(5) $CH_3COCH_2CO\text{–}S\text{–}ACP + NADPH + H^+ \rightleftharpoons$
$D\text{-}CH_3CHOHCH_2CO\text{–}S\text{–}ACP + NADP^+$
D-β-Hydroxybutyryl-S-ACP

(6) $D\text{-}CH_3CHOHCH_2CO\text{–}S\text{–}ACP \rightleftharpoons trans\text{-}CH_3CH{=}CHCO\text{–}S\text{–}ACP + H_2O$
Crotonyl-S-ACP

(7) $trans\text{-}CH_3CH{=}CHCO\text{–}S\text{–}ACP + NADPH + H^+ \rightleftharpoons$
$CH_3CH_2CH_2CO\text{–}S\text{–}ACP + NADP^+$
n-Butyryl-S-ACP

We now have a straight-chain saturated fatty acid, and with this the cycle begins again: reaction of it with malonyl-S-ACP, decarboxylation, reduction, dehydration, hydrogenation. After seven such cycles we arrive at the 16-carbon acid, palmitic acid—and here, for some reason, the process stops. Additional carbons can be added, but by a different process. Double bonds can be introduced, to produce unsaturated acids. Finally, glycerol esters are formed: triacylglycerols, to be stored and, when needed, oxidized to provide energy; and phosphoglycerides (Sec. 33.8) to help make up cell walls.

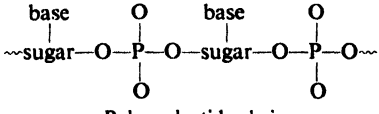
Enzymes are marvelous catalysts. Yet, even with their powerful help, these biological reactions seek the easiest path. In doing this, they take advantage of the same structural effects that the organic chemist does: the acidity of α-hydrogens, the leaving ability of a particular group, the ease of decarboxylation of β-keto acids.

37.7 Nucleoproteins and nucleic acids

In every living cell there are found **nucleoproteins**: substances made up of proteins combined with natural polymers of another kind, the **nucleic acids**. Of all fields of chemistry, the study of the nucleic acids is perhaps the most exciting, for these compounds are the substance of heredity. Let us look very briefly at the structure of nucleic acids and, then, in the next section, see how this structure may be related to their literally vital role in heredity.

Although chemically quite different, nucleic acids resemble proteins in a fundamental way: there is a long chain—a backbone—that is the same (except for length) in all nucleic acid molecules; and attached to this backbone are various groups, which by their nature and sequence characterize each individual nucleic acid.

Where the backbone of the protein molecule is a polyamide chain (a polypeptide chain), the backbone of the nucleic acid molecule is a polyester chain (called a *polynucleotide* chain). The ester is derived from phosphoric acid (the acid portion) and a sugar (the alcohol portion).



Polynucleotide chain

The sugar is D-ribose (p. 1086) in the group of nucleic acids known as ribonucleic

acids (RNA), and D-2-deoxyribose in the group known as deoxyribonucleic acids (DNA). (The prefix *2-deoxy* simply indicates the lack of an —OH group at the 2-position.) The sugar units are in the furanose form, and are joined to phosphate through the C–3 and C–5 hydroxyl groups (Fig. 37.4).

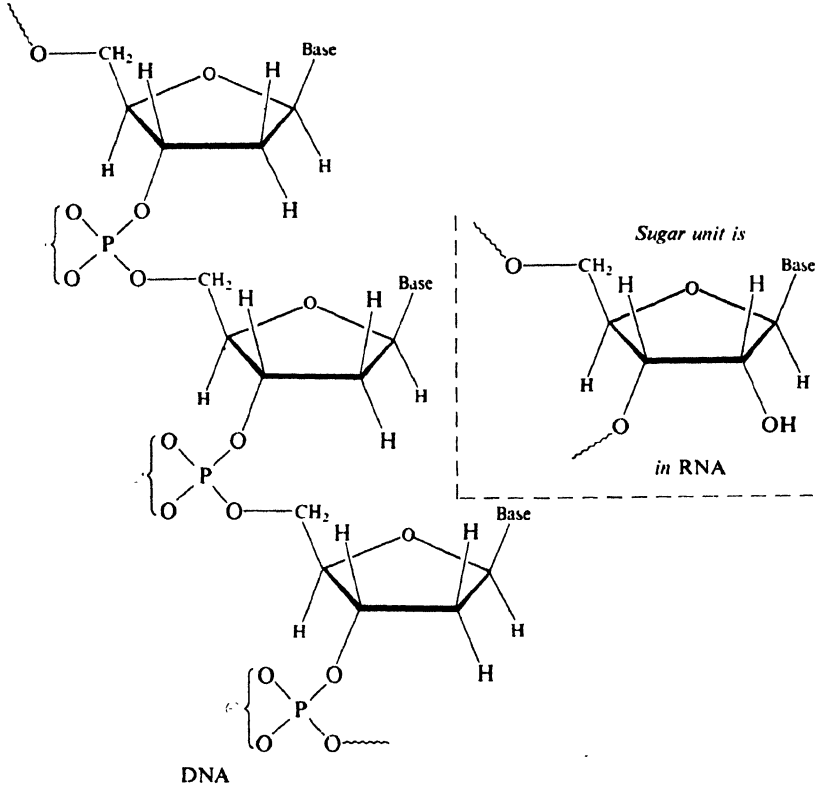


Figure 37.4. Deoxyribonucleic acid (DNA) and ribonucleic acid (RNA).

Attached to C–1 of each sugar, through a β-linkage, is one of a number of heterocyclic bases. A base–sugar unit is called a *nucleoside*; a base–sugar–phosphoric acid unit is called a *nucleotide*. An example of a nucleotide is shown in Fig. 37.5.

The bases found in DNA are *adenine* and *guanine*, which contain the purine ring system, and *cytosine*, *thymine*, and *5-methylcytosine*, which contain the pyrimidine ring system. RNA contains adenine, guanine, cytosine, and *uracil*. (See Fig. 37.6.)

The proportions of these bases and the sequence in which they follow each other along the polynucleotide chain differ from one kind of nucleic acid to another. This primary structure is studied in essentially the same way as the structure of proteins: by hydrolytic degradation and identification of the fragments. In this way, and after *seven years* of work, Robert W. Holley and his collaborators at

Cornell University. determined the exact sequence of the 77 nucleotides in the molecule of one kind of transport RNA (p. 1181).

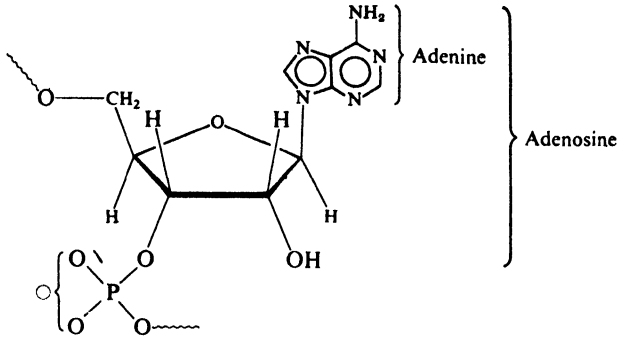


Figure 37.5. A nucleotide: an adenylic acid unit of RNA. Here, the nucleoside is adenosine, and the heterocyclic base is adenine.

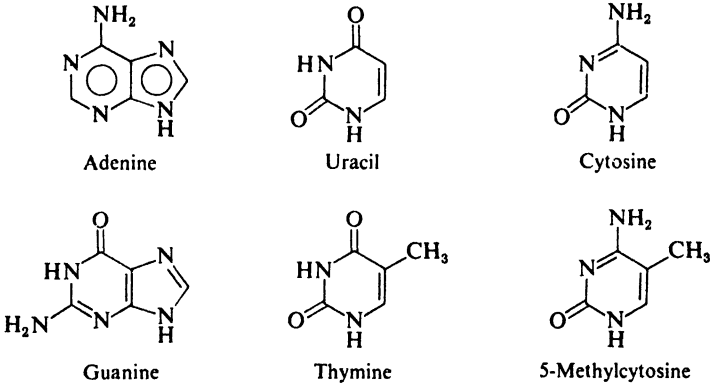


Figure 37.6. The heterocyclic bases of DNA and RNA.

What can we say about the secondary structure of nucleic acids? The following picture of DNA fits both chemical and x-ray evidence. Two polynucleotide chains, identical but heading in opposite directions, are wound about each other to form a double helix 18 A in diameter (shown schematically in Fig. 37.7). Both helixes are right-handed and have ten nucleotide residues per turn.

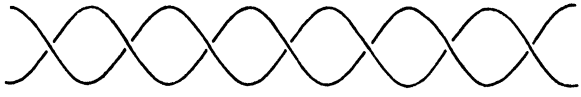


Figure 37.7. Schematic representation of the double helix structure proposed for DNA. Both helixes are right-handed and head in opposite directions; ten residues per turn. Hydrogen bonding between the helixes.

The two helixes in DNA are held to each other at intervals by hydrogen bonding between bases. From study of molecular models, it is believed that these hydrogen bonds can form only between adenine and thymine and between guanine and cytosine; hydrogen bonding between other pairs of bases would not allow them to fit into the double helical structure. In agreement with this idea, the adenine:thymine and guanine:cytosine ratios are found to be 1:1.

Less is known about structures of the various kinds of RNA, although here, too, helixes are involved. In 1973, the precise shape of one RNA molecule—the transport RNA that delivers phenylalanine—was reported: two short segments of double helix at right angles to each other and held together by two loops, the whole making a sort of four-leaf-clover pattern.

So far we have discussed only the nucleic acid portion of nucleoproteins. There is evidence that in one nucleoprotein (found in fish sperm), a polyarginine chain lies in one of the grooves of the double helix, held by electrostatic forces between the negative phosphate groups of the polynucleotide (which face the outside of the helix) and the positive guanidium groups of the arginine residues.

37.8 Chemistry and heredity. The genetic code

Just how is the structure of nucleic acids related to their function in heredity? Nucleic acids control heredity *on the molecular level.* The double helix of DNA is the repository of the hereditary information of the organism. The information is stored as the sequence of bases along the polynucleotide chain; it is a message "written" in a language that has only four letters, A, G, T, C (adenine, guanine, thymine, cytosine).

DNA must both *preserve* this information and *use* it. It does these things through two properties: (a) DNA molecules can duplicate themselves, that is, can bring about the synthesis of other DNA molecules identical with the originals; and (b) DNA molecules can control the synthesis, in an exact and specific way, of the proteins that are characteristic of each kind of organism.

First, there is the matter of self-duplication. The sequence of bases in one chain of the double helix controls the sequence in the other chain. The two chains fit together (as F. H. C. Crick of Cambridge University puts it) like a hand and a glove. They separate, and about the hand is formed a new glove, and inside the glove is formed a new hand. Thus, the pattern is preserved, to be handed down to the next generation.

Next, there is the matter of guiding the synthesis of proteins. A particular sequence of bases along a polynucleotide chain leads to a particular sequence of amino acid residues along a polypeptide chain. A protein has been likened to a long sentence written in a language of 20 letters: the 20 different amino acid residues. But the hereditary message is written in a language of only four letters; it is written in a *code*, with each word standing for a particular amino acid.

The genetic code has been broken, but research continues, aimed at tracking down the lines of communication. DNA serves as a template on which molecules of RNA are formed. It has been suggested that the double helix of DNA partially uncoils, and about the individual strands are formed chains of RNA; the process thus resembles self-duplication of DNA, except that these new chains contain

ribose instead of deoxyribose. The base sequence along the RNA chain is different from that along the DNA template, but is determined *by it*: opposite each adenine of DNA, there appears on RNA a uracil; opposite guanine, cytosine; opposite thymine, adenine; opposite cytosine, guanine. Thus, AATCAGTT on DNA becomes UUAGUCAA on RNA.

One kind of RNA—called, fittingly, *messenger RNA*—carries a message to the ribosome, where protein synthesis actually takes place. At the ribosome, messenger RNA calls up a series of *transport RNA* molecules, each of which is loaded with a particular amino acid. The order in which the transport RNA molecules are called up—the sequence in which the amino acids are built into the protein chain—depends upon the sequence of bases along the messenger RNA chain. Thus, GAU is the code for aspartic acid; UUU, phenylalanine; GUG, valine. There are 64 three-letter code words (*codons*) and only 20-odd amino acids, so that more than one codon can call up the same amino acids: CUU and CUC, leucine; GAA and GAG, glutamic acid.

A difference of a single base in the DNA molecule, or a single error in the "reading" of the code can cause a change in the amino acid sequence. The tiny defect in the hemoglobin molecule that results in sickle-cell anemia (p. 1152) has been traced to a single gene—a segment of the DNA chain—where, perhaps, the codon GUG appears instead of GAG. There is evidence that antibiotics, by altering the ribosome, cause misreading of the code and death to the organism.

Thus, the structure of nucleic acid molecules determines the structure of protein molecules. The structure of protein molecules, we have seen, determines the way in which they control living processes. Biology is becoming more and more a matter of shapes and sizes of molecules.

At the beginning of this book, we said that the structural theory is the basis of the science of organic chemistry. It is much more than that: the structural theory is the basis of our understanding of life.

PROBLEMS

1. Carbon dioxide is required for the conversion of acetyl CoA into fatty acids. Yet when carbon dioxide labeled with ^{14}C is used, none of the labeled carbon appears in the fatty acids that are formed. How do you account for these facts?

2. Taken together, what do these two facts show about chymotrypsin action? (a) The two esters, *p*-nitrophenyl acetate and *p*-nitrophenyl thiolacetate, $p\text{-}NO_2C_6H_4SCOCH_3$, undergo chymotrypsin-catalyzed hydrolysis at the same rate and with the same pH-dependence of rate, despite the fact that —SR is a much better leaving group than —OR. (b) There is no oxygen exchange (Sec. 20.17) in chymotrypsin-catalyzed hydrolysis of an ester RCOOR .

3. In DNA, the bases are bonded to deoxyribose at the following positions (that is, a hydrogen in Fig. 37.6, p. 1179, is replaced by C–1 of the sugar): adenine and guanine, NH in the five-membered ring; cytosine and thymine, NH.

(a) Draw structures to show likely hydrogen bonding between adenine and thymine; between guanine and cytosine. (b) Can you account for the fact that guanine and cytosine pairs hold the chains together more strongly than do adenine and thymine pairs?