem.* 58:913–949, 1989.

O'Gorman, S.; Fox, D.T.; Wahl, G.M. Recombinase-mediated gene activation and site-specific integration in mammalian cells. *Science* 251(4999):1351–1355, 1991.

Stark, W.M.; Boocock, M.R.; Sherratt, D.J. Catalysis by site-specific recombinases. *Trends Genet.* 8(12):432–439, 1992.

47. Mizuuchi, K. Transpositional recombination: mechanistic insights from studies of mu and other elements. *Annu. Rev. Biochem.* 61:1011–1051, 1992.

48. Borg, D.E.; Howe, M.M., ed. Mobile DNA. Washington, DC: American Society for Microbiology, 1989.

Joklik, W.K., ed. Virology, 2nd ed. Norwalk, CT: Appleton & Lange, 1985.

Levine, A. Viruses. New York: Scientific American Library, 1992.

49. Hershey, A.D.; Chase, M. Independent functions of viral protein and nucleic acid in growth of bacteriophage. *J. Gen. Physiol.* 36:39–56, 1952.

Brock, T.D. The Emergence of Bacterial Genetics. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 1990. (Chapter 6 provides a historical overview of bacteriophage research.)

50. Fields, B.N., ed. Virology. New York: Raven Press, 1985. (Chapters 3 and 4 discuss virus capsid structure and viral membranes, respectively.)

Simons, K.; Garoff, H.; Helenius, A. How an animal virus gets into and out its host cell. *Sci. Am.* 246(2):58–66, 1982.

51. Fields, B.N., ed. Virology. New York: Raven Press, 1985. (Chapter 5 summarizes types of viral genomes.)

Gierer, A.; Schramm, G. Infectivity of ribonucleic acid from tobacco mosaic virus. *Nature* 177:702–703, 1956.

Strauss, E.G.; Strauss, J.H. RNA viruses: genome structure and evolution. *Curr. Opin. Genet. Dev.* 1(4):485–493, 1991.

52. Borowiec, J.A.; Dean, F.B.; Bullock, P.A.; Hurwitz, J. Binding and unwinding—how T antigen engages the SV40 origin of DNA replication. *Cell* 60(2):181–184, 1990.

Carlson, K.; Overvatn, A. Bacteriophage T4 endonucleases II and IV, oppositely affected by dCMP hydroxymethylase activity, have different roles in the degradation and in the RNA polymerase-dependent replication of T4 cytosine containing DNA. *Genetics* 114(3):669–685, 1986.

Cohen, S.S. Virus-induced Enzymes. New York: Columbia University Press, 1968.

53. David, C.; Gargouri-Bouzid, R.; Haenni, A.-L. RNA replication of plant viruses containing an RNA genome. *Prog. Nucleic Acid Res. Mol. Biol.* 42:157–227, 1992.

Kornberg, A.; Baker, T.A. DNA Replication, 2nd ed. New York: W.H. Freeman, 1992. (Chapters 17 and 19 describe the replication of DNA phages and animal viruses.)

54. Kielian, M.; Jungerwirth, S. Mechanisms of enveloped virus entry into cells. *Mol. Biol. Med.* 7(1):17–31, 1990.

White, J.M. Membrane fusion. *Science* 258(5084):917–924, 1992.

Zhao, H.; Garoff, H. Role of cell surface spikes in alphavirus budding. *J. Virol.* 66(12):7089–7095, 1992.

55. Griffiths, G.; Rottier, P. Cell biology of viruses that assemble along the biosynthetic pathway. *Semin. Cell Biol.* 3(5):367–381, 1992.

56. Campbell, A.M. Thirty years ago in genetics: prophage insertion into bacterial chromosomes. *Genetics* 133(3):433–438, 1993.

Friedman, D.I. Interaction between bacteriophage lambda and its *Escherichia coli* host. *Curr. Opin. Genet. Dev.* 2(5):727–738, 1992.

57. Tooze, J. DNA Tumor Viruses. Molecular Biology of Tumor Viruses, 2nd ed., Part 2. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 1980. (Chapter 4 describes transformation by SV40 and polyoma viruses.)

zur Hausen, H. Viruses in human cancers. *Science* 254(5035):1167–1173, 1991.

58. Baltimore, D. Viral RNA-dependent DNA polymerase. *Nature* 226:1209–1211, 1970.

Varmus, H. Retroviruses. *Science* 240:1427–1435, 1988.

Whitcomb, J.M.; Hughes, S.H. Retroviral reverse transcription and integration: progress and problems. *Annu. Rev. Cell Biol.* 8:275–306, 1992.

59. Gallo, R.C.; Montagnier, L. The chronology of AIDS research. *Nature* 326(6112):435–436, 1987.

60. Boeke, J.D.; Chapman, K.B. Retrotransposition mechanisms. *Curr. Opin. Cell Biol.* 3(3):502–507, 1991.

Corces, V.G.; Geyer, P.K. Interactions of retrotransposons with the host genome: the case of the gypsy element of *Drosophila*. *Trends Genet.* 7(3):86–90, 1991.

Sandmeyer, S.B. Yeast retrotransposons. *Curr. Opin. Genet. Dev.* 2(5):705–711, 1992.

61. Gierl, A.; Frey, M. Eukaryotic transposable elements with short terminal inverted repeats. *Curr. Opin. Genet. Dev.* 1(4):494–497, 1991.

Haniford, D.B.; Chaconas, G. Mechanistic aspects of DNA transposition. *Curr. Opin. Genet. Dev.* 2(5):698–704, 1992.

Mizuuchi, K. Mechanism of transposition of bacteriophage mu: polarity of the strand transfer reaction at the initiation of transposition. *Cell* 39:395–404, 1984.

62. Levine, A. Viruses. New York: Scientific American Library, 1992. (Chapter 10 discusses the evolution of viruses.)

Novick, R.P. Plasmids. *Sci. Am.* 243(6):102–127, 1980.

Symons, R.H. The intriguing viroids and virusoids: what is their information content and how did they evolve? *Mol. Plant-Microbe Interact.* 4(2):111–121, 1991.

Sarepta Exhibit 1040, page 127