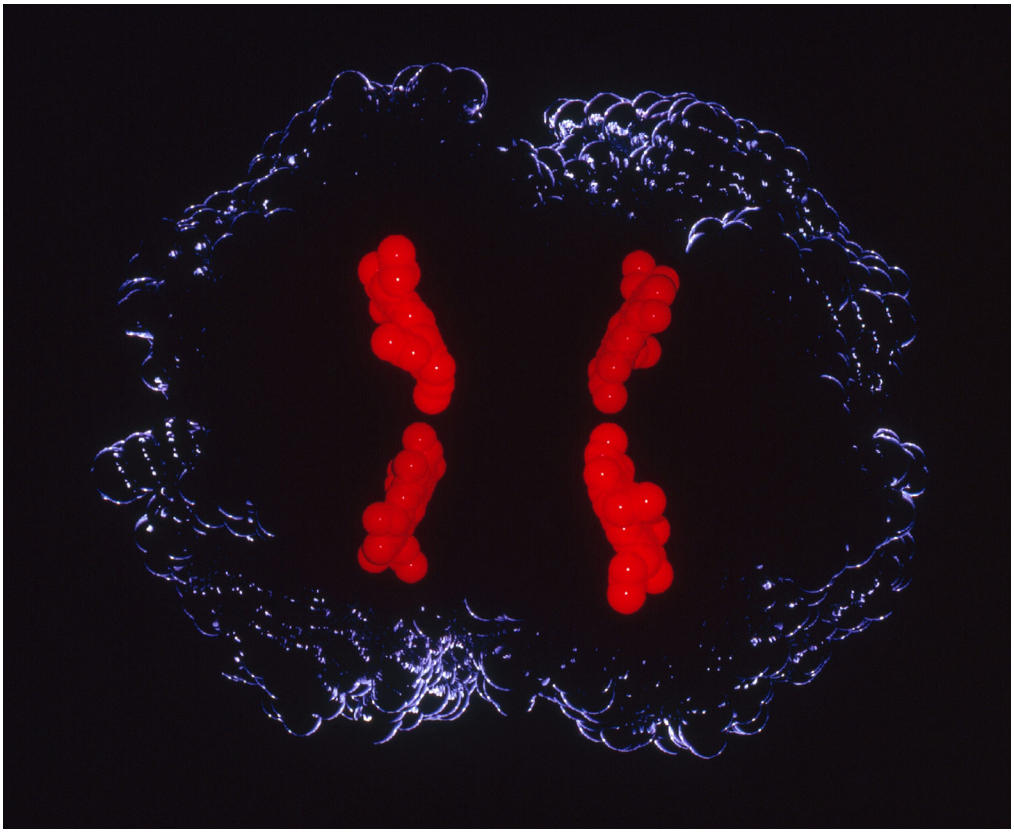
INTERCHAPTER T

# Biological Polymers



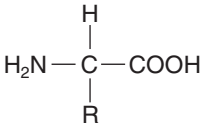
Hemoglobin is a large biological polymer found in red blood cells that transports oxygen to the various tissues in our bodies. Hemoglobin is made up of four subunits, each folded around a group called a heme that contains an iron ion that binds to oxygen. Thus, each hemoglobin molecule can bind up to four oxygen molecules. The molecular structure of hemoglobin was first determined using X-ray crystallography by Max Perutz in 1959, who shared the 1962 Nobel Prize in Chemistry.

The fields of biochemistry, medicine, and molecular biology have been profoundly influenced by discoveries in polymer chemistry. In exploring the relationship between the three-dimensional structures of biomolecules and their biological function, biochemists have elucidated how nerve impulses travel, how enzymes catalyze biological reactions, and the molecular mechanisms underlying many diseases. An understanding of polymers helped elucidate how DNA and RNA molecules store and transmit genetic information and direct the synthesis of proteins. The understanding of the structure and function of proteins stands as one of the greatest achievements of modern science and is still a highly active area of research. In this Interchapter, we shall briefly consider the structure and function of proteins and DNA.

## T-1. Amino Acids Are the Monomer Units of Biological Polymers Called Proteins

**Proteins** are biological polymers**.** The word *protein* was coined in 1838 by the Swedish chemist Jöns Berzelius, drawing on the Greek word *proteios,* which means "of the first rank." As their name suggests, proteins are essential to life. Hemoglobin (Frontispiece), which transports oxygen in the blood and hydrogen carbonate ions from cells, is a protein. Other globular (or roughly spherical) proteins act as catalysts (enzymes) in living organisms. The fibrous protein collagen provides the high tensile strength of skin and bone; other fibrous proteins include the antibodies that protect the body by combining with viruses, foreign bacteria, and cells from other organisms. All told, proteins constitute about 12% to 15% by mass of the human body.

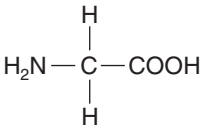
Proteins are polymers whose monomeric units are **amino acids**. The general formula of an amino acid is



These compounds are called amino acids because they contain both an amino group, $-NH_2$, and an acidic group, –COOH. Amino acids differ from one another only in the **side group** (denoted by R in the formula shown above) attached to the central carbon atom. A total of 20 amino acids are commonly found in proteins. At least some of these 20 amino acids are found in proteins at all levels of life, from the simplest bacteria to humans. Most natural proteins contain between 50 and 2000 of these monomer units, and the molecular mass of most protein polymer chains ranges from 550 Da to 220 000 Da.

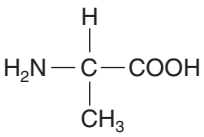
In biology it is common to refer to atomic mass units as Daltons, abbreviated Da.

Except for glycine,



glycine

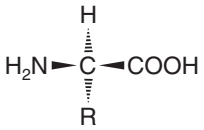
which is the simplest amino acid, all the amino acids have four different groups attached to the central carbon atom. For example, the structural formula for the amino acid alanine is



alanine

Notice that alanine has a methyl ($-CH_3$) side group. The side groups and the names of the corresponding amino acids are shown in Table T.1.

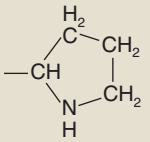
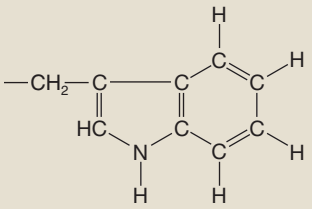
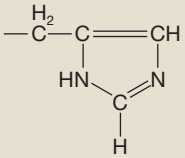
The four bonds about the central carbon atom in an amino acid are tetrahedrally oriented, a geometry that can be represented as



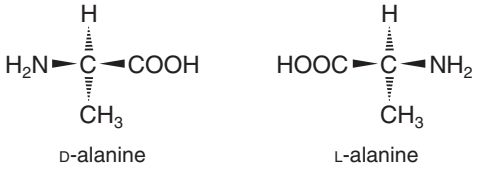
The dashed, wedge-shaped bonds indicate that –H and –R lie below the page; and the dark, wedge-shaped bonds indicate that the $-NH_2$ and –COOH groups lie above the page.

Because the tetrahedral carbon atom at the center of the amino acid backbone is bonded to four distinct groups, all amino acids (except for glycine) are optically active. Recall from Section 8-10 that **optical isomers** are pairs of nonsuperimposable isomers that are mirror images of each other. The optical isomers of amino acids are typically distinguished from each other by a D or L placed in front of the name of the

TABLE T.1 The side groups and names of the 20 common amino acids of proteins

| Side group | Amino acid |
|---|---|
| ***Nonpolar side groups*** | |
| —H | glycine (Gly) |
| —$CH_3$ | alanine (Ala) |
| —CH(—$CH_3$)—$CH_3$ | valine (Val) |
| —$CH_2$CH(—$CH_3$)—$CH_3$ | leucine (Leu) |
| —CH(—$CH_3$)—$CH_2CH_3$ | isoleucine (Ile) |
|  | proline (Pro) |
| —$CH_2$—(benzene) | phenylalanine (Phe) |
|  | tryptophan (Trp) |
| —$CH_2CH_2$—S—$CH_3$ | methionine (Met) |
| ***Uncharged polar side groups*** | |
| —$CH_2OH$ | serine (Ser) |
| —CH(—$CH_3$)—OH | threonine (Thr) |
| —$CH_2SH$ | cysteine (Cys) |
| —$CH_2$—(benzene)—OH | tyrosine (Tyr) |
| —$CH_2$C(=O)—$NH_2$ | asparagine (Asn) |
| —$CH_2CH_2$C(=O)—$NH_2$ | glutamine (Gln) |
| ***Acidic side groups*** | |
| —$CH_2$C(=O)—OH | aspartic acid (Asp) |
| —$CH_2CH_2$C(=O)—OH | glutamic acid (Glu) |
| ***Basic side groups*** | |
| —$CH_2CH_2CH_2CH_2NH_2$ | lysine (Lys) |
| —$CH_2CH_2CH_2$N(H)—C(=NH)—$NH_2$ | arginine (Arg) |
|  | histidine (His) |

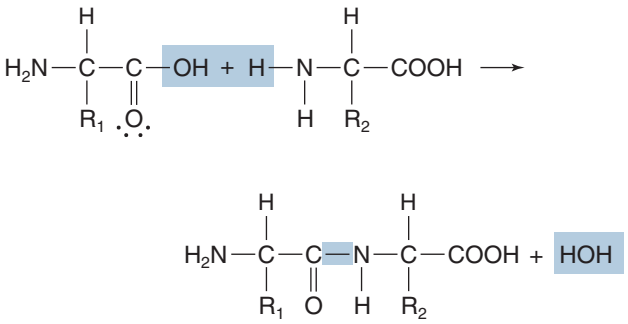
amino acid. The D and L are derived from *dextro-* (meaning right) and *levo-* (meaning left).




Optical isomers ordinarily display the same chemical properties; but, with few exceptions, only the L-isomers of the amino acids occur in biological systems. Biochemical reactions are exceptionally **stereospecific;** that is, they are extremely dependent on the shape of the reactants. Apparently, most of the life on earth originated from L-amino acids; and once the process started, it continued to use only L-isomers, which, unlike the D-isomers, are recognized by our enzymes. This has great significance in biochemistry. For example, the antibiotic penicillin attacks proteins containing D-alanine that are found in the cell walls of certain bacteria, but does not attack the L-isomer found in humans. As a result penicillin kills bacteria but not people. Fragrances and flavors are another example where the stereochemistry of the molecules plays an important role. For instance, the molecule D-carvone is the component of oil from caraway seeds that smells like rye, whereas its mirror image, L-carvone, is the component of spearmint oil that smells like spearmint.

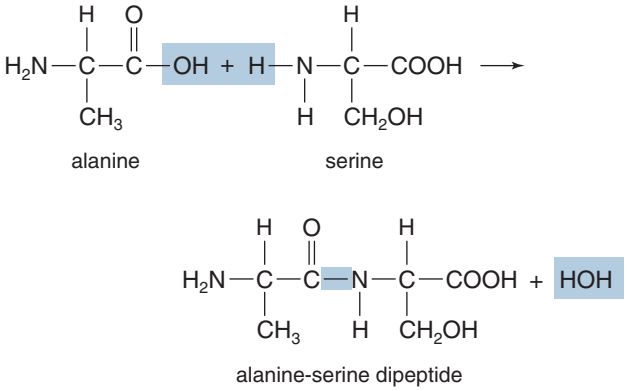
## T-2. Proteins Are Formed by Condensation Reactions of Amino Acids

Proteins are formed by condensation reactions similar to the reaction that results in the formation of nylon (Interchapter S). The carboxyl group on one amino acid reacts with the amino group on another, thereby forming a **peptide bond.** For example,

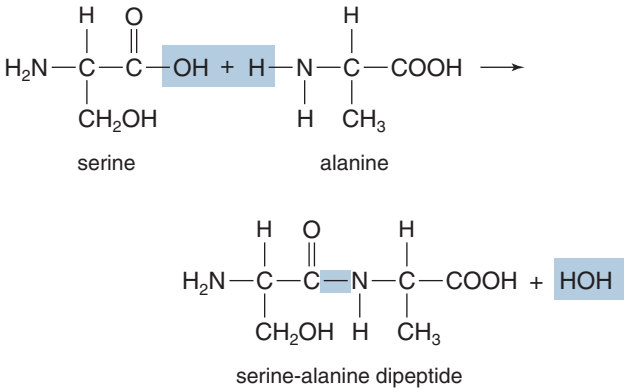



The portion of the amino acid that remains in the chain after the water molecule is split out is called an **amino acid residue.** The product of the reaction is called a **dipeptide** because it contains two amino acid residues.

For example, the reaction equation that describes the formation of a dipeptide from the two amino acids alanine and serine (Table T.1) is

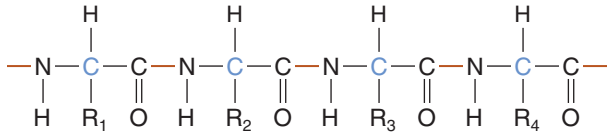



This result is not the only one possible, however. A different dipeptide is formed when the carboxyl group on serine reacts with the amino group on alanine according to




Thus, we see that it is necessary to specify the order of the amino acids in a peptide.

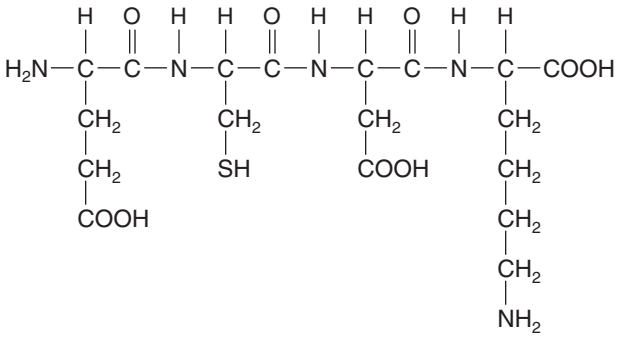
Further condensation reactions of a dipeptide with additional amino acid molecules produces a **polypeptide,** which is a polymer having amino acids as monomers. Polypeptides are thus long chains of amino acid residues joined together by peptide bonds. The chain to which the amino acid side groups are attached is called the **polypeptide backbone.** An example of a portion of a polypeptide is



where the carbon atoms that are bonded to the amino acid side groups are shown in blue and the peptide bonds are shown in red.

As the number of amino acids in a polypeptide increases, it becomes unwieldy to write out the complete chemical formula of the polypeptide. For this reason, three-letter abbreviations are commonly used to designate the amino acids in a polypeptide chain (Table T.1). For example, the amino acid alanine is designated Ala, and serine is designated Ser. The two dipeptides formed from alanine and serine can be designated by Ala-Ser and Ser-Ala.

Using the entries in Table T.1, let's write out the chemical formula of the tetrapeptide ($H_2N$-end) Glu-Cys-Asp-Lys:



Proteins are naturally occurring polypeptides. Each protein is characterized by a specific number and variety of amino acid units that occur in a specific order (sequence) along the polypeptide backbone. Table T.2 lists some proteins and the number of amino acid units in each.

The sequence of the amino acid units in a polypeptide defines the **primary structure** of that polypeptide. The primary structure uniquely characterizes a protein. The primary structures of hundreds of proteins have been determined since the 1950s. Figure T.1 shows the primary structure of the protein beef insulin, which is a polypeptide hormone that regulates carbohydrate metabolism. A deficiency of insulin in humans leads to the disease *diabetes mellitus*.

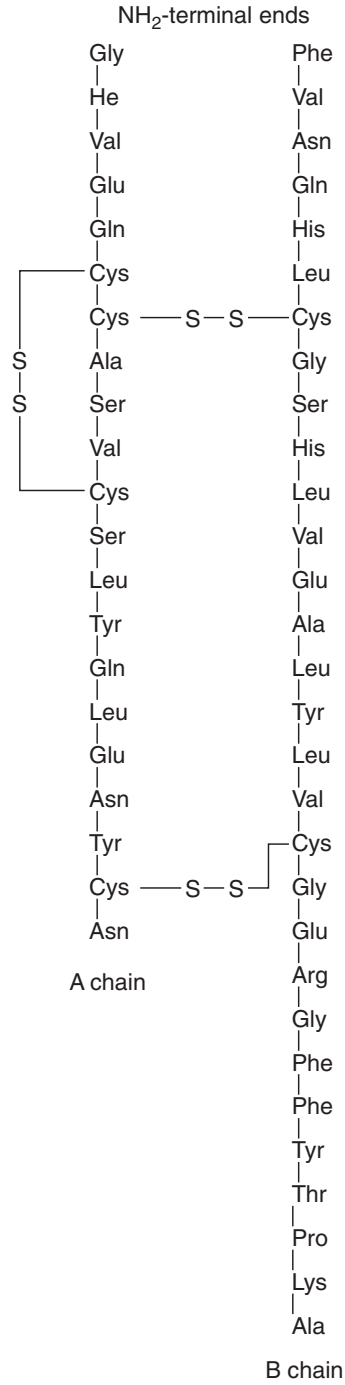



Figure T.1 The primary structure of the protein beef insulin. The amino acids are designated by standard three-letter abbreviations. The determination of the primary structure of a protein is like a complicated chemical jigsaw puzzle. The protein is hydrolyzed into shorter chains, which are separated and analyzed individually. The first primary structure determination was completed in 1953 by the British chemist Frederick Sanger, who received the 1958 Nobel Prize in Chemistry for this work. He received a second Nobel Prize in 1980 for the sequencing of DNA.

TABLE T.2 Number of amino acids in and formula mass of common proteins

| Protein | Number of amino acids | Formula mass | Number of polypeptide chains |
|---|---|---|---|
| insulin (hormone) | 51 | 5700 | 2 |
| cobratoxin (snake toxin) | 62 | 7000 | 1 |
| myoglobin (carries oxygen in muscles) | 153 | 16900 | 1 |
| keratin (wool protein) | 204 | 21000 | 1 |
| actin (muscle protein) | 410 | 46000 | 1 |
| hemoglobin (transports oxygen in bloodstream) | 574 | 64500 | 4 |
| alcohol dehydrogenase (metabolism of ethanol) | 748 | 80000 | 2 |
| γ-globulin (antibody) | 1250 | 150000 | 4 |
| collagen (skin, tendons, cartilage) | 3000 | 300000 | 3 |

## T-3. The Shape of a Protein Molecule Is Called Its Tertiary Structure

A key step in understanding how a particular protein functions is the determination of its shape. Because many proteins are extremely large molecules, this task is not easy. The definitive method for determining a protein's structure is X-ray crystallography. X-ray patterns can be used to determine the arrangement of atoms in crystalline solids. The X-ray patterns obtained from proteins, however, are more difficult to analyze and interpret because so many atoms are involved (Figure T.2).

In the 1950s two American chemists, Linus Pauling and Robert B. Corey, were able to interpret X-ray patterns of proteins to show that many proteins have regions in which the chain twists into a helix (which is the shape of a spiral staircase). Pauling and Corey called the helix an **α-helix** (Figure T.3). The helical shape results from the formation of hydrogen bonds between peptide linkages in the peptide chain. Individually, these hydrogen bonds are relatively weak, but collectively they combine to bend the protein chain into the α-helix. This coiled, helical shape in different regions of a protein chain is called **secondary structure.**

The shape of a protein molecule in water results from a complicated interplay between the amino acid side groups along the protein chain and the solvent, water. This interplay causes the protein to coil, fold, and bend into a three-dimensional shape called the **tertiary structure.** The tertiary structure of a protein can be obtained from X-ray analysis. Because proteins play crucial roles in nearly all biological processes, an active area of biochemical research is an understanding of how amino acid sequences determine the conformations of proteins.

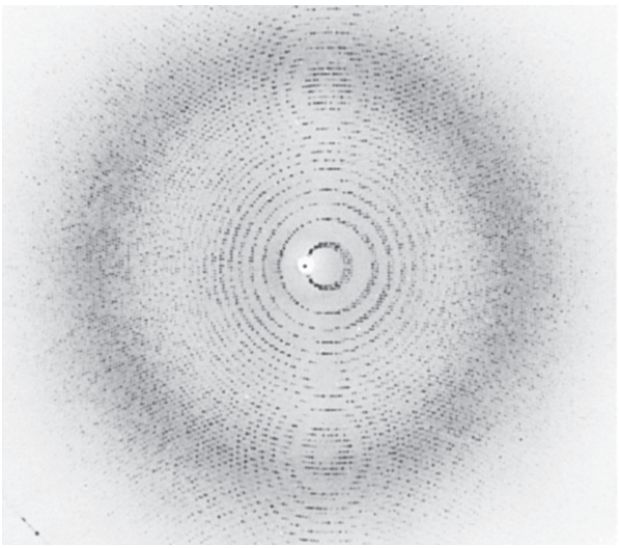


Figure T.2 X-ray diffraction patterns like this one from a polio virus can be used to determine the structures of proteins and other biological molecules.

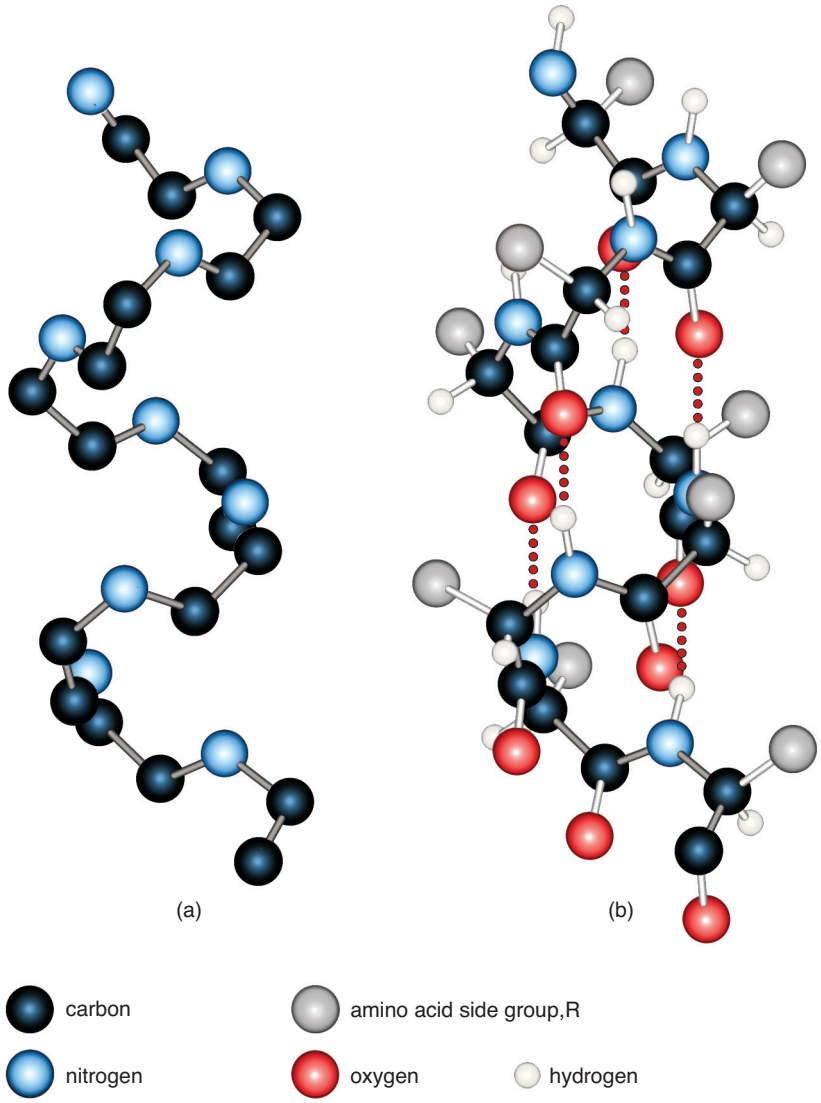



Figure T.3 A segment of an α-helical region along a polypeptide chain. The chain is held in a helical shape by hydrogen bonds (dotted lines). The bond is formed between a hydrogen atom in one peptide bond and an oxygen atom in the fourth peptide bond further along the polypeptide chain. Part (a) shows just the backbone of the chain, while (b) shows the groups attached to the backbone and the hydrogen bonds.

## T-4. DNA Is a Double Helix

The final class of biological polymers, or **biopolymers,** that we study in this Interchapter are the **polynucleotides.** The two most important polynucleotides are **DNA** (deoxyribonucleic acid) and **RNA** (ribonucleic acid). DNA occurs in the nuclei of cells and the genome of viruses and is the principal component of chromosomes (Figure T.4). Genetic information that is passed from one generation to another is stored in DNA molecules. The discovery in 1953 of just how this is done has led to a revolution in biology that produced the fields of molecular biology and genetic engineering. In order to see how DNA can store and pass on information, we must examine its molecular structure.

DNA is a polynucleotide, by which we mean that it is a polymer made up of nucleotides. **Nucleotides,** the monomers of DNA and RNA, consist of three parts: a sugar (carbohydrate), a phosphate group, and a nitrogen-containing ring compound called a

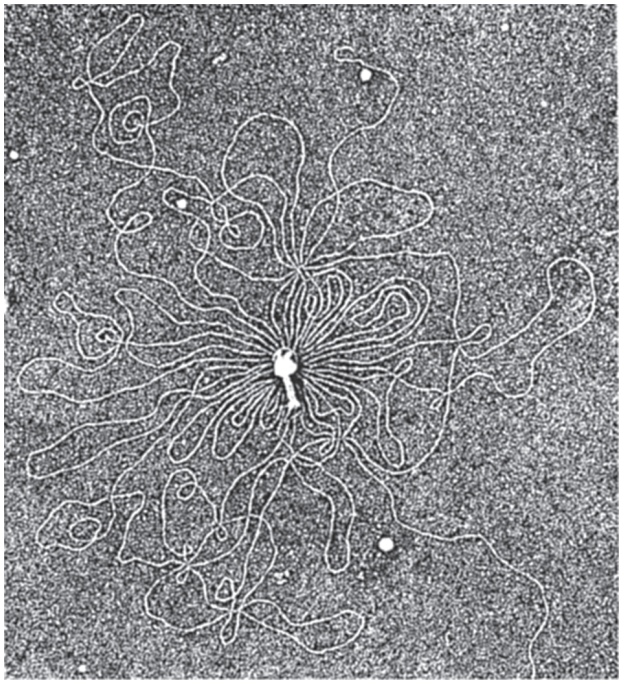
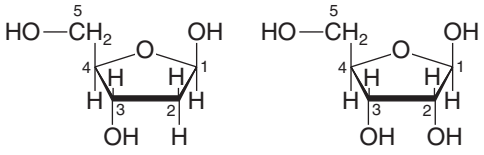


Figure T.4 An electron micrograph of a virus particle that has burst and released strands of DNA. The long, cylindrical molecule is revealed beautifully in the photo.

**base.** The sugar in DNA is 2-deoxyribose and that in RNA is ribose:




Carbon atoms, which are understood to constitute the vertices of these rings, are numbered 1 to 4 in the formulas. Notice that the difference between 2-deoxyribose and ribose is that 2-deoxyribose is lacking an oxygen atom at the number 2 carbon atom.

In both DNA and RNA, a phosphate group (red) is attached to the number 5 carbon atom in the sugar, as shown in the box below. The group labeled X is –OH in ribose and –H in 2-deoxyribose.

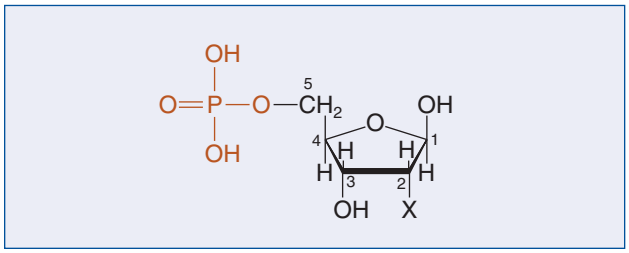
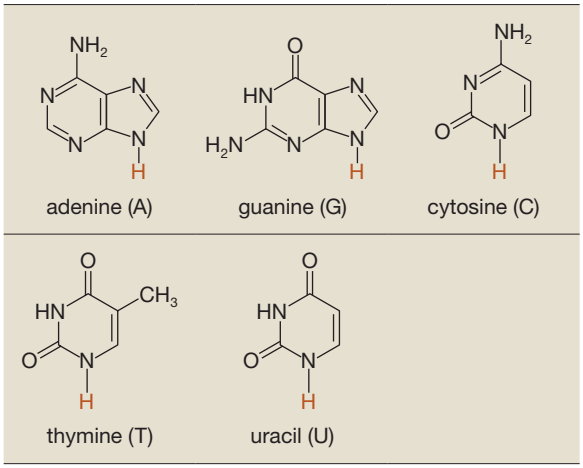
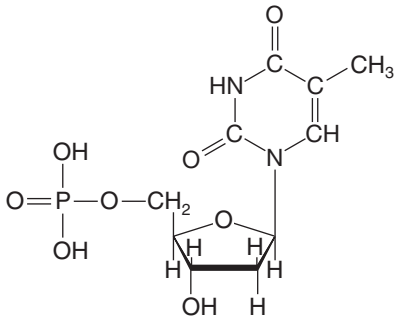


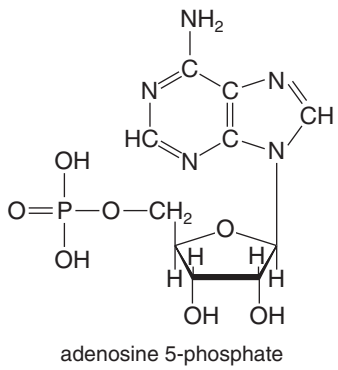
TABLE T.3 The five bases that occur in DNA and RNA*




*Thymine is found only in DNA; uracil only in RNA.

The five bases that occur in DNA and RNA are given in Table T.3. The bases are bonded to the ribose or deoxyribose rings by condensation reactions involving the hydrogen atoms shown in red on the bases in Table T.3 and the –OH group on the number 1 carbon atom in the ribose and deoxyribose rings. For example, if thymine or adenine is the base, we have

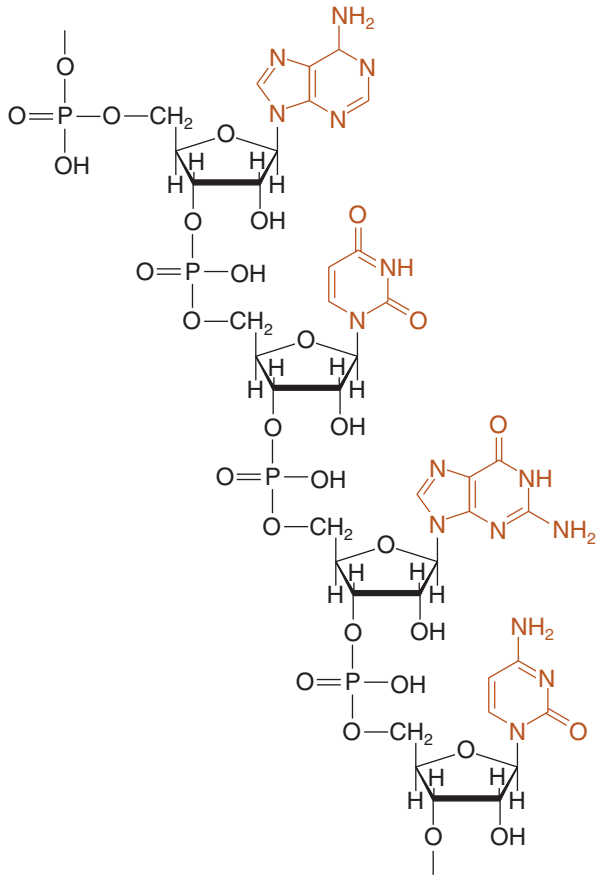



Both of these molecules are nucleotides. Deoxythymidine 5-phosphate is one of four monomers

of DNA; and adenosine 5-phosphate is one of four monomers of RNA. DNA contains only the four bases adenine (A), guanine (G), cytosine (C), and thymine (T); and RNA contains adenine (A), guanine (G), cytosine (C), and uracil (U).

Nucleotides can be joined together by a condensation reaction between the phosphate group of one nucleotide and the 3-hydroxyl group of another. The result is a polynucleotide, part of which might look like this:



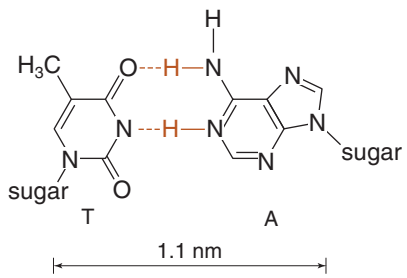
Thus, we see that DNA and RNA consist of a **sugar-phosphate backbone** (shown in black) with bases attached at intervals (shown in red). Let's see now how a molecule like DNA can store and pass on genetic information.

The key to understanding how DNA works lies in its three-dimensional structure. In 1953, James Watson and Francis Crick (see sidebar, page T9) proposed that DNA consists of two polynucleotide chains intertwined in a **double helix** (Figure T.5). Their proposal was based on two principal observations: X-ray data indicated that DNA is helical; and chemical analysis revealed that, regardless of the source of DNA, be

Figure T.5 The double helix structure of DNA consists of two polynucleotide strands twisted about each other.

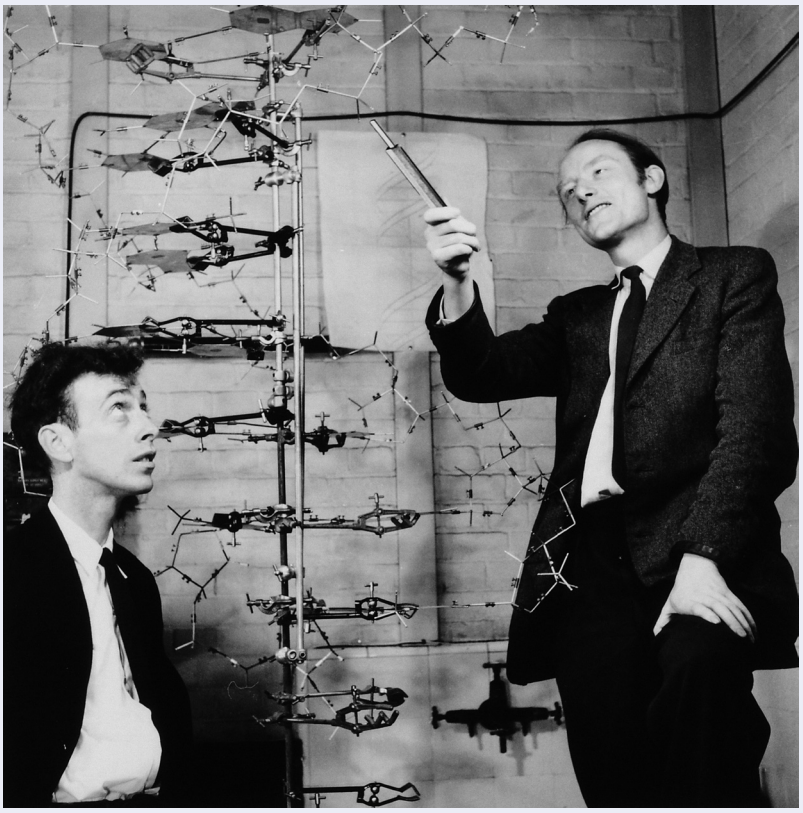
it a simple bacterium or the higher vertebrates, the amount of guanine is always equal to the amount of cytosine and the amount of adenine is always equal to the amount of thymine.

Watson and Crick realized that the bases in DNA must somehow be paired. Working with molecular models, they discovered that thymine (T) and adenine (A) were of the right shape and size to form two hydrogen bonds:

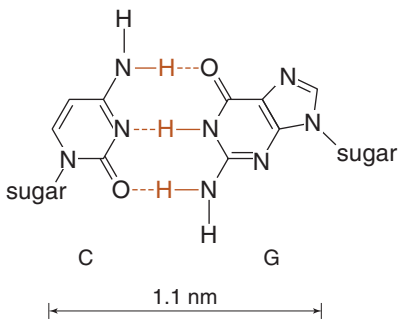


### THE DOUBLE HELIX

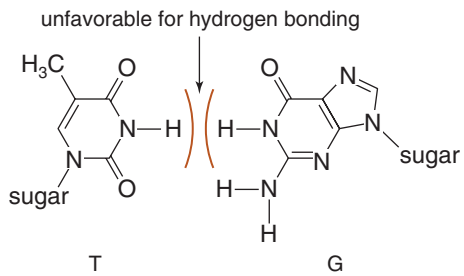
In the early 1950s, James Watson (*left*), who had recently received his Ph.D. in zoology from Indiana University, went to Cambridge University on a post-doctoral research fellowship. He and the British physicist Francis Crick (*right*) worked together on the molecular structure of DNA. In 1953 they proposed the double helix model of DNA, which explains elegantly how DNA can store and transmit genetic information. Their proposal is one of the most important scientific breakthroughs of modern times. Watson and Crick were awarded the Nobel Prize in Physiology and Medicine in 1962. The details of their discovery are given by Watson in his book *The Double Helix*.



Similarly, cytosine (C) and guanine (G) were found to form three hydrogen bonds:




Notice that both the T–A and the C–G base pairs encompass a distance of 1.1 nm, thus allowing the two strands of the double helix to be evenly separated by 1.1 nm. Other possible base pairs, such as C–C, T–T, and C–T, encompass a narrower distance and A–A, G–G, and A–G a wider one. Others that theoretically would be the right size (A–C and G–T) cannot pair because their atoms are not in suitable positions to form hydrogen bonds:




Thus, only A–T and G–C base pairs form, and this restricted base pairing accounts for the structure of DNA. The two strands of the double helix are said to be **complementary.** For example, if the base sequence along a portion of one strand of a double helix is ··· AGCCTCG ···, then the corresponding sequence on the other strand must be ··· TCGGAGC ··· because the two sequences must be complementary to each other, meaning that a T and an A must be opposite each other and a G and a C must be opposite each other. The two strands of the DNA double helix are held together electrostatically by the hydrogen bonds between complementary bases (Figure T.6).

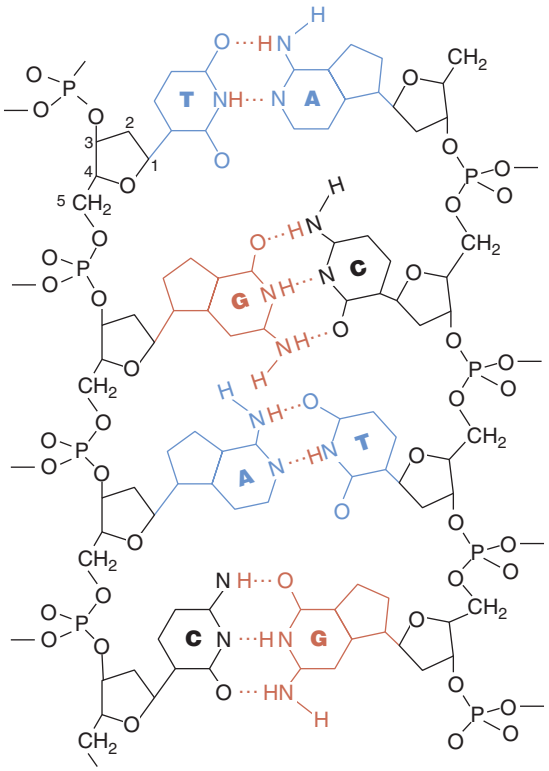


Figure T.6 Hydrogen bonds between complementary base pairs hold the two strands of DNA together in a double helix configuration.

## T-5. DNA Can Duplicate Itself

Hydrogen bonds, unlike covalent bonds, are weak enough to allow the double helix to uncoil into two separate strands at moderate temperatures (~40°C). Each strand can then act as a template for building a complementary strand, and the result is two double helices that are identical to the first. In this way genetic information is transmitted. Thus, the Watson-Crick model for DNA explains not only DNA structure but also DNA replication.

Living systems differ from one another by the myriad biochemical processes characteristic of each system. Almost all these biochemical reactions are controlled by enzymes, and many of them involve other proteins as well. In a sense, each living system is a reflection of its various proteins. What we mean by genetic information is the information that calls for the production of all the proteins characteristic of a given organism. This is the information stored in DNA.

Researchers discovered in the 1950s that each series of three bases along a DNA segment represents a code for binding to a particular amino acid during protein synthesis. For example, the triplet AAA is a DNA code for phenylalanine, and TTC is a DNA code for lysine. Thus, the segment AAATTC in DNA would give rise to a segment Phe-Lys in a protein. Because these code words are represented by a sequence of three bases, the code is called the **triplet code.**

Because the nature and the order of the bases are equivalent to genetic information, the bases are in the interior of the double helix for protection. A **gene** is a segment along a DNA molecule that codes the synthesis of one polypeptide. A DNA molecule can have a molecular mass of over 100 000 000.

The chemical reactions that are involved in transcribing the base sequence along a DNA strand into a protein molecule are complicated but fairly well understood. They involve several types of RNA and numerous enzymes and other proteins. If you go on to take a course in biochemistry or biology, you will study the DNA-protein pathway.

## TERMS YOU SHOULD KNOW

protein *T1*
amino acid *T1*
side group *T1*
optical isomer *T1*
stereospecific *T3*
peptide bond *T3*
amino acid residue *T3*
dipeptide *T3*
polypeptide *T3*
polypeptide backbone *T3*
primary structure *T4*
α-helix *T5*
secondary structure *T5*
tertiary structure *T5*
biopolymer *T6*
polynucleotide *T6*
DNA *T6*
RNA *T6*
nucleotide *T6*
base *T7*
sugar-phosphate backbone *T8*
double helix *T8*
complementary *T9*
triplet code *T10*
gene *T10*

## QUESTIONS

**T-1.** Why isn't the amino acid glycine optically active?

**T-2.** What is the primary, secondary, and tertiary structure of a protein?

**T-3.** List three differences between RNA and DNA.

**T-4.** Write the equations for the reactions between tyrosine and valine.

**T-5.** Draw the structural formulas for the two possible dipeptides that can be formed from the reaction between glycine and alanine.

**T-6.** How many different tripeptides can be formed from two different amino acids? How many different tripeptides can be formed from three different amino acids?

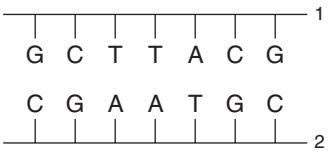
**T-7.** Draw the structural formula for the tripeptide glu-val-cys.

**T-8.** Draw the structural formula for the DNA triplet GAT.

**T-9.** If the base sequence along a portion of one strand of a double helix is AAGTCTCGA, what must the corresponding sequence on the other strand be?

**T-10.** Determine the complementary base sequence that corresponds to the following sequence of DNA bases:


**T-11.** Suppose a segment along a double helix is




Draw the segments obtained when the DNA duplicates itself.