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

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.

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

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.)

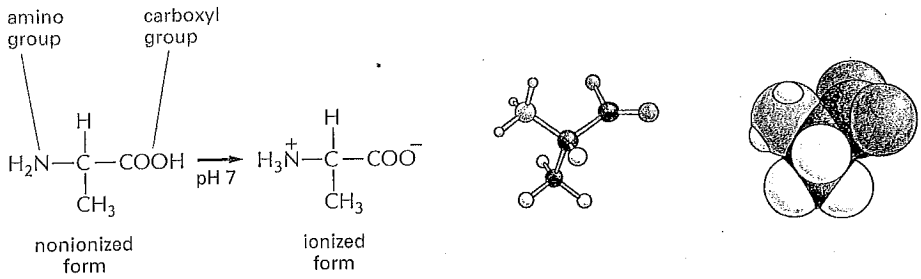


Figure 2–6 The amino acid alanine. In the cell, where the pH is close to 7, the free amino acid exists in its ionized form; but when it is incorporated into a polypeptide chain, the charges on the amino and carboxyl groups disappear. A ball-and-stick model and a space-filling model are shown to the right of the structural formulas. For alanine, the side chain is a $-CH_3$ group.

the hydrophobic tail regions pack together very closely facing the air and the hydrophilic head groups are in contact with the water (Panel 2–4, pp. 54–55). Two such films can combine tail to tail in water to make a phospholipid sandwich, or **lipid bilayer,** an extremely important assembly that is the structural basis of all cell membranes (discussed in Chapter 10).

Amino Acids Are the Subunits of Proteins [5]

The common amino acids are chemically varied, but they all contain a carboxylic acid group and an amino group, both linked to a single carbon atom (called the α-carbon; Figure 2–6). They serve as subunits in the synthesis of **proteins,** which are long linear polymers of amino acids joined head to tail by a *peptide bond* between the carboxylic acid group of one amino acid and the amino group of the next (Figure 2–7). Although there are many different possible amino acids, only 20 are common in proteins, each with a different *side chain* attached to the α-carbon atom (Panel 2–5, pp. 56–57). The same 20 amino acids occur over and over again in all proteins, including those made by bacteria, plants, and animals. Although the choice of these particular 20 amino acids probably occurred by chance in the course of evolution, the chemical versatility they provide is vitally important. For example, 5 of the 20 amino acids have side chains that can carry a charge (Figure 2–8), whereas the others are uncharged but reactive in specific ways (Panel 2–5, pp. 56–57). As we shall see, the properties of the amino acid side chains, in aggregate, determine the properties of the proteins they constitute and underlie all of the diverse and sophisticated functions of proteins.

Nucleotides Are the Subunits of DNA and RNA [6]

In nucleotides one of several different nitrogen-containing ring compounds (often referred to as *bases* because they can combine with H^+ in acidic solutions) is linked to a five-carbon sugar (either *ribose* or *deoxyribose*) that carries a phosphate group. There is a strong family resemblance between the different nitrogen-containing rings found in nucleotides. *Cytosine* (C), *thymine* (T), and *uracil* (U) are called *pyrimidine* compounds because they are all simple derivatives of a six-membered pyrimidine ring; *guanine* (G) and *adenine* (A) are *purine* compounds, with a second five-membered ring fused to the six-membered ring. Each nucleotide is named by reference to the unique base that it contains (Panel 2–6, pp.58–59).

Nucleotides can act as carriers of chemical energy. The triphosphate ester of adenine, **ATP** (Figure 2–9), above all others, participates in the transfer of energy in hundreds of individual cellular reactions. Its terminal phosphate is added using energy from the oxidation of foodstuffs, and this phosphate can be split off readily by hydrolysis to release energy that drives energetically unfavorable biosynthetic reactions elsewhere in the cell. As we discuss later, other nucleotide derivatives serve as carriers for the transfer of particular chemical groups, such as hydrogen atoms or sugar residues, from one molecule to another. And a cyclic phosphate-

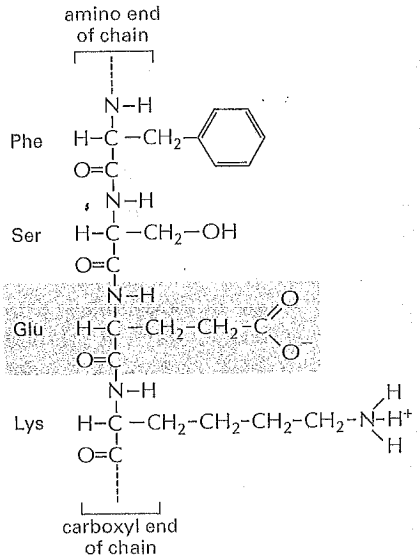


Figure 2–7 A small part of a protein molecule, showing four amino acids. Each amino acid is linked to the next by a covalent *peptide bond,* one of which is shaded *yellow.* A protein is therefore also sometimes referred to as a *polypeptide.* The amino acid *side chains* are shown in *red,* and the atoms of one amino acid (glutamic acid) are outlined by the *gray box.*

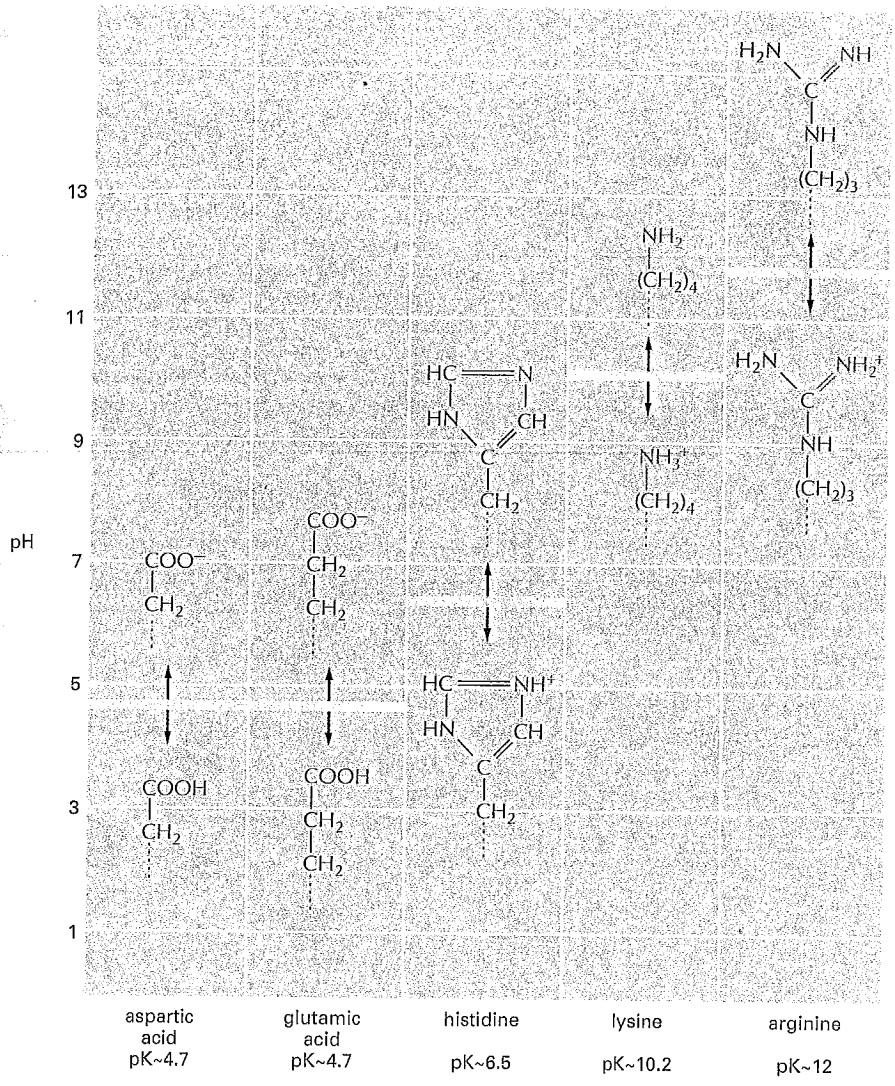


Figure 2–8 The charge on amino acid side chains depends on the pH. Carboxylic acids readily lose H^+ in aqueous solution to form a negatively charged ion, which is denoted by the suffix "-ate," as in aspart*ate* or glutam*ate*. A comparable situation exists for amines, which in aqueous solution take up H^+ to form a positively charged ion (which does not have a special name). These reactions are rapidly reversible, and the amounts of the two forms, charged and uncharged, depend on the pH of the solution. At a high pH, carboxylic acids tend to be charged and amines uncharged. At a low pH, the opposite is true—the carboxylic acids are uncharged and amines are charged. The pH at which exactly *half* of the carboxylic acid or amine residues are charged is known as the pK of that amino acid side chain.

In the cell the pH is close to 7, and almost all carboxylic acids and amines are in their fully charged form.

containing adenine derivative, *cyclic AMP,* serves as a universal signaling molecule within cells.

The special significance of nucleotides is in the storage of biological information. Nucleotides serve as building blocks for the construction of **nucleic acids,** long polymers in which nucleotide subunits are covalently linked by the formation of a phosphate ester between the 3′-hydroxyl group on the sugar residue of one nucleotide and the 5′-phosphate group on the next nucleotide (Fig-

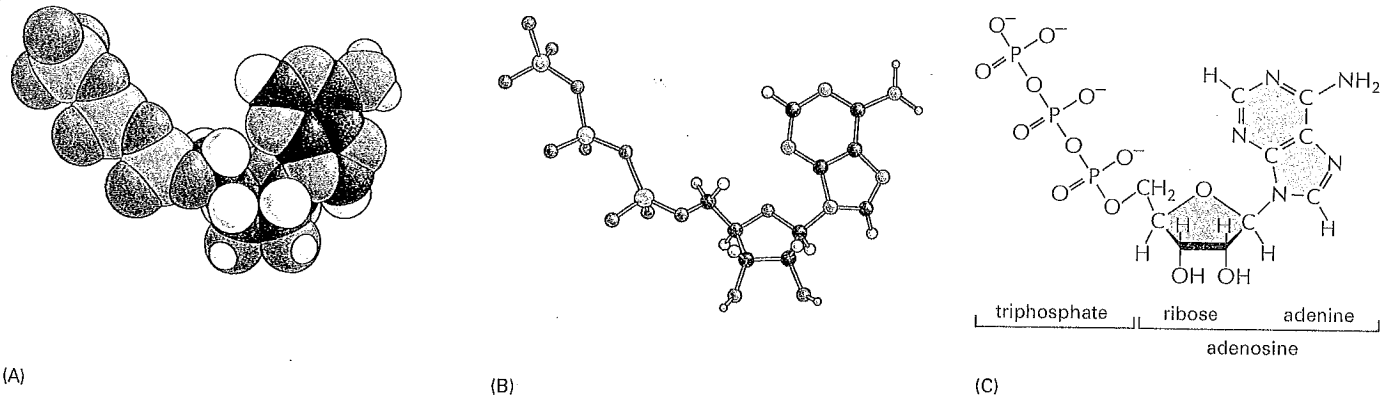
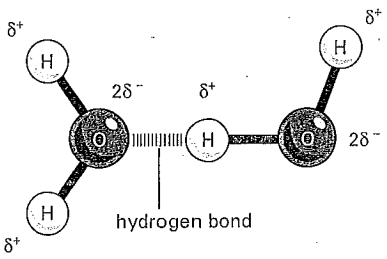


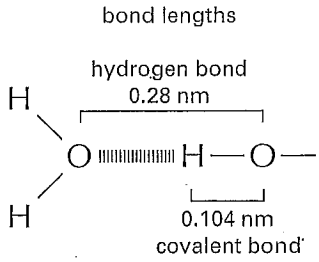
Figure 2–9 Chemical structure of adenosine triphosphate (ATP). A space-filling model (A), a ball-and-stick model (B), and the structural formula (C) are shown. Note the negative charges on each of the three phosphates.

HYDROGEN BONDS

Because they are polarized, two adjacent H_2O molecules can form a linkage known as a hydrogen bond. Hydrogen bonds have only about 1/20 the strength of a covalent bond.

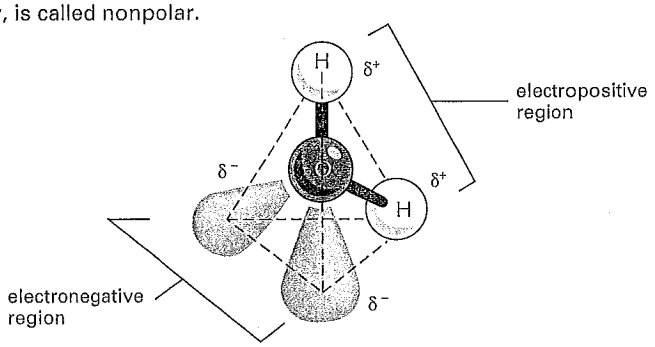
Hydrogen bonds are strongest when the three atoms lie in a straight line.





WATER

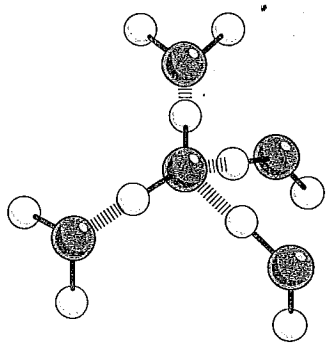
Two atoms, connected by a covalent bond, may exert different attractions for the electrons of the bond. In such cases the bond is dipolar, with one end slightly negatively charged (δ^-) and the other slightly positively charged (δ^+). A bond in which both atoms are the same, or in which they attract electrons equally, is called nonpolar.



Although a water molecule has an overall neutral charge (having the same number of electrons and protons), the electrons are asymmetrically distributed, which makes the molecule polar. The oxygen nucleus draws electrons away from the hydrogen nuclei, leaving these nuclei with a small net positive charge. The excess of electron density on the oxygen atom creates weakly negative regions at the other two corners of an imaginary tetrahedron.

WATER STRUCTURE

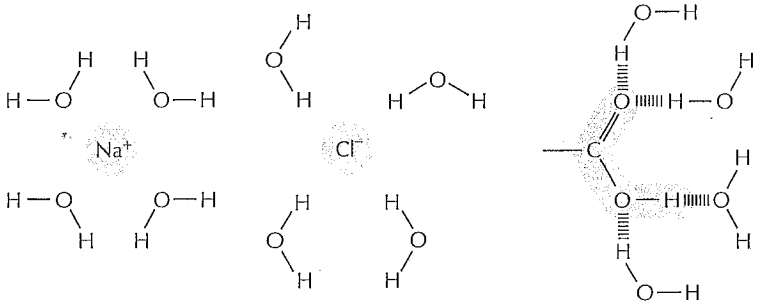
Molecules of water join together transiently in a hydrogen-bonded lattice. Even at 37°C, 15% of the water molecules are joined to four others in a short-lived assembly known as a "flickering cluster."



The cohesive nature of water is responsible for many of its unusual properties, such as high surface tension, specific heat, and heat of vaporization.

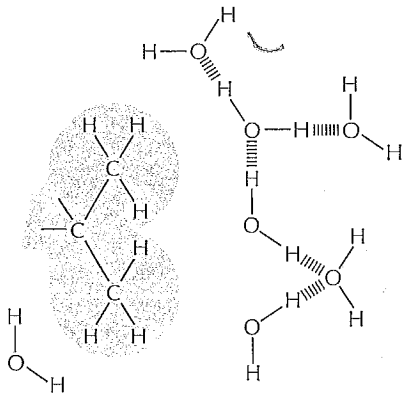
HYDROPHILIC AND HYDROPHOBIC MOLECULES

Because of the polar nature of water molecules, they will cluster around ions and other polar molecules.



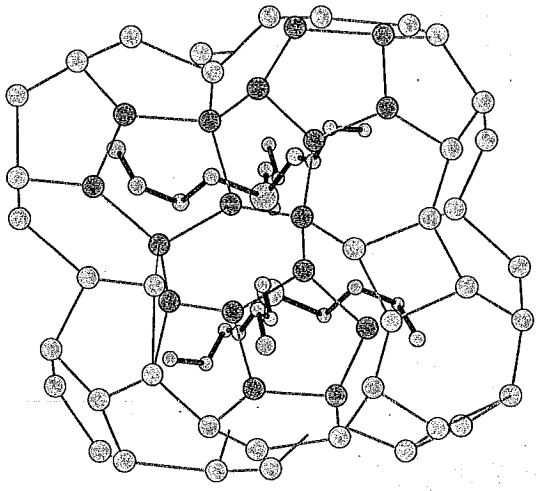
Molecules that can thereby be accommodated in water's hydrogen-bonded structures are hydrophilic and relatively water-soluble.

Nonpolar molecules interrupt the H-bonded structure of water without forming favorable interactions with water molecules. They are therefore hydrophobic and quite insoluble in water.



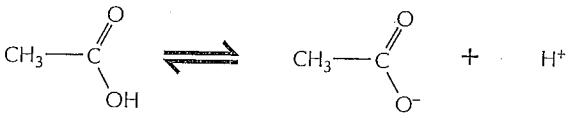
HYDROPHOBIC MOLECULES AND CLATHRATE WATER STRUCTURES

Molecules that are nonpolar and cannot form hydrogen bonds—such as hydrocarbons—have only limited solubility in water and are called hydrophobic. In water, ordered cages of water molecules are formed around hydrocarbons. These icelike cages, called "clathrate structures," are relatively more ordered than water and cause an entropy decrease of the mixture. Part of a clathrate cage (*red*) surrounding a hydrocarbon (*black*) is shown. In the intact cage, each oxygen atom (*red circles*) would be tetrahedrally coordinated to four others.



ACIDS AND BASES

An acid is a molecule that releases an H^+ ion (proton) in solution. For example,

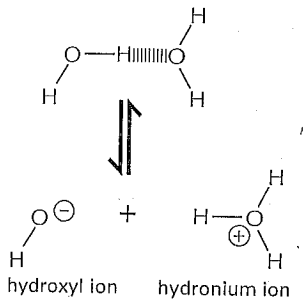


acid base proton

A base is a molecule that accepts an H^+ ion (proton) in solution. For example,

base proton acid

Water itself has a slight tendency to ionize and can act both as a weak acid and as a weak base. When it acts as an acid, it releases a proton to form a hydroxyl ion. When it acts as a base, it accepts a proton to form a hydronium ion. Most protons in aqueous solutions exist as hydronium ions.



hydroxyl ion hydronium ion

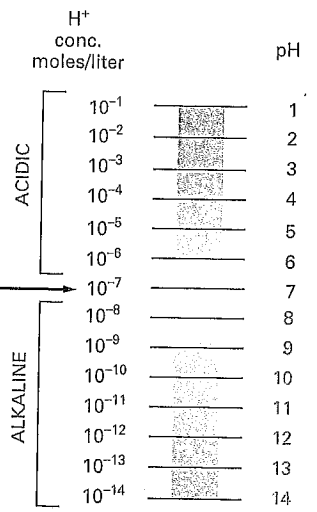
pH

The acidity of a solution is defined by the concentration of H^+ ions it possesses. For convenience we use the pH scale where

$$pH = -\log_{10}[H^+]$$

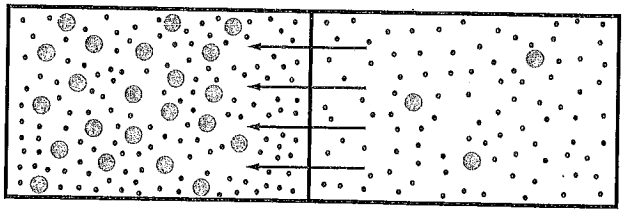
For pure water

$[H^+] = 10^{-7}$ moles/liter



OSMOSIS

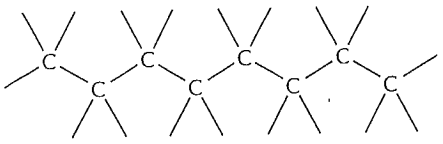
If two aqueous solutions are separated by a membrane that allows only water molecules to pass, water will move into the solution containing the greatest concentration of solute molecules by a process known as osmosis.



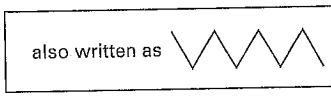
This movement of water from a hypotonic to a hypertonic solution can cause an increase in hydrostatic pressure in the hypertonic compartment. Two solutions that have identical solute concentrations and are therefore osmotically balanced are said to be isotonic.

CARBON SKELETONS

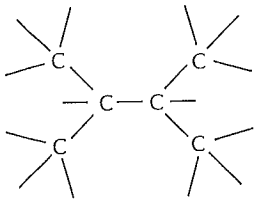
The unique role of carbon in the cell comes from its ability to form strong covalent bonds with other carbon atoms. Thus carbon atoms can join to form chains.



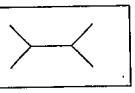
also written as



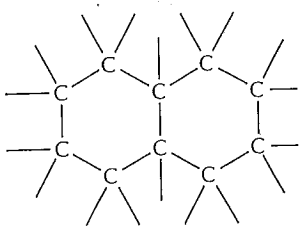
or branched trees



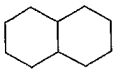
also written as



or rings



also written as

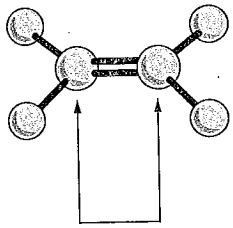


COVALENT BONDS

A covalent bond forms when two atoms come very close together and share one or more of their electrons. In a single bond one electron from each of the two atoms is shared; in a double bond a total of four electrons are shared.

Each atom forms a fixed number of covalent bonds in a defined spatial arrangement. For example, carbon forms four single bonds arranged tetrahedrally, whereas nitrogen forms three single bonds and oxygen forms two single bonds arranged as shown below.

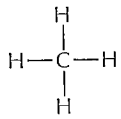
Double bonds exist and have a different spatial arrangement.



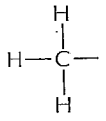
Atoms joined by two or more covalent bonds cannot rotate freely about the bond axis. This restriction is a major influence on the three-dimensional shape of many macromolecules.

HYDROCARBONS

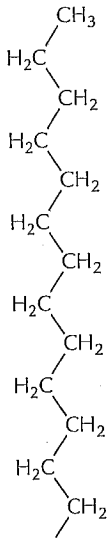
Carbon and hydrogen together make stable compounds called hydrocarbons. These are nonpolar, do not form hydrogen bonds, and are generally insoluble in water.



methane



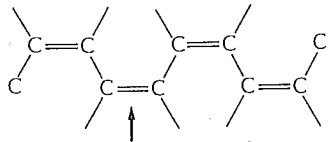
methyl group



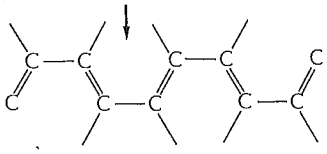
part of a fatty acid chain

RESONANCE AND AROMATICITY

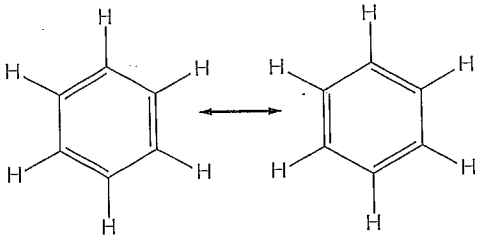
The carbon chain can include double bonds. If these are on alternate carbon atoms, the bonding electrons move within the molecule, stabilizing the structure by a phenomenon called resonance.



the truth is somewhere between these two structures



When resonance occurs throughout a ring compound, an aromatic ring is generated.

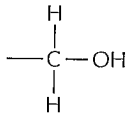


often written as

C—O COMPOUNDS

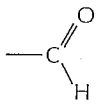
Many biological compounds contain a carbon bonded to an oxygen. For example,

alcohol

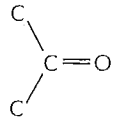


The —OH is called a hydroxyl group.

aldehyde

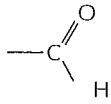


ketone



The C=O is called a carbonyl group.

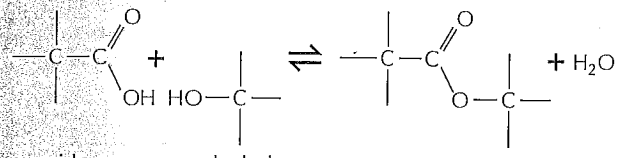
carboxylic acid



The —COOH is called a carboxyl group. In water this loses an H^+ ion to become $—COO^-$.

esters

Esters are formed by combining an acid and an alcohol.

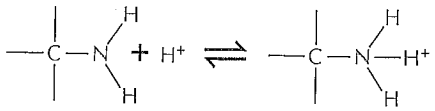


acid alcohol ester

C—N COMPOUNDS

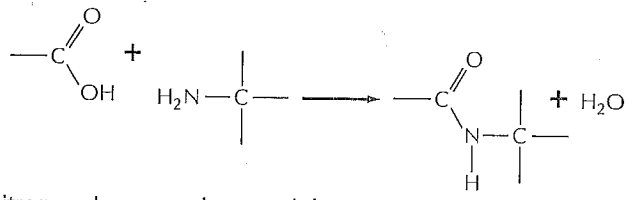
Amines and amides are two important examples of compounds containing a carbon linked to a nitrogen.

Amines in water combine with an H^+ ion to become positively charged.

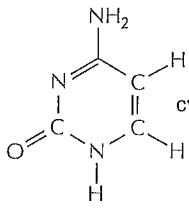


They are therefore basic.

Amides are formed by combining an acid and an amine. They are more stable than esters. Unlike amines, they are uncharged in water. An example is the peptide bond.



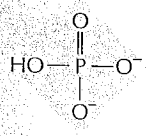
Nitrogen also occurs in several ring compounds, including important constituents of nucleic acids: purines and pyrimidines.



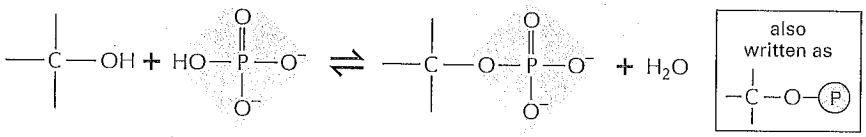
cytosine (a pyrimidine)

PHOSPHATES

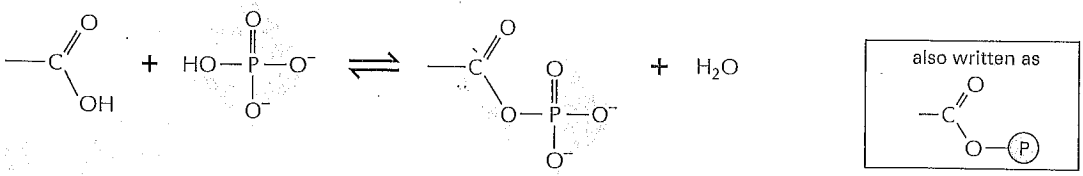
Inorganic phosphate is a stable ion formed from phosphoric acid, H_3PO_4. It is often written as P_i.

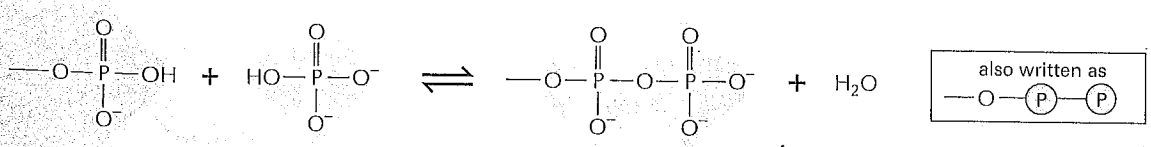


Phosphate esters can form between a phosphate and a free hydroxyl group.



The combination of a phosphate and a carboxyl group, or two or more phosphate groups, gives an acid anhydride.

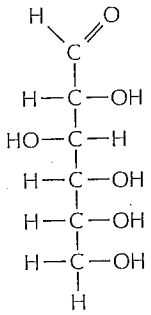




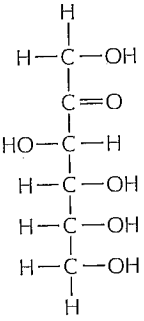
HEXOSES $n = 6$

Two common hexoses are

glucose

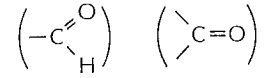


fructose

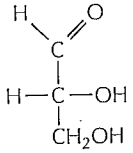


MONOSACCHARIDES

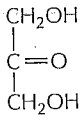
Monosaccharides are aldehydes or ketones



that also have two or more hydroxyl groups. Their general formula is $(CH_2O)n$. The simplest are trioses ($n = 3$) such as



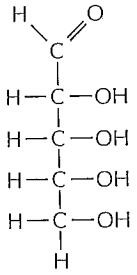
glyceraldehyde (an aldose)



dihydroxyacetone (a ketose)

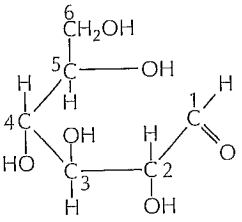
PENTOSES $n = 5$

A common pentose is



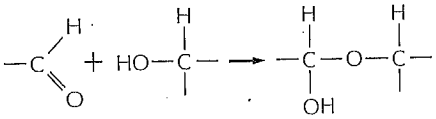
ribose

D-glucose (open-chain form)

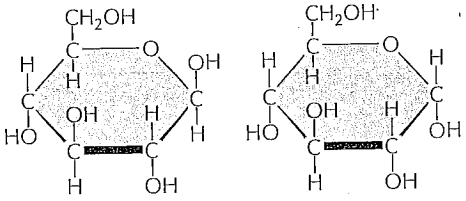


RING FORMATION

The aldehyde or ketone group of a sugar can react with a hydroxyl group



For the larger sugars ($n > 4$) this happens within the same molecule to form a 5- or 6-membered ring.



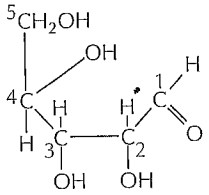
β-D-glucose α-D-glucose

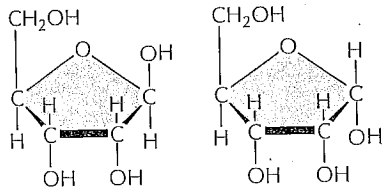
STEREOISOMERS

NUMBERING

The carbon atoms of a sugar are numbered from the end closest to the aldehyde or ketone.

D-ribose (open-chain form)



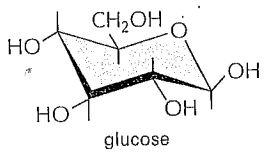


β-D-ribose α-D-ribose

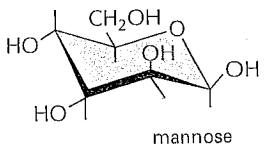
STEREOISOMERS

ISOMERS

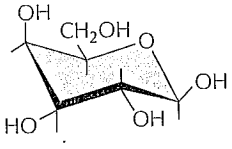
Monosaccharides have many isomers that differ only in the orientation of their hydroxyl groups—e.g., glucose, mannose, and galactose are isomers of each other.



glucose



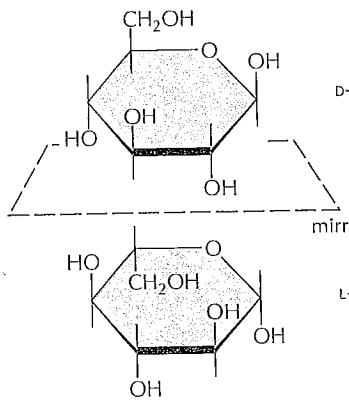
mannose



galactose

D AND L FORMS

Two isomers that are mirror images of each other have the same chemistry and therefore are given the same name and distinguished by the prefix D or L.



D-glucose

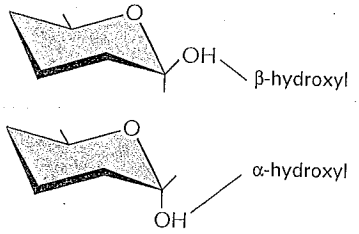
mirror plane

L-glucose

52 Panel 2–3 An outline of some of the types of sugars commonly found in cells.

α- AND β-LINKS

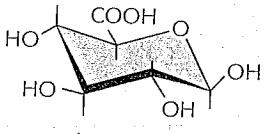
The hydroxyl group on the carbon that carries the aldehyde or ketone can rapidly change from one position to another. These two positions are called α- and β-.



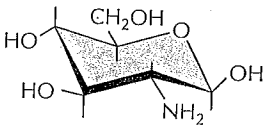
As soon as one sugar is linked to another, the α- or β-form is frozen.

SUGAR DERIVATIVES

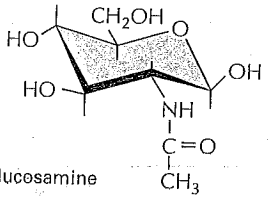
The hydroxyl groups of a simple monosaccharide can be replaced by other groups. For example,



D-glucuronic acid



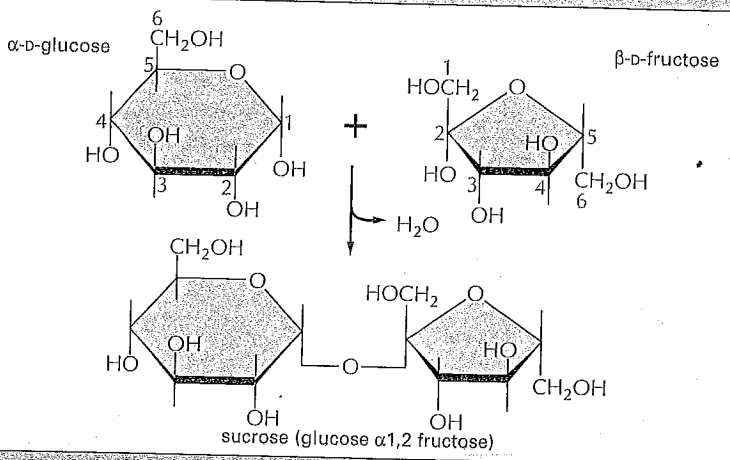
D-glucosamine



N-acetyl-D-glucosamine

DISACCHARIDES

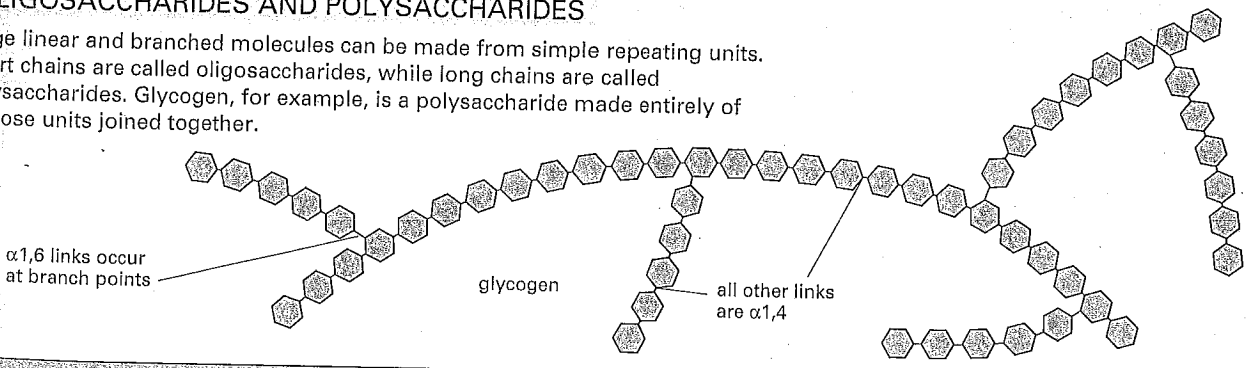
The carbon that carries the aldehyde or the ketone can react with any hydroxyl group on a second sugar molecule to form a glycosidic bond. Three common disaccharides are maltose (glucose α1,4 glucose), lactose (galactose β1,4 glucose), and sucrose (glucose α1,2 fructose). Sucrose is shown here.



sucrose (glucose α1,2 fructose)

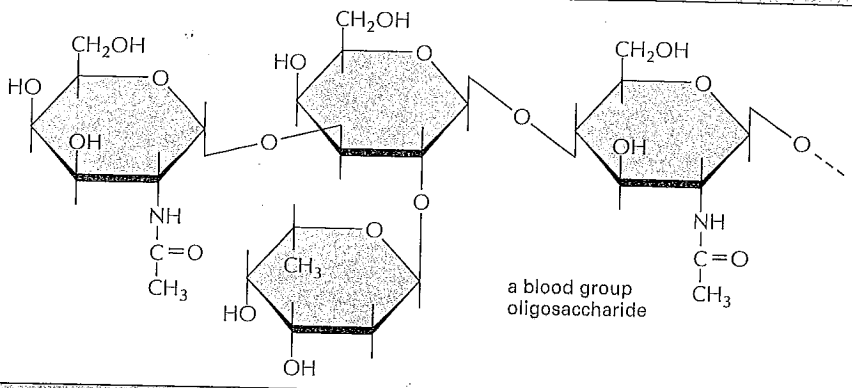
OLIGOSACCHARIDES AND POLYSACCHARIDES

Large linear and branched molecules can be made from simple repeating units. Short chains are called oligosaccharides, while long chains are called polysaccharides. Glycogen, for example, is a polysaccharide made entirely of glucose units joined together.



COMPLEX OLIGOSACCHARIDES

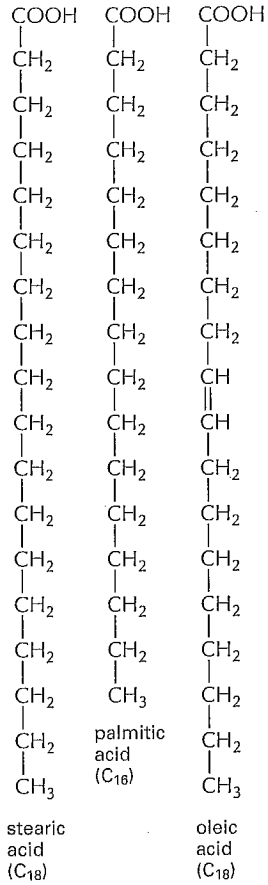
In many cases a sugar sequence is nonrepetitive. Many different molecules are possible. Such complex oligosaccharides are usually linked to proteins or to lipids.



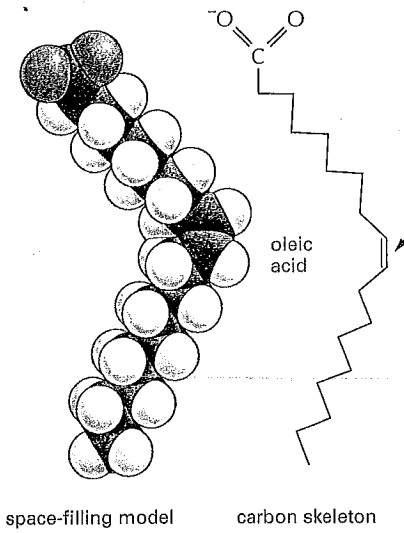
a blood group oligosaccharide

COMMON FATTY ACIDS

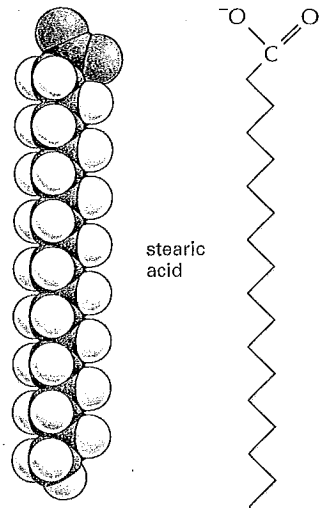
These are carboxylic acids with long hydrocarbon tails.



Hundreds of different kinds of fatty acids exist. Some have one or more double bonds and are said to be unsaturated.

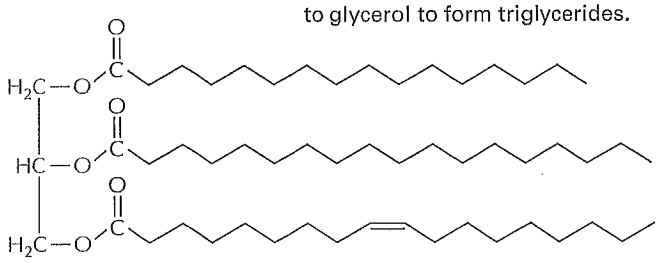


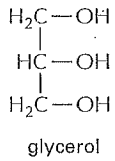
This double bond is rigid and creates a kink in the chain. The rest of the chain is free to rotate about the other C—C bonds.



TRIGLYCERIDES

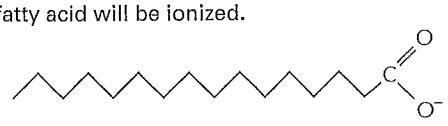
Fatty acids are stored as an energy reserve (fat) through an ester linkage to glycerol to form triglycerides.



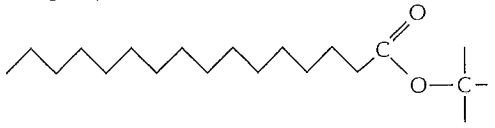


CARBOXYL GROUP

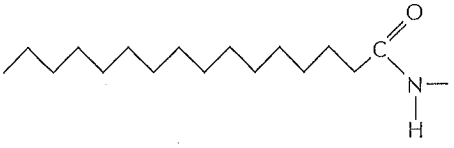
If free, the carboxyl group of a fatty acid will be ionized.



But more usually it is linked to other groups to form either esters

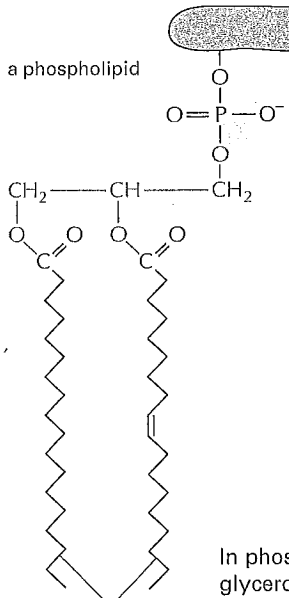


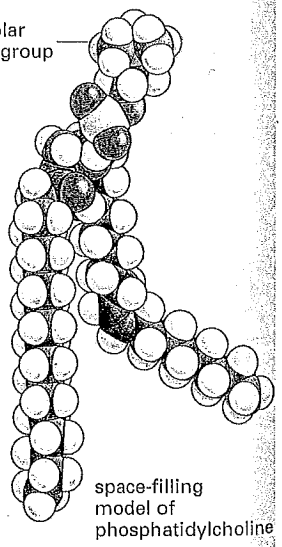
or amides.



PHOSPHOLIPIDS

Phospholipids are the major constituents of cell membranes.



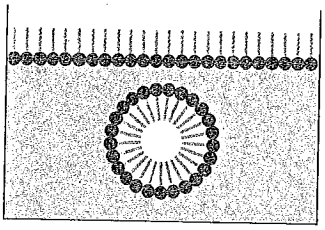


In phospholipids two of the —OH groups in glycerol are linked to fatty acids while the thi —OH group is linked to phosphoric acid. The phosphate is further linked to one of a variet of small polar head groups (alcohols).

54 Panel 2–4 An outline of some of the types of fatty acids commonly encountered in cells and the structures that they form.

LIPID AGGREGATES

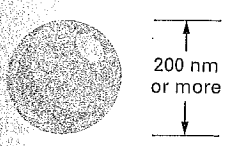
Fatty acids have a hydrophilic head and a hydrophobic tail.



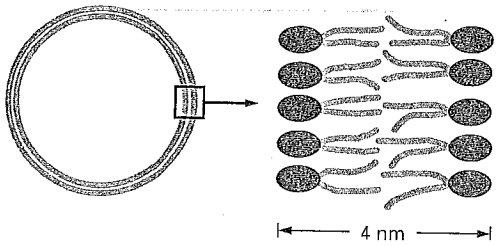
In water they can form a surface film or form small micelles.

Their derivatives can form larger aggregates held together by hydrophobic forces:

Triglycerides form large spherical fat droplets in the cell cytoplasm.

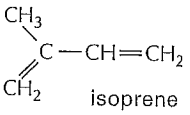


Phospholipids and glycolipids form self-sealing lipid bilayers that are the basis for all cellular membranes.



OTHER LIPIDS

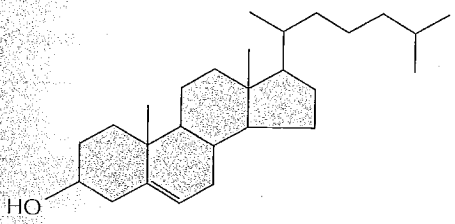
Lipids are defined as the water-insoluble molecules in cells that are soluble in organic solvents. Two other common types of lipids are steroids and polyisoprenoids. Both are made from isoprene units.



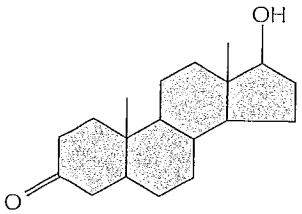
isoprene

STEROIDS

Steroids have a common multiple-ring structure.



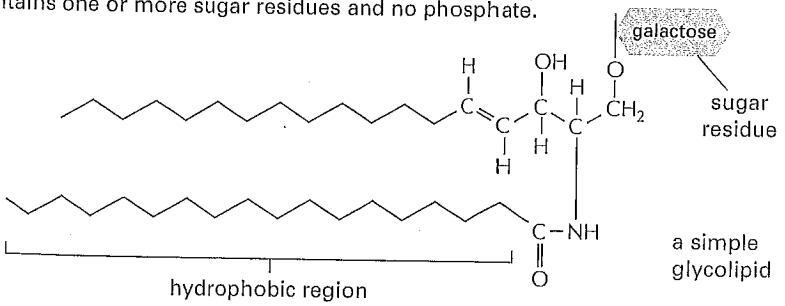
cholesterol—found in many membranes



testosterone—male steroid hormone

GLYCOLIPIDS

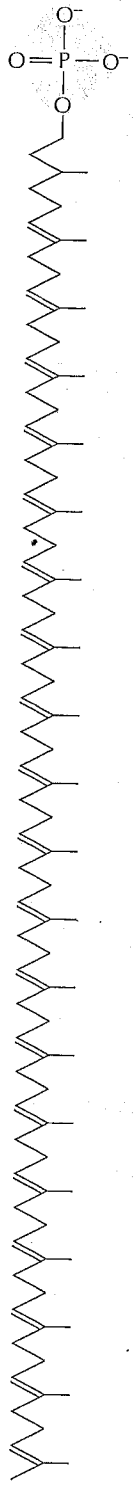
Like phospholipids, these compounds are composed of a hydrophobic region, containing two long hydrocarbon tails, and a polar region, which now contains one or more sugar residues and no phosphate.



a simple glycolipid

POLYISOPRENOIDS

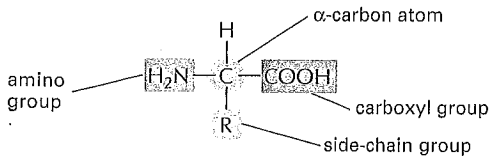
long chain polymers of isoprene



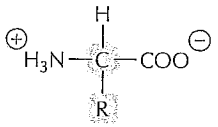
dolichol phosphate—used to carry activated sugars in the membrane-associated synthesis of glycoproteins and some polysaccharides

THE AMINO ACID

The general formula of an amino acid is

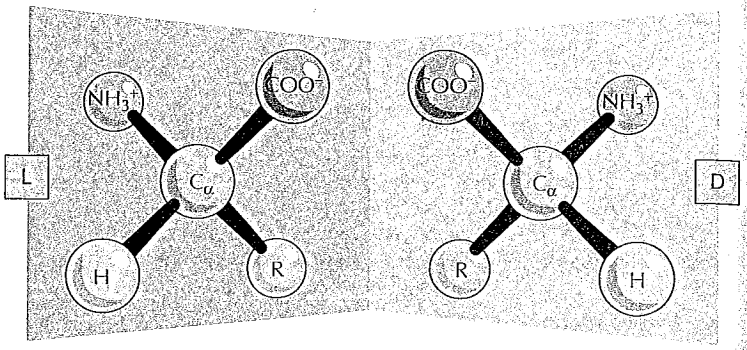


R is commonly one of 20 different side chains. At pH 7 both the amino and carboxyl groups are ionized.



OPTICAL ISOMERS

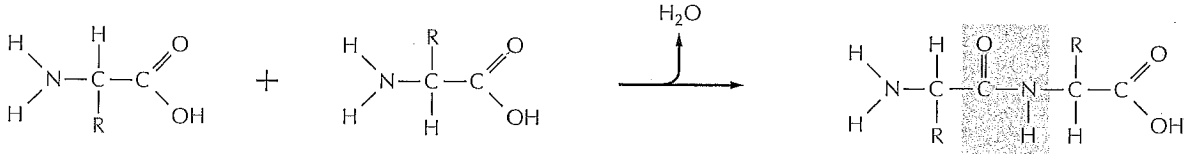
The α-carbon atom is asymmetric, which allows for two mirror image (or stereo-) isomers, D and L.



Proteins consist exclusively of L-amino acids.

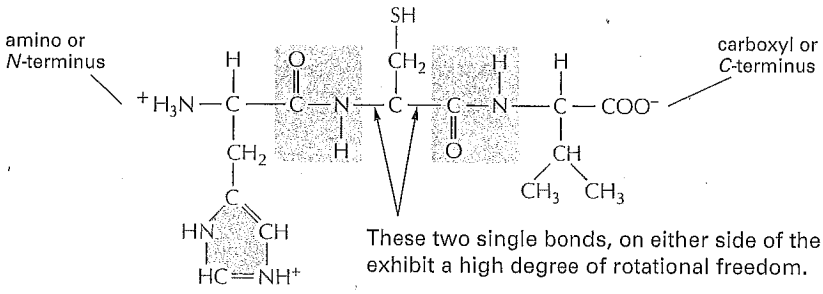
PEPTIDE BONDS

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

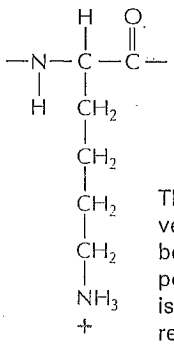
- acidic
- basic
- uncharged polar
- nonpolar

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

Thus: alanine = Ala = A

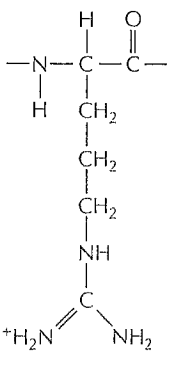
BASIC SIDE CHAINS

lysine
(Lys, or K)

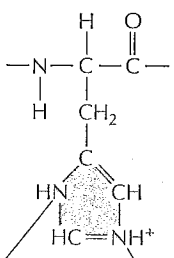


This group is very basic because its positive charge is stabilized by resonance.

arginine
(Arg, or R)



histidine
(His, or H)

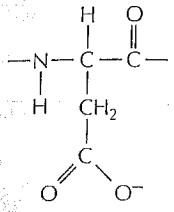


These nitrogens have a relatively weak affinity for an H^+ and are only partly positive at neutral pH.

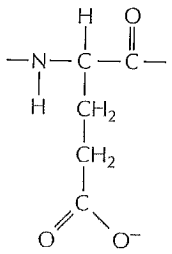
Panel 2–5 The 20 amino acids involved in the synthesis of proteins.

ACIDIC SIDE CHAINS

aspartic acid
(Asp, or D)



glutamic acid
(Glu, or E)



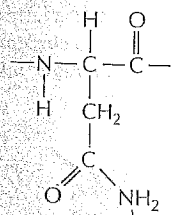
Amino acids with uncharged polar side chains are relatively hydrophilic and are usually on the outside of proteins, while the side chains on nonpolar amino acids tend to cluster together on the inside. Amino acids with basic or acidic side chains are very polar, and they are nearly always found on the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

The one letter code in alphabetical order:

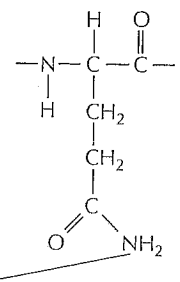
A = Ala	G = Gly	M = Met	S = Ser
C = Cys	H = His	N = Asn	T = Thr
D = Asp	I = Ile	P = Pro	V = Val
E = Glu	K = Lys	Q = Gln	W = Trp
F = Phe	L = Leu	R = Arg	Y = Tyr

UNCHARGED POLAR SIDE CHAINS

asparagine
(Asn, or N)

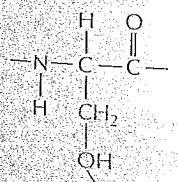


glutamine
(Gln, or Q)

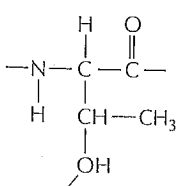


Although the amide N is not charged at neutral pH, it is polar.

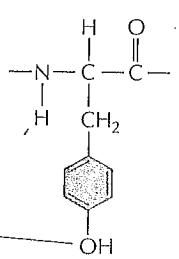
serine
(Ser, or S)



threonine
(Thr, or T)

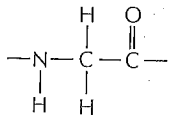


tyrosine
(Tyr, or Y)



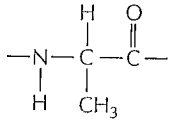
The —OH group is polar.

NONPOLAR SIDE CHAINS

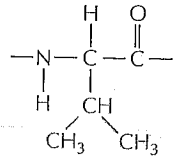


glycine
(Gly, or G)

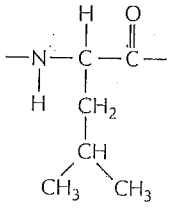
alanine
(Ala, or A)



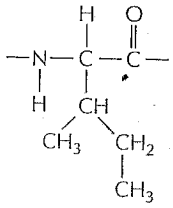
valine
(Val, or V)



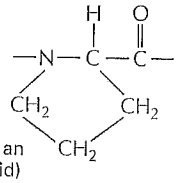
leucine
(Leu, or L)



isoleucine
(Ile, or I)

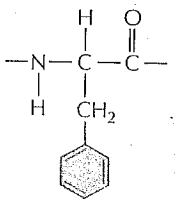


proline
(Pro, or P)

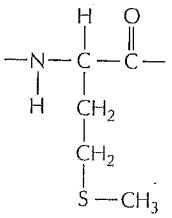


(actually an imino acid)

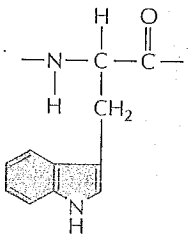
phenylalanine
(Phe, or F)

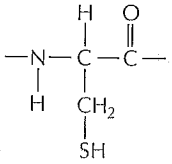


methionine
(Met, or M)



tryptophan
(Trp, or W)

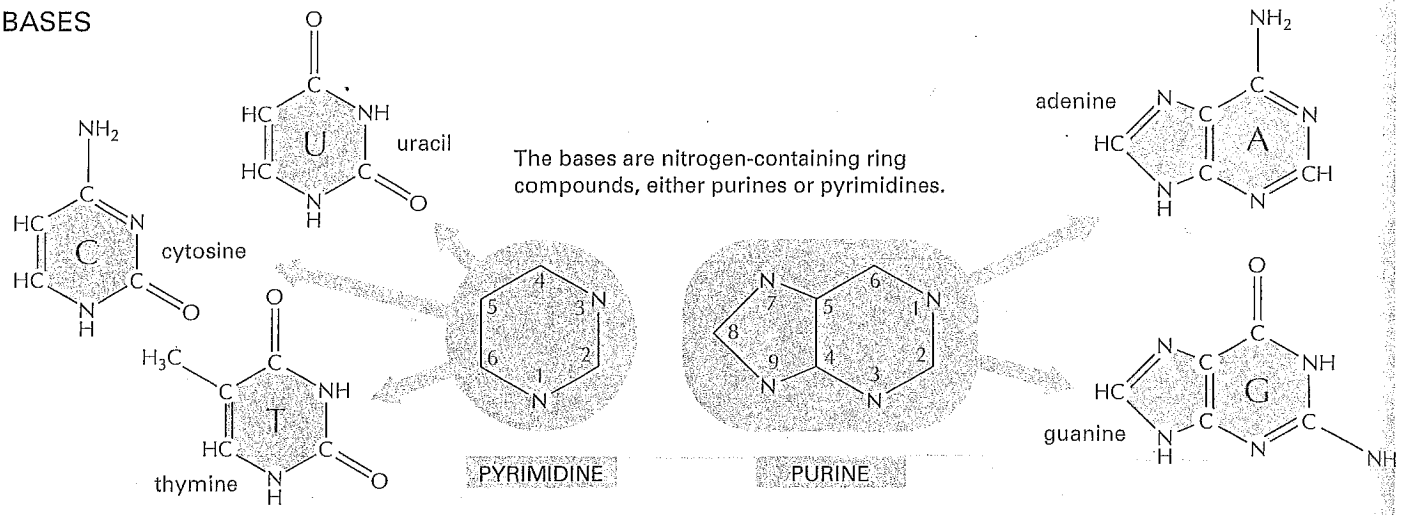




cysteine
(Cys, or C)

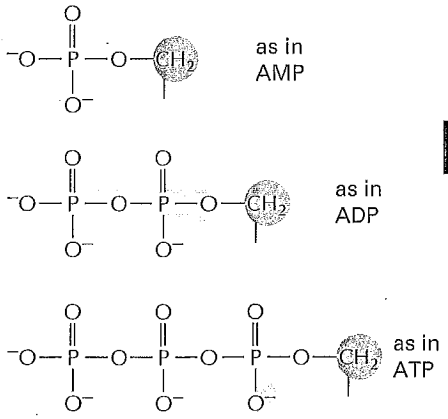
Paired cysteines allow disulfide bonds to form in proteins.

BASES



PHOSPHATES

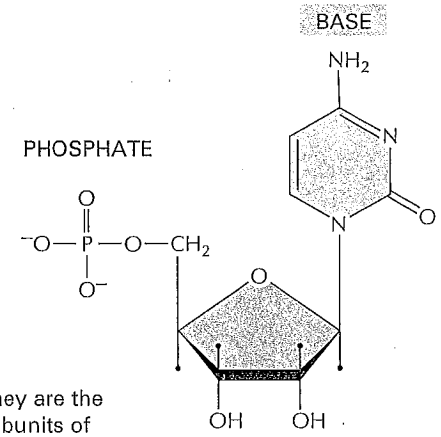
The phosphates are normally joined to the C5 hydroxyl of the ribose or deoxyribose sugar. Mono-, di-, and triphosphates are common.



The phosphate makes a nucleotide negatively charged.

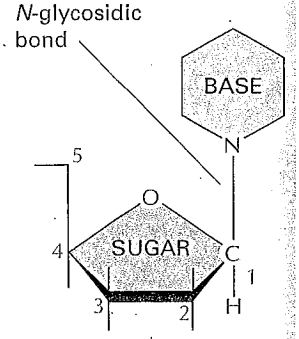
NUCLEOTIDES

A nucleotide consists of a nitrogen-containing base, a 5-carbon sugar, and one or more phosphate groups.



They are the subunits of the nucleic acids.

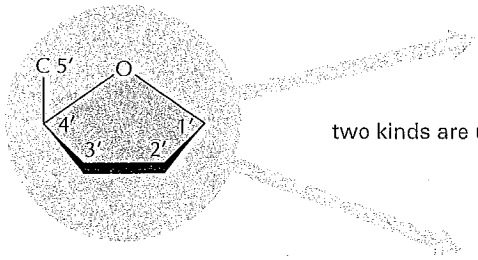
BASIC SUGAR LINKAGE



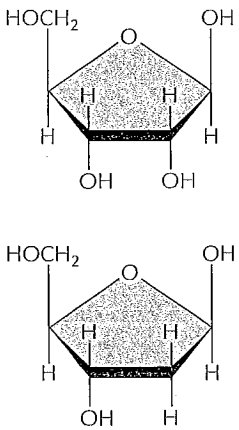
The base is linked to the same carbon (C1) used in sugar-sugar bonds.

SUGARS

PENTOSE
a 5-carbon sugar



two kinds are used



β-D-RIBOSE
used in ribonucleic acid

β-D-2-DEOXYRIBOSE
used in deoxyribonucleic acid

Each numbered carbon on the sugar of a nucleotide is followed by a prime mark; therefore, one speaks of the "5-prime carbon," etc.

58 Panel 2-6 A survey of the major types of nucleotides and their derivatives encountered in cells.

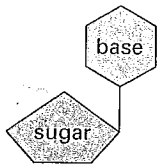
NOMENCLATURE

The names can be confusing, but the abbreviations are clear.

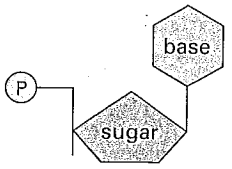
BASE	NUCLEOSIDE	ABBR.
adenine	adenosine	A
guanine	guanosine	G
cytosine	cytidine	C
uracil	uridine	U
thymine	thymidine	T

Nucleotides are abbreviated by three capital letters. Some examples follow:

AMP = adenosine monophosphate
dAMP = deoxyadenosine monophosphate
UDP = uridine diphosphate
ATP = adenosine triphosphate



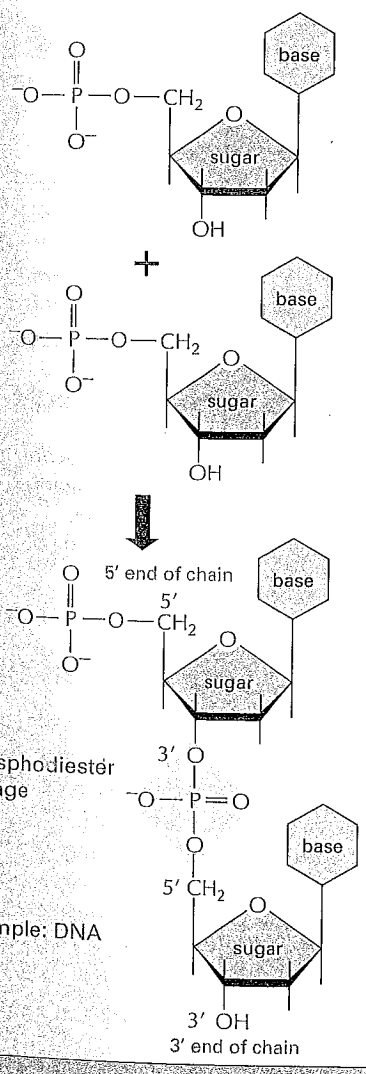
BASE + SUGAR = NUCLEOSIDE



BASE + SUGAR + PHOSPHATE = NUCLEOTIDE

NUCLEIC ACIDS

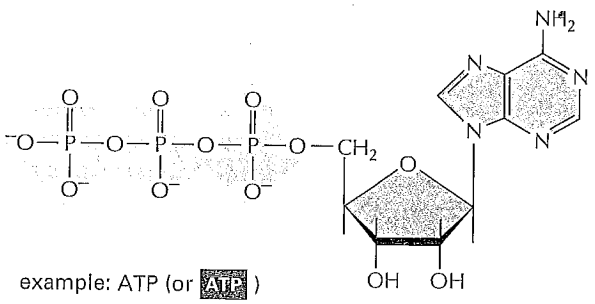
Nucleotides are joined together by a phosphodiester linkage between 5′ and 3′ carbon atoms to form nucleic acids. The linear sequence of nucleotides in a nucleic acid chain is commonly abbreviated by a one-letter code, A—G—C—T—T—A—C—A, with the 5′ end of the chain at the left.



example: DNA

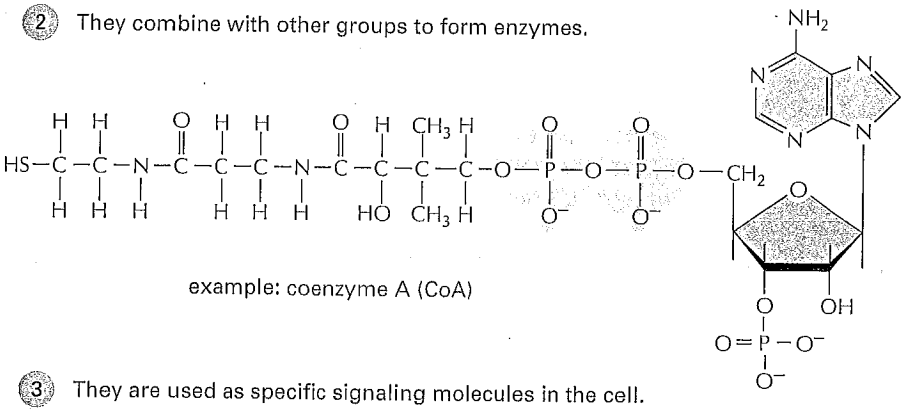
NUCLEOTIDES HAVE MANY OTHER FUNCTIONS

1. They carry chemical energy in their easily hydrolyzed acid-anhydride bonds.



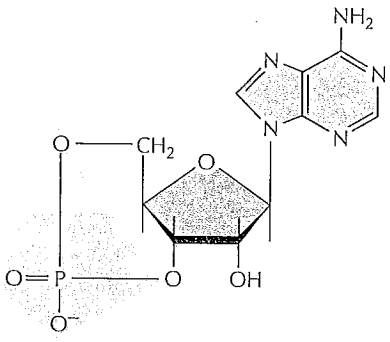
example: ATP (or ATP)

2. They combine with other groups to form enzymes.



example: coenzyme A (CoA)

3. They are used as specific signaling molecules in the cell.



example:
cyclic AMP (cAMP)

ure 2–10). There are two main types of nucleic acids, differing in the type of sugar that forms their polymeric backbone. Those based on the sugar *ribose* are known as **ribonucleic acids,** or **RNA,** and contain the four bases A, U, G, and C. Those based on *deoxyribose* (in which the hydroxyl at the 2′ position of ribose is replaced by a hydrogen) are known as **deoxyribonucleic acids,** or **DNA,** and contain the four bases A, T, G, and C. The sequence of bases in a DNA or RNA polymer represents the genetic information of the living cell. The ability of the bases from different nucleic acid molecules to recognize each other by noncovalent interactions (called **base-pairing**)—G with C, and A with either T (in DNA) or U (in RNA)—underlies all of heredity and evolution, as explained in Chapter 3.

Summary

Living organisms are autonomous, self-propagating chemical systems. They are made from a distinctive and restricted set of small carbon-based molecules that are essentially the same for every living species. The main categories are sugars, fatty acids, amino acids, and nucleotides. Sugars are a primary source of chemical energy for cells and can be incorporated into polysaccharides for energy storage. Fatty acids are also important for energy storage, but their most significant function is in the formation of cell membranes. Polymers consisting of amino acids constitute the remarkably diverse and versatile macromolecules known as proteins. Nucleotides play a central part in energy transfer and also are the subunits from which the informational macromolecules, RNA and DNA, are made.

Biological Order and Energy [7]

Cells must obey the laws of physics and chemistry. The rules of mechanics and of the conversion of one form of energy to another apply just as much to a cell as to a steam engine. There are, however, puzzling features of a cell that, at first sight, seem to place it in a special category. It is common experience that things left to themselves eventually become disordered: buildings crumble, dead organisms become oxidized, and so on. This general tendency is expressed in the *second law of thermodynamics,* which states that the degree of disorder in the universe (or in any isolated system in the universe) can only increase.

The puzzle is that living organisms maintain, at every level, a very high degree of order; and as they feed, develop, and grow, they appear to create this order out of raw materials that lack it. Order is strikingly apparent in large structures such as a butterfly wing or an octopus eye, in subcellular structures such as a mitochondrion or a cilium, and in the shape and arrangement of molecules from which these structures are built. The constituent atoms have been captured, ultimately, from a relatively disorganized state in the environment and locked together into a precise structure. Even a nongrowing cell requires constant ordering processes for survival since all of its organized structures are subject to spontaneous accidents and must be repaired continually. How is this possible thermodynamically? The answer is that the cell draws in fuel from its environment and releases heat as a waste product. The cell is therefore not an isolated system in the thermodynamic sense.

Biological Order Is Made Possible by the Release of Heat Energy from Cells [8]

As already mentioned, the second law of thermodynamics states that the amount of order in the universe (that is, in a cell plus its environment) must always decrease. Therefore, the continuous increase in order inside a living cell must be accompanied by an even greater increase in *dis*order in the cell's environment. Heat is energy in its most disordered form—the random commotion of mol-

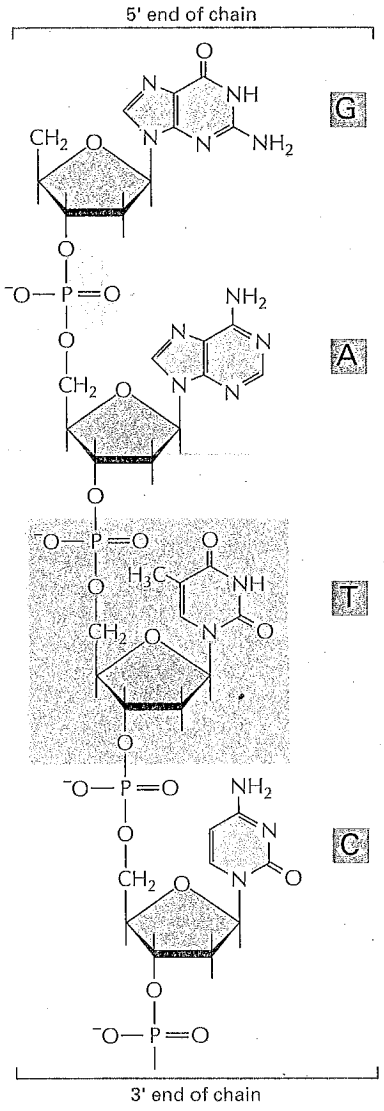


Figure 2–10 A short length of deoxyribonucleic acid (DNA), showing four nucleotides. One of the phosphodiester bonds that link adjacent nucleotides is shaded *yellow,* and one of the nucleotides is enclosed in a *gray box.* DNA and its close relative RNA are the nucleic acids of the cell.

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

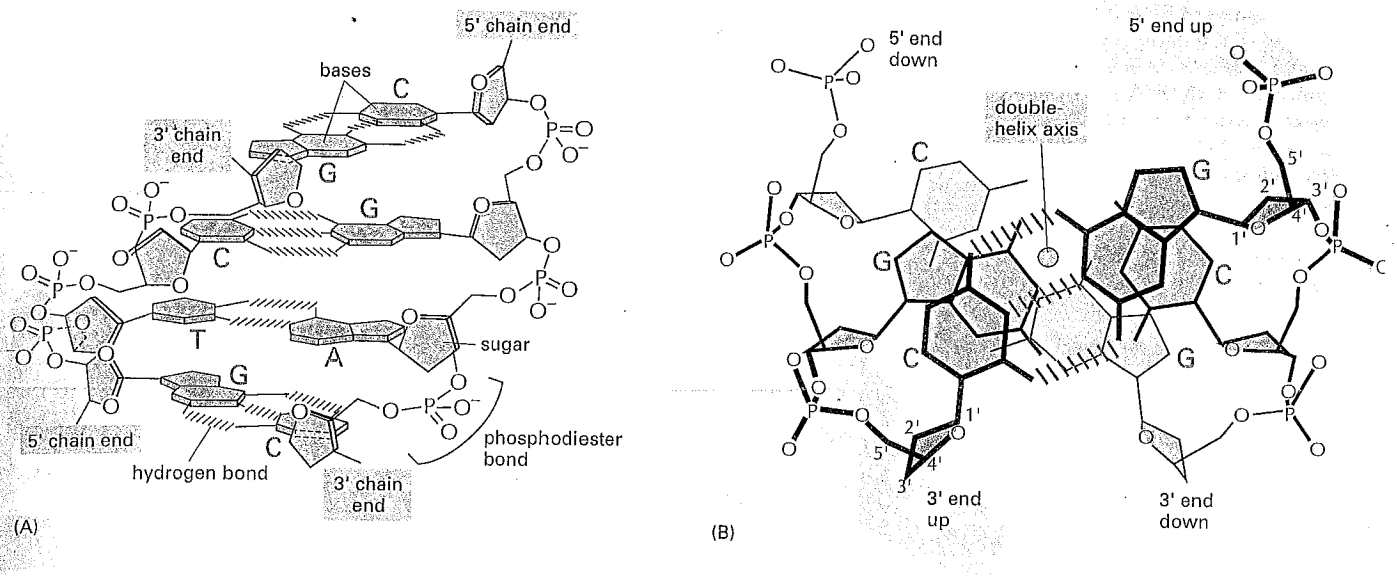


Figure 3–10 The DNA double helix. (A) A short section of the helix viewed from its side. Four complementary base pairs are shown. The bases are shown in *green,* while the deoxyribose sugars are *blue.* (B) The helix viewed from an end. Note that the two DNA strands run in opposite directions and that 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]

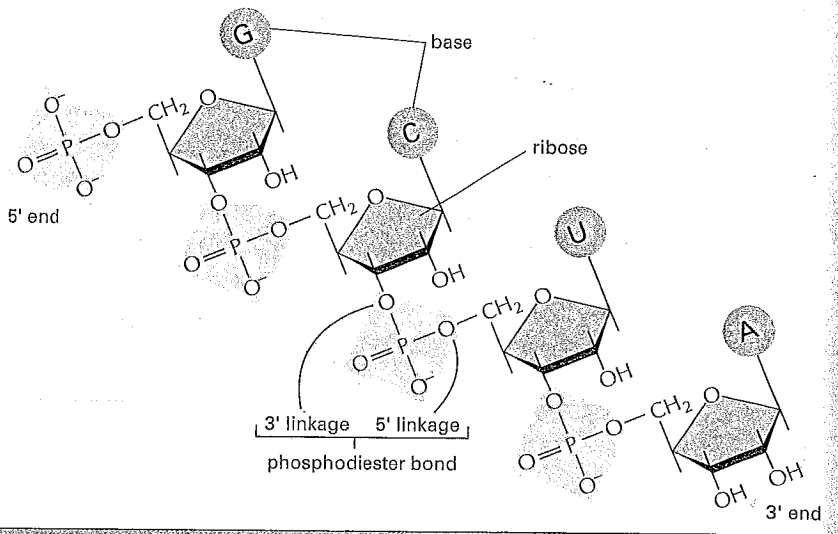
A gene carries biological information in a form that must be precisely copied and transmitted from each cell to all of its progeny. The implications of the discov-

DNA AND RNA

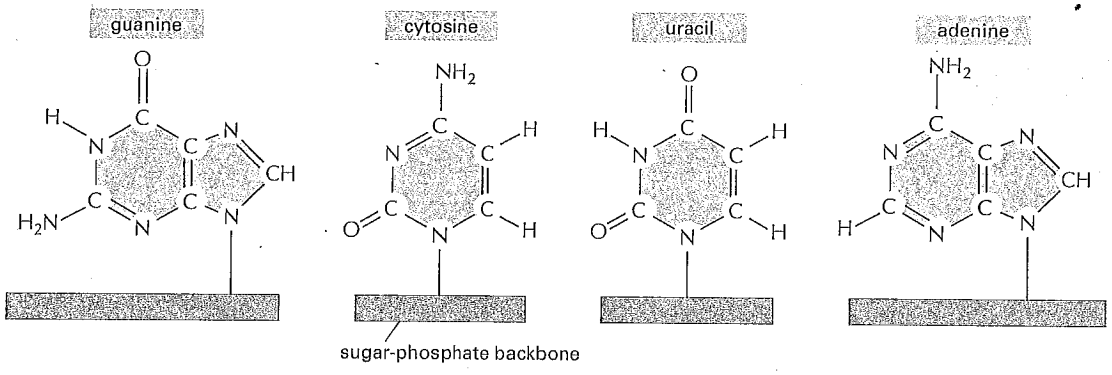
The structure of RNA is shown in this half of the panel, while the structure of DNA is shown in the other half. Both DNA and RNA are linear polymers of nucleotides (see Panel 2–6, pp. 58–59). RNA differs from DNA in three ways:

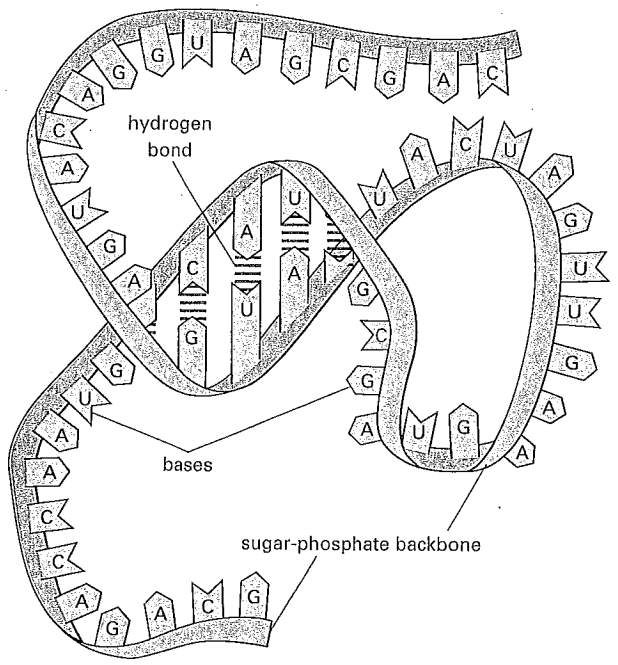
1. the sugar phosphate backbone contains ribose rather than deoxyribose
2. it contains the base uracil (U) instead of thymine (T)
3. it exists as a single strand rather than a double-stranded helix

SUGAR-PHOSPHATE BACKBONE OF RNA



FOUR BASES OF RNA

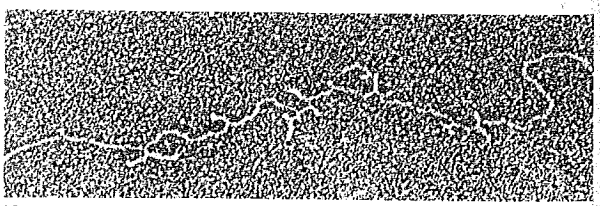




RNA SINGLE STRAND

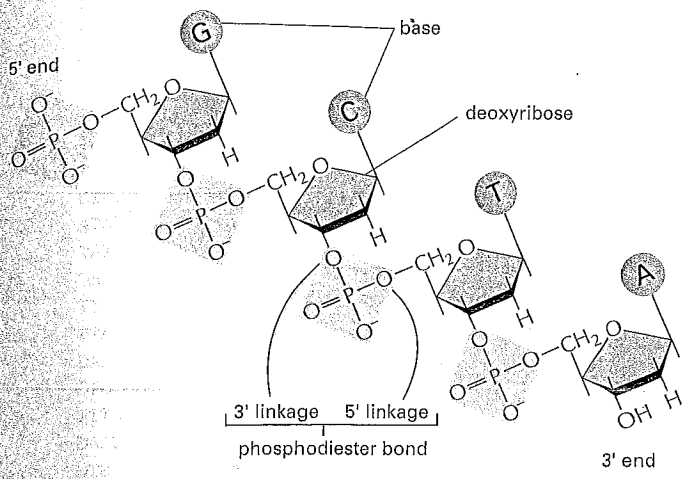
RNA is single-stranded, but it contains local regions of short complementary base-pairing that can form from a random matching process. Regions of base-pairing can be seen in the electron micrograph as branches off the stretched-out chain.

ELECTRON MICROGRAPH OF RNA

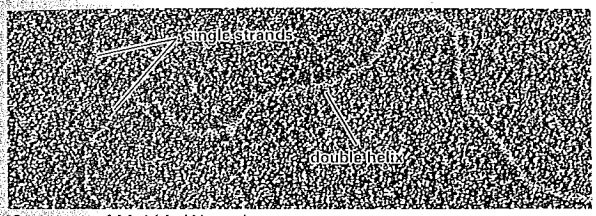


(Courtesy of Peter Wellauer.)

SUGAR-PHOSPHATE BACKBONE OF DNA

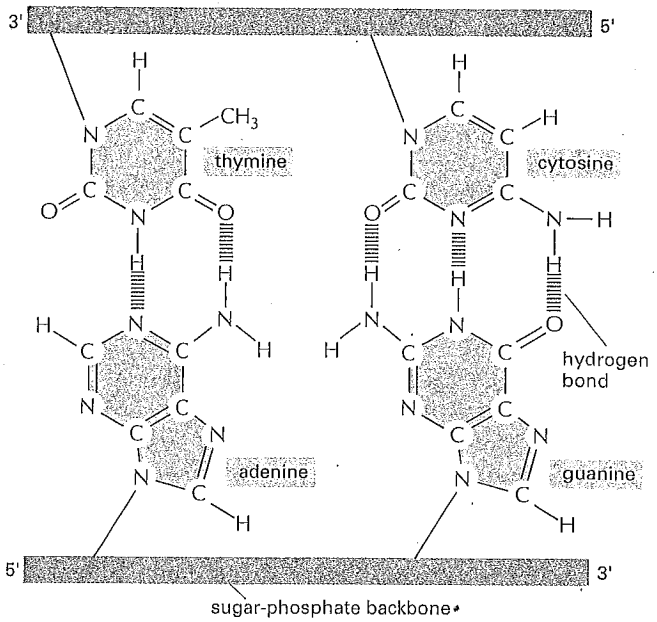


ELECTRON MICROGRAPH OF DNA



(Courtesy of Mei Lie Wong.)

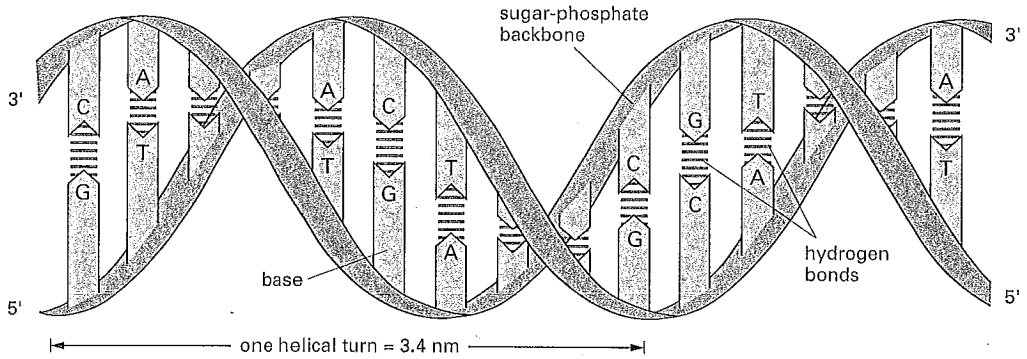
FOUR BASES AS BASE PAIRS OF DNA

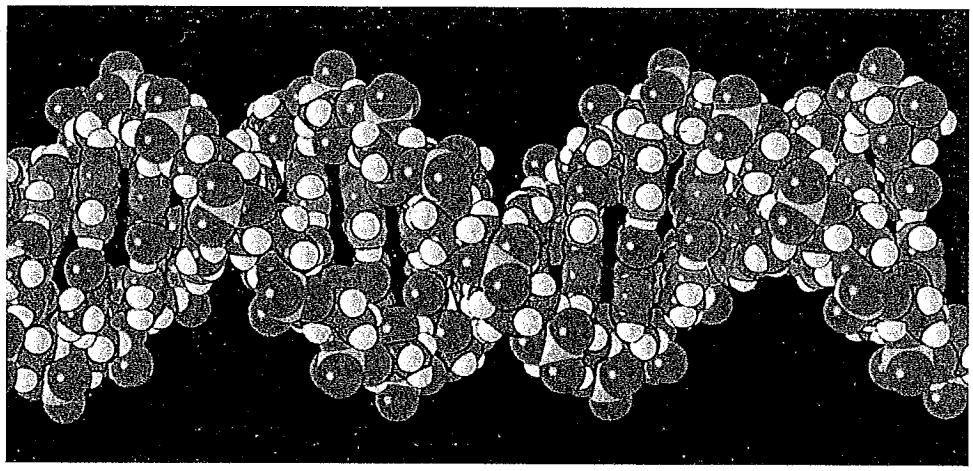


Specific hydrogen bonding between G and C and between A and T (A and U in RNA) generates complementary base-pairing.

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

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.

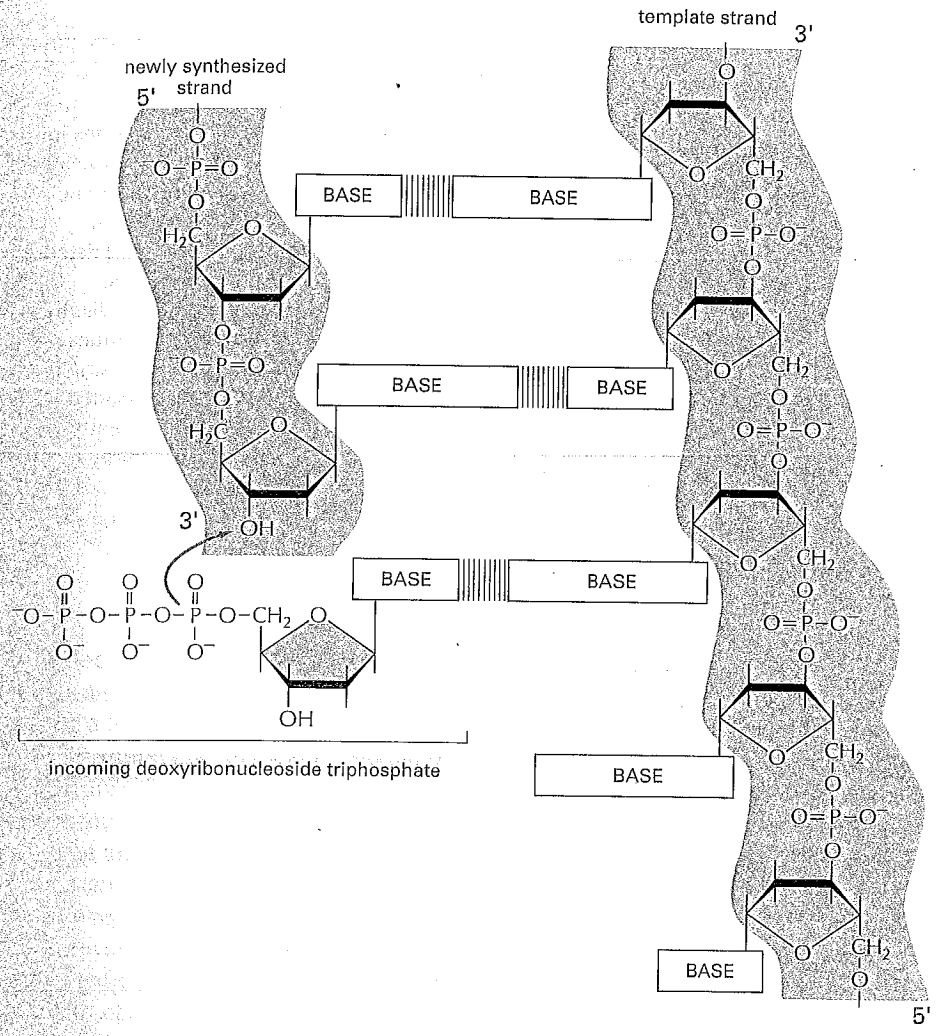


Figure 3–12 DNA synthesis. The addition of a deoxyribonucleotide to the 3′ end of a polynucleotide chain is the fundamental reaction by which DNA is synthesized. As shown, base-pairing between this incoming deoxyribonucleotide and an existing strand of DNA (the *template* strand) guides the formation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

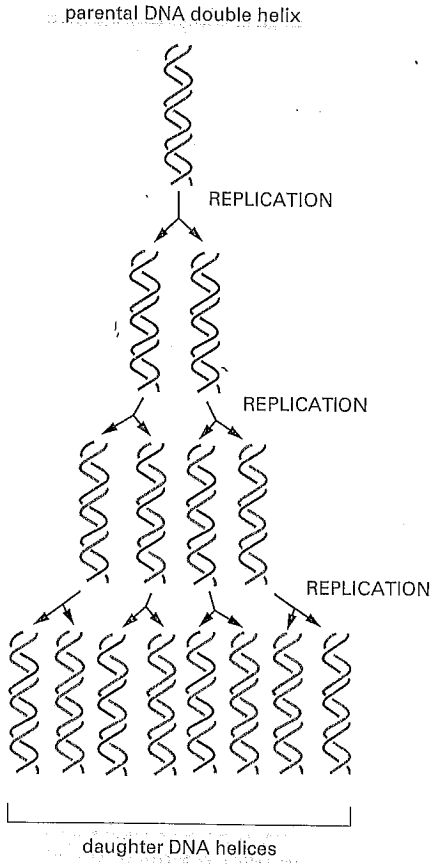


Figure 3–13 The semiconservative replication of DNA. In each round of replication each of the two strands of DNA is used as a template for the formation of a complementary DNA strand. The original strands therefore remain intact through many cell generations.

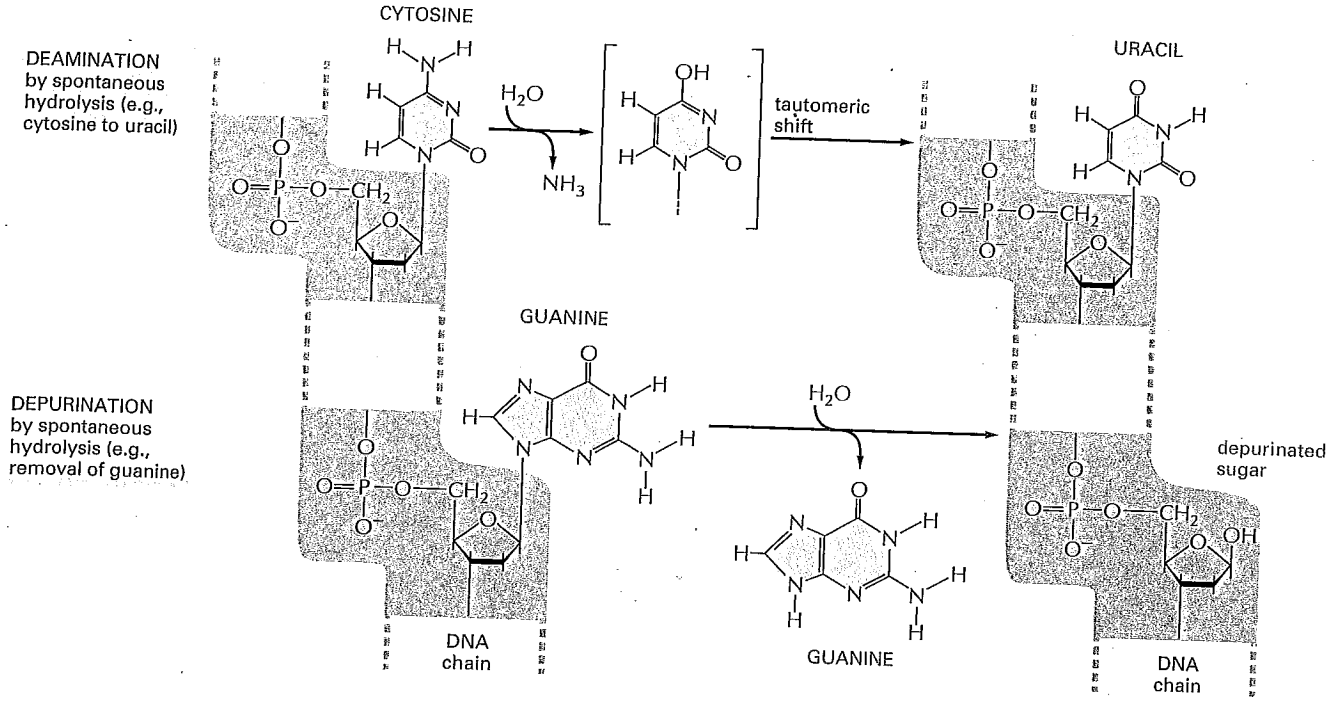


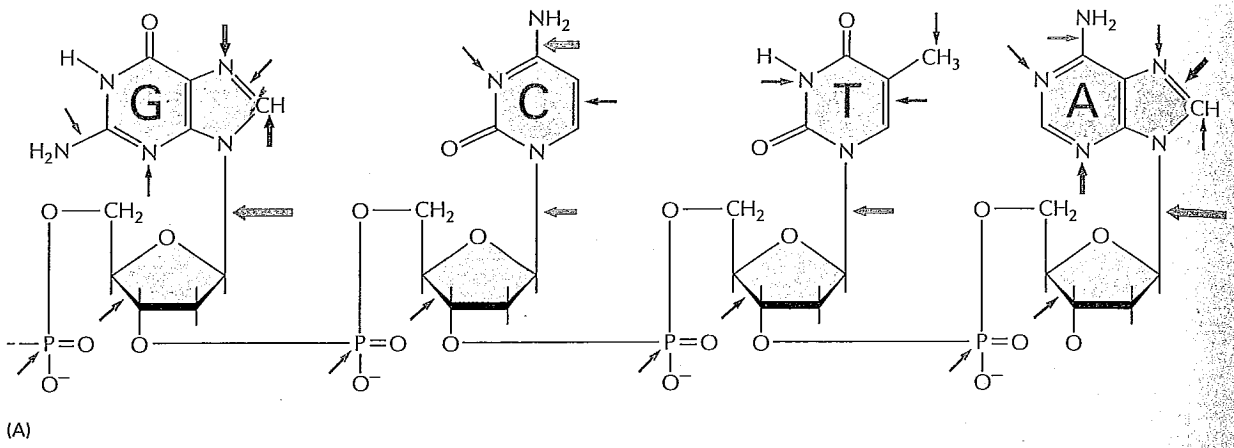
Figure 6–33 Deamination and depurination. These hydrolytic reactions are the two most frequent spontaneous chemical reactions known to create serious DNA damage in cells. Only a single example is shown for each type of reaction. (See also Figure 6–39.)

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.**



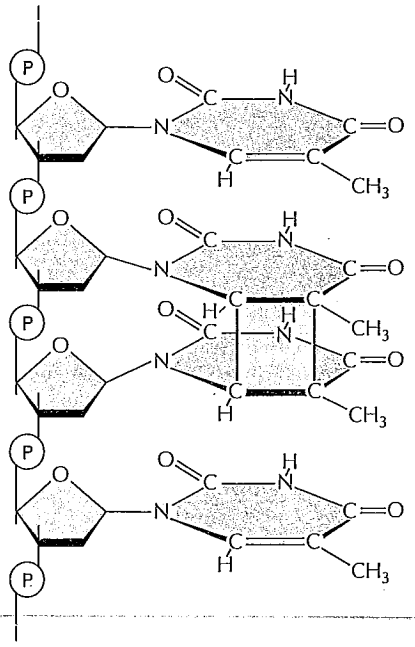


Figure 6–34 A summary of spontaneous alterations likely to require DNA repair. (A) The sites on each nucleotide that are known to be modified by spontaneous oxidative damage (*red arrows*), hydrolytic attack (*blue arrows*), and uncontrolled methylation by the methyl group donor S-adenosylmethionine (*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.

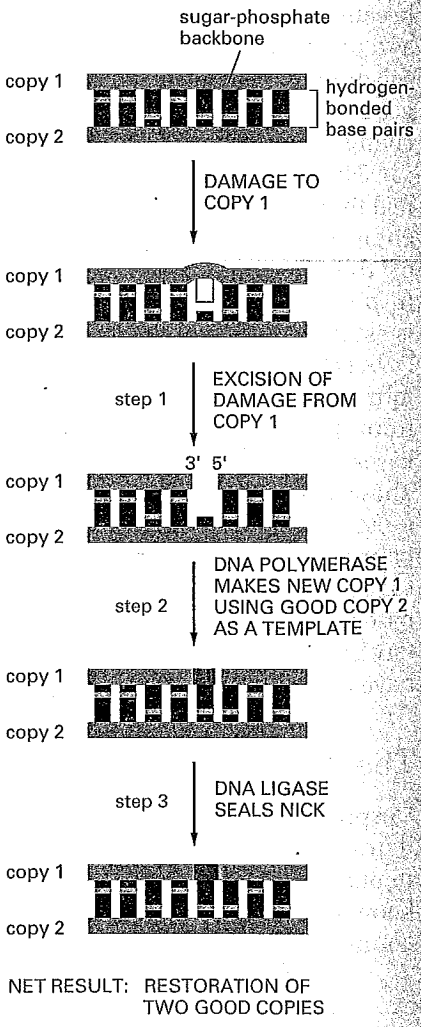
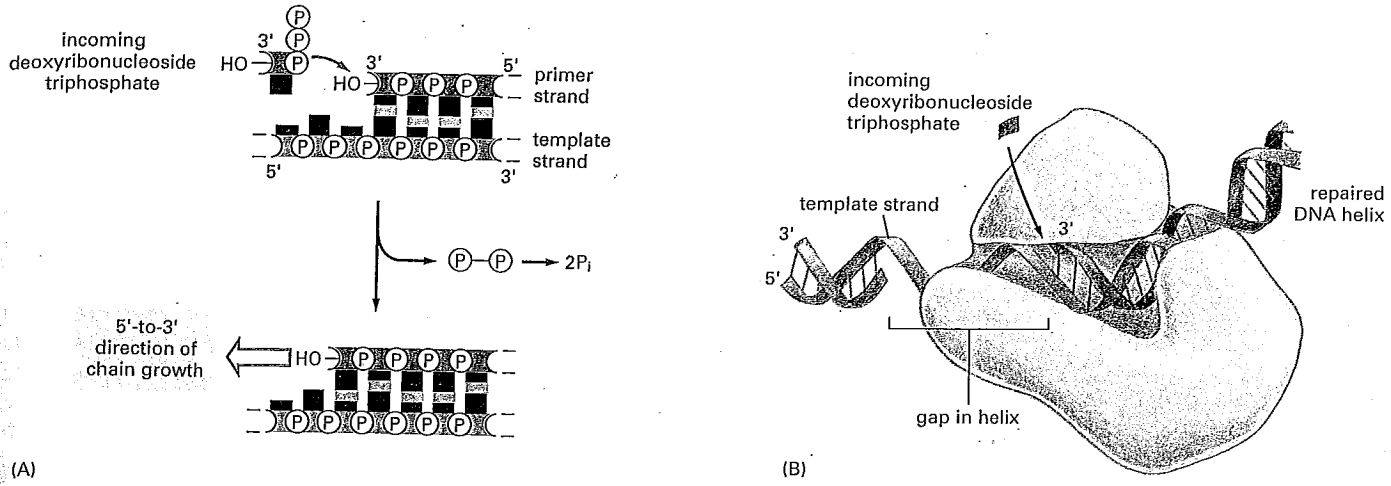


Figure 6–35 DNA repair. The three steps common to most types of repair are excision (step 1), resynthesis (step 2), and ligation (step 3). In step 1 the damage is excised; in steps 2 and 3 the original DNA sequence is restored. *DNA polymerase* fills in the gap created by the excision events, and *DNA 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).



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).

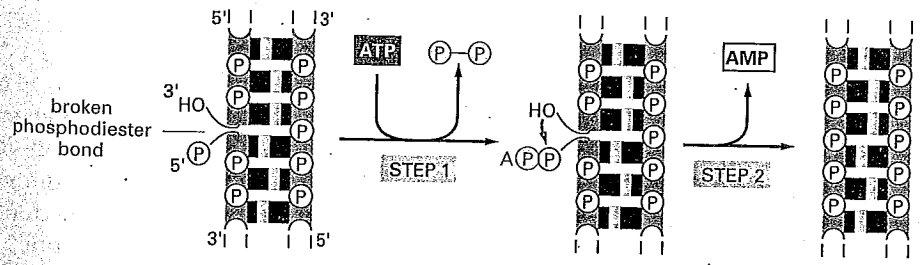


Figure 6–37 The reaction catalyzed by DNA ligase. This enzyme seals a broken phosphodiester bond. As shown, DNA ligase uses a molecule of ATP to activate the 5′ end at the nick (step 1) before forming the new bond (step 2). In this way the energetically unfavorable nick-sealing reaction is driven by 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.

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-

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

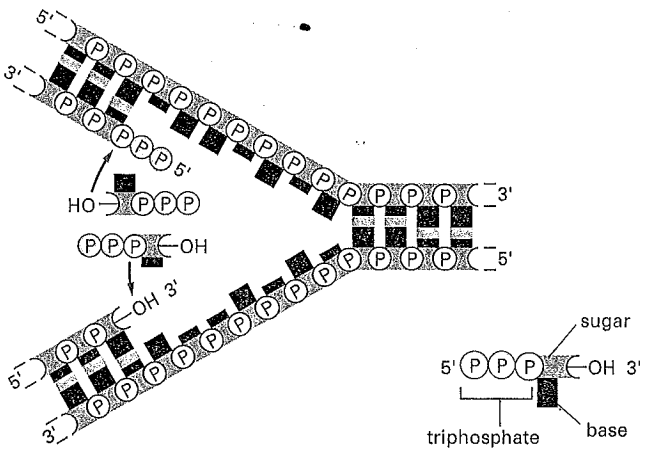


Figure 6–40 An incorrect model for DNA replication. Although it might appear to be the simplest mechanism for DNA replication, the mechanism illustrated here is not the one that cells use. Note that in this scheme both daughter DNA strands would grow continuously, using the energy of hydrolysis of the *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.

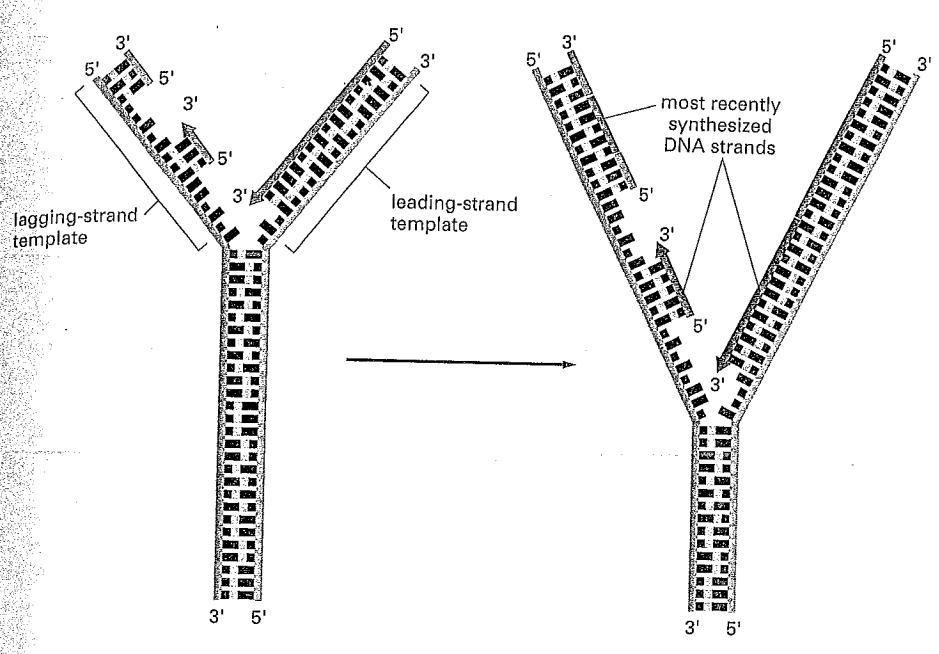


Figure 6–41 The structure of a DNA replication fork. Because both daughter DNA strands (*colored*) are synthesized in the 5′-to-3′ direction, the DNA synthesized on the lagging strand must be made initially as a series of short DNA molecules, called *Okazaki fragments.*

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

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).

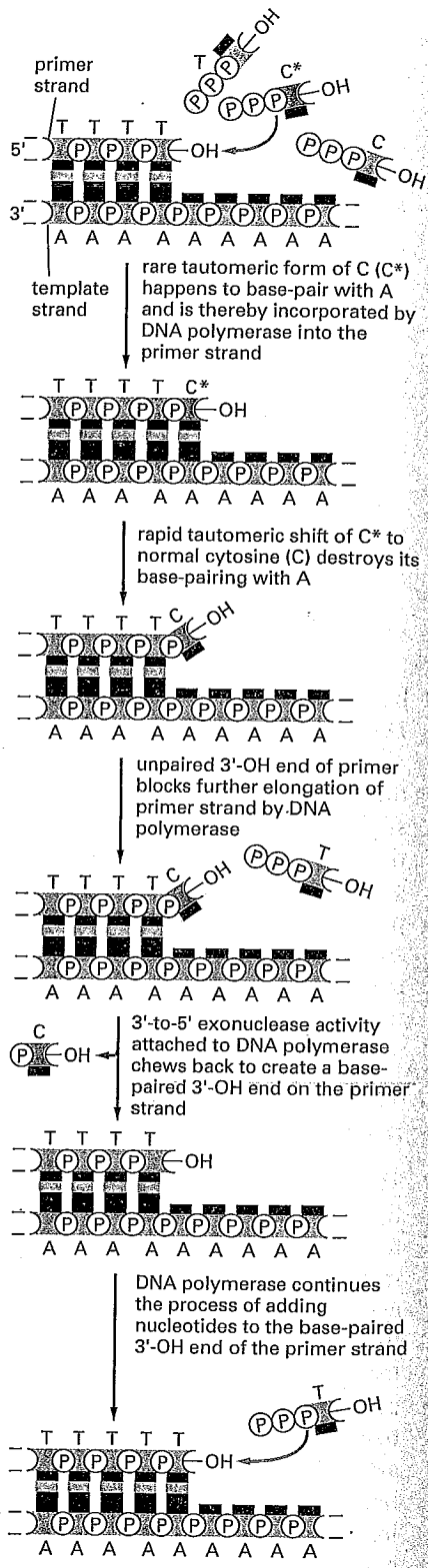


Figure 6–42 Proofreading during DNA replication.

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.

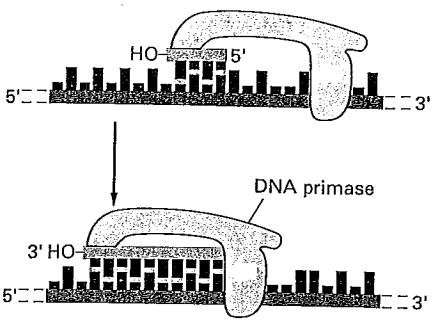


Figure 6–43 RNA primer synthesis. A schematic view of the reaction catalyzed by DNA primase, the enzyme that synthesizes the short RNA primers made on the lagging strand. Unlike DNA polymerase, this enzyme can start a new polynucleotide chain by joining two nucleoside triphosphates together. The primase 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-

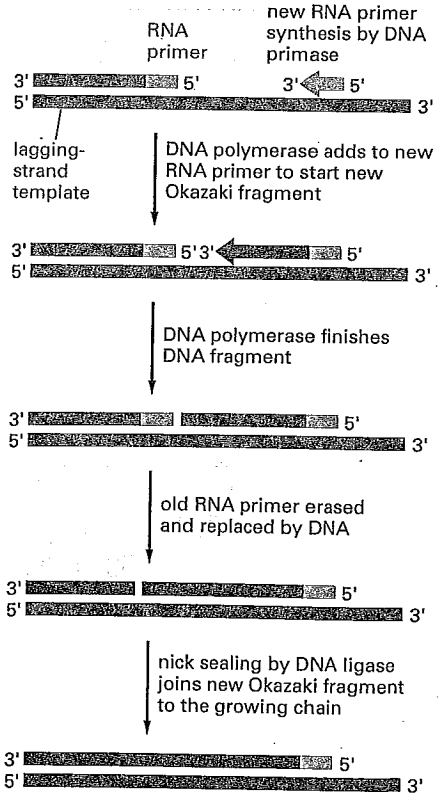
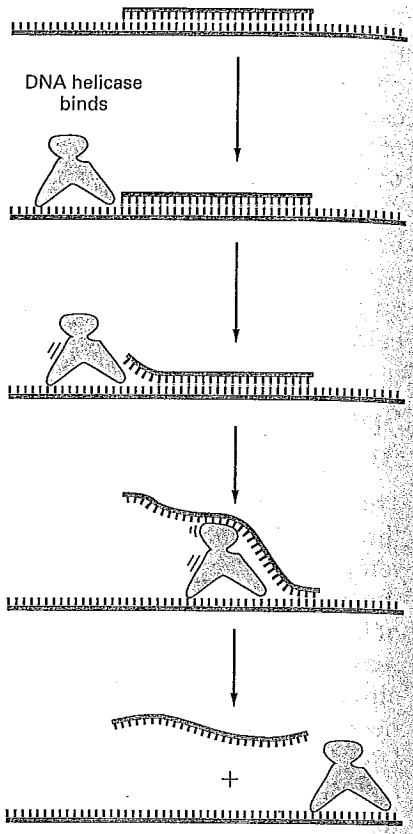


Figure 6–44 The synthesis of one of the many DNA fragments on the lagging strand. In eucaryotes the RNA primers are made at intervals spaced by about 200 nucleotides on the lagging strand, and each RNA primer is 10 nucleotides long. This primer is erased by a special DNA repair enzyme that recognizes 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).

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

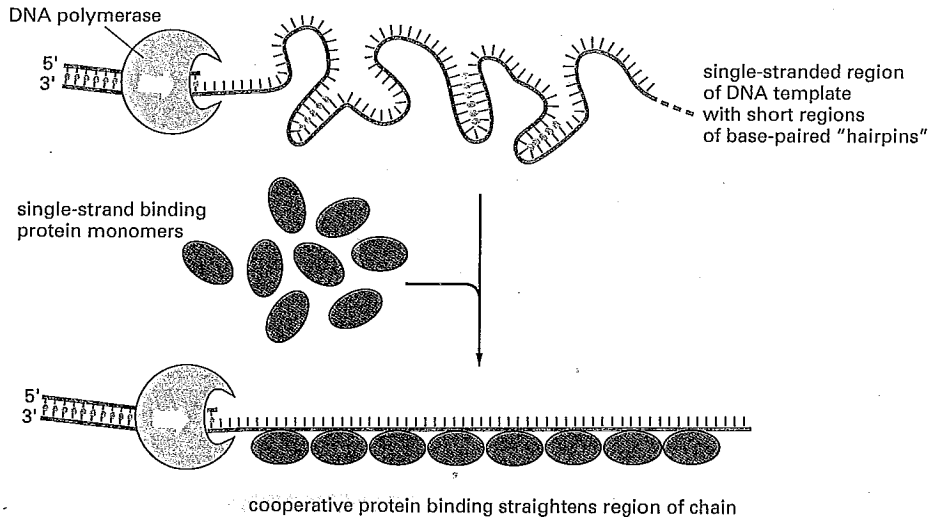
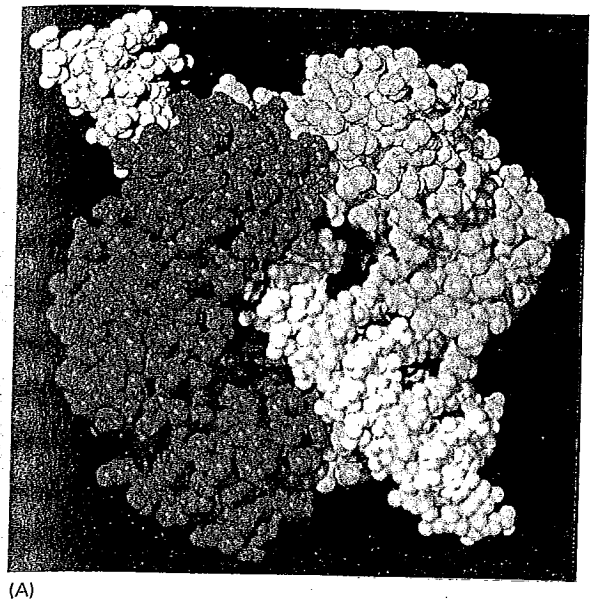
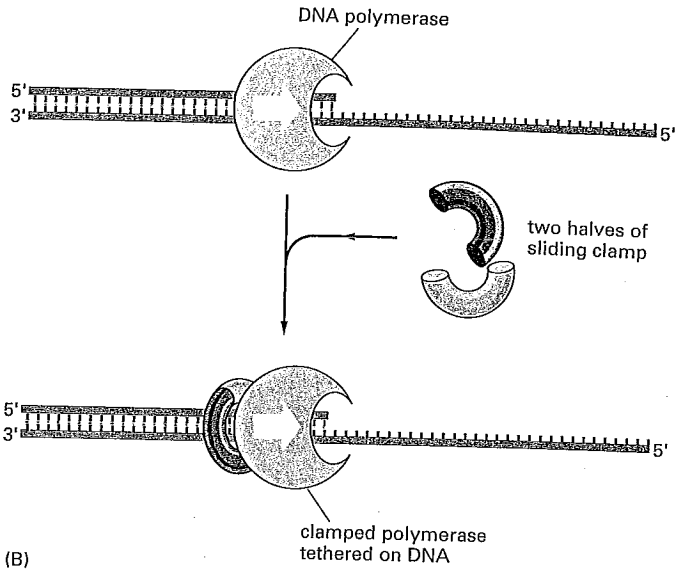


Figure 6–46 The effect of single-strand binding proteins on the structure of single-stranded DNA. Because each protein molecule prefers to bind next to a previously bound molecule (*cooperative binding*) long rows of this protein will form on a DNA single strand. This cooperative binding straightens 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.



(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

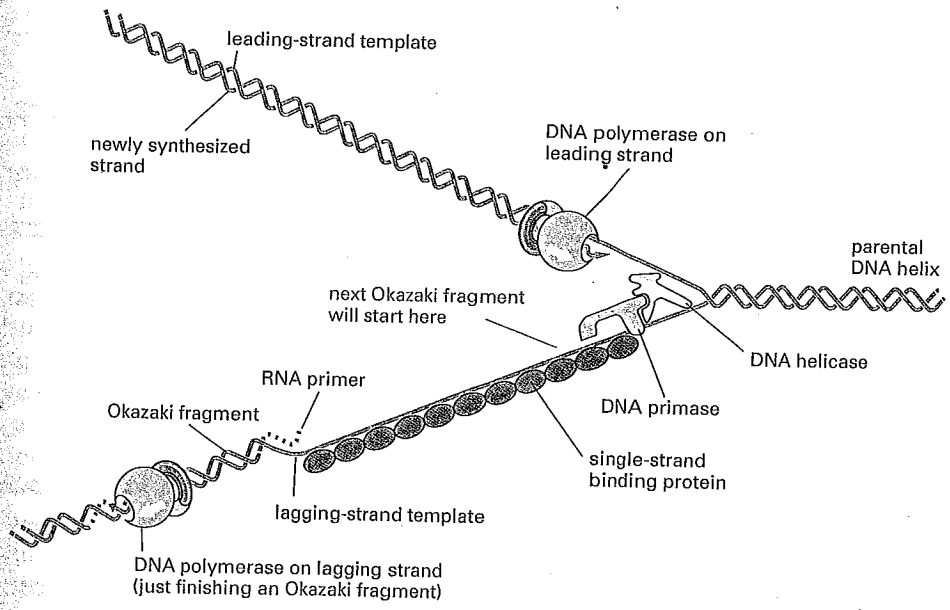


Figure 6–48 The proteins at a DNA replication fork. The major types of proteins that act at a DNA replication fork are illustrated, showing their positions on the DNA.

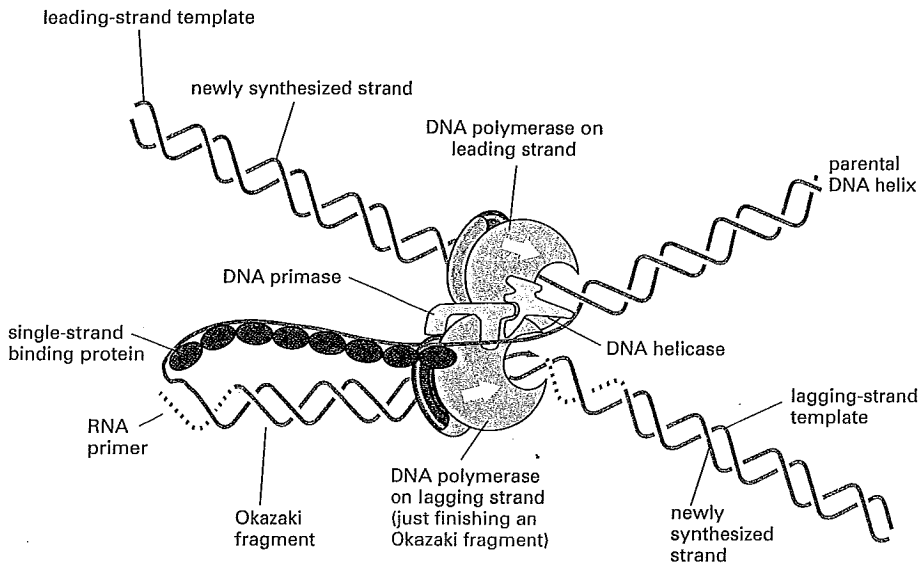


Figure 6–49 A replication fork in three dimensions. This diagram shows a current view of how the replication proteins are arranged at a replication fork when the fork is moving. The two-dimensional structure of Figure 6–48 has been altered by folding the DNA on the lagging strand to bring the lagging-strand 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-

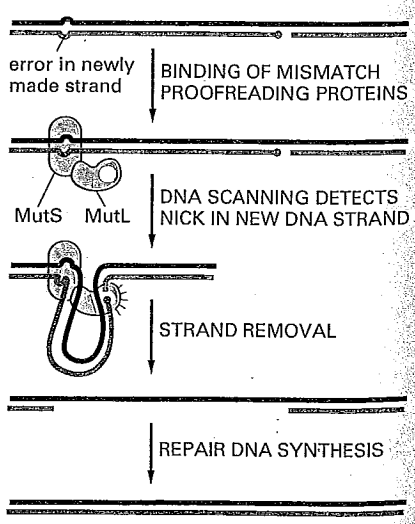


Figure 6–50 A model for mismatch proofreading in eucaryotes. The two proteins shown are present in both bacteria and eucaryotic cells: MutS binds specifically to a mismatched base pair, while MutL scans the nearby DNA for a nick. Once a nick is found, MutL triggers the degradation of the nicked strand 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.