3 Nucleic Acids, Proteins, and Enzymes

KEY CONCEPTS

The bark of the willow tree (*Salix alba*) was the original source of salicylic acid, later modified to aspirin.

Despite suffering from the "ague," the Reverend Edward Stone went walking in the English countryside. Feverish, tired, with aching muscles and joints, he came across a willow tree. Although apparently unaware that many ancient healers used willow bark extracts to reduce fever, the clergyman knew of the tradition of natural remedies for various diseases. The willow reminded him of the bitter extracts from the bark of South American trees then being sold (at high prices) to treat fevers. Removing some willow bark, Stone sucked on it and found it did indeed taste bitter—and that it relieved his symptoms.

Later he gathered a pound of willow bark and ground it into a powder, which he gave to about 50 people who complained of pain; all said they felt better. Stone reported the results of this "clinical test" in a letter to the Royal Society, England's most respected scientific body. Stone had discovered the main source of salicylic acid, the basis of the most widely used drug in the world. The date of his letter (which still exists) was April 25, 1763.

The chemical structure of salicylic acid (named for *Salix*, the willow genus) was worked out about 70 years later, and soon chemists could synthesize it in the laboratory. Although the compound alleviated pain, its acidity irritated the digestive system. In the late 1890s, the German chemical company Bayer synthesized a milder yet equally effective form, acetylsalicylic acid, which it marketed as aspirin. The new medicine's success launched Bayer to world prominence as a pharmaceutical company, a position it maintains today.

In the 1960s and 1970s, aspirin use declined when two alternative medications, acetaminophen (Tylenol) and ibuprofen (Motrin and Advil), became widely available. But over this same time, clinical studies revealed a new use for aspirin: it is an effective anticoagulant, shown to prevent heart attacks and strokes caused by blood clots. Today many people take a daily low dose of aspirin as a preventive agent against clotting disorders.

Fever, joint pain, headache, blood clots. What do these symptoms have in common? They all are mediated by fatty acid products called prostaglandins and molecules derived from them. Salicylic acid blocks the synthesis of the primary prostaglandin. The biochemical mechanism by which aspirin works was described in 1971. As we will see, an understanding of this mechanism requires an understanding of protein and enzyme function—two subjects of this chapter.

Q How does an understanding of proteins and enzymes help explain how aspirin works?

You will find the answer to this question on page 57.

CONCEPT 3.1 Nucleic Acids Are Informational Macromolecules

Nucleic acids are polymers that store, transmit, and express hereditary (genetic) information. This information is encoded in the sequences of monomers that make up nucleic acids. There are two types of nucleic acids: **DNA** (*d*eoxyribo*n*ucleic *a*cid) and **RNA** (*r*ibo*n*ucleic *a*cid). DNA stores and transmits genetic information. Through RNA intermediates, the information encoded in DNA is used to specify the amino acid sequences of proteins. As you will see later in this chapter, proteins are essential for both metabolism and structure. Certain specialized RNA molecules also play roles in metabolism. Ultimately, *nucleic acids and the proteins encoded by them determine the metabolic functions of an organism.*

Nucleotides are the building blocks of nucleic acids

Nucleic acids are polymers composed of monomers called nucleotides. A **nucleotide** consists of three components: a nitrogen-containing **base**, a pentose sugar, and one to three phosphate groups (FIGURE 3.1). Molecules consisting of a pentose sugar and a base—but no phosphate group—are called nucleosides. The nucleotides that make up nucleic acids contain just one phosphate group—they are nucleoside monophosphates.

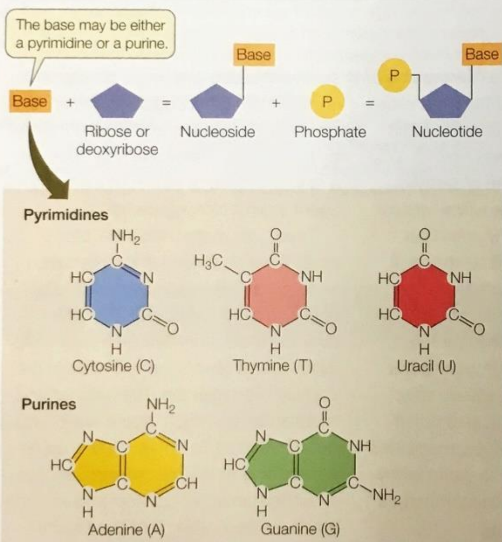


FIGURE 3.1 Nucleotides Have Three Components Nucleotide monomers are the building blocks of DNA and RNA polymers. Nucleotides may have one to three phosphate groups; those in DNA and RNA have one.

Go to ACTIVITY 3.1 Nucleic Acid Building Blocks
PoL2e.com/ac3.1

TABLE 3.1 Distinguishing RNA from DNA

Nucleic acid	Sugar	Bases	Strands
RNA	Ribose	Adenine Cytosine Guanine Uracil	Single
DNA	Deoxyribose	Adenine Cytosine Guanine Thymine	Double

The bases of the nucleic acids take one of two chemical forms: a six-membered single-ring structure called a **pyrimidine**, or a fused double-ring structure called a **purine** (see Figure 3.1). In DNA, the pentose sugar is **deoxyribose**, which differs from the **ribose** found in RNA by the absence of one oxygen atom (see Figure 2.9).

During the formation of a nucleic acid, new nucleotides are added to an existing chain one at a time. The pentose sugar in the last nucleotide of the existing chain and the phosphate on the new nucleotide undergo a condensation reaction (see Figure 2.8) and the resulting linkage is called a **phosphodiester bond**. The phosphate on the new nucleotide is attached to the 5′ (5 prime) carbon atom of its sugar, and the bond occurs between it and the 3′ (3 prime) carbon on the last sugar of the existing chain. Because each nucleotide is added to the 3′ carbon of the last sugar, nucleic acids are said to *grow in the 5′ to 3′ direction* (FIGURE 3.2).

Nucleic acids can be oligonucleotides, with a few to about 20 nucleotide monomers, or longer polynucleotides:

- *Oligonucleotides* include RNA molecules that function as "primers" to begin the duplication of DNA; RNA molecules that regulate the expression of genes; and synthetic DNA molecules used for amplifying and analyzing other, longer nucleotide sequences.
- *Polynucleotides*, more commonly referred to as nucleic acids, include DNA and most RNA. Polynucleotides can be very long, and indeed are the longest polymers in the living world. Some DNA molecules in humans contain hundreds of millions of nucleotides.

Base pairing occurs in both DNA and RNA

In addition to differing in their sugar groups, DNA and RNA also differ in their bases and general structures (TABLE 3.1). Four bases are found in DNA: **adenine (A)**, **cytosine (C)**, **guanine (G)**, and **thymine (T)**. RNA also contains adenine, cytosine, and guanine, but the fourth base in RNA is **uracil (U)** rather than thymine. The lack of a hydroxyl group at the 2′ position of the deoxyribose sugar in DNA makes the structure of DNA less flexible than that of RNA. As we describe below, DNA is composed of two polynucleotide strands whereas RNA is usually single-stranded. However, a long RNA can fold up on itself, forming a variety of structures.

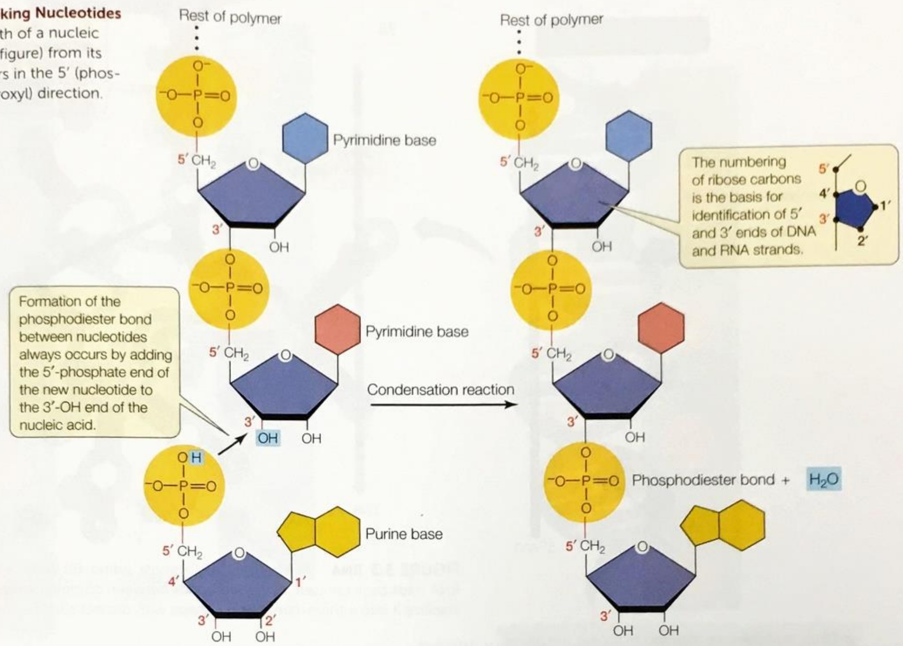
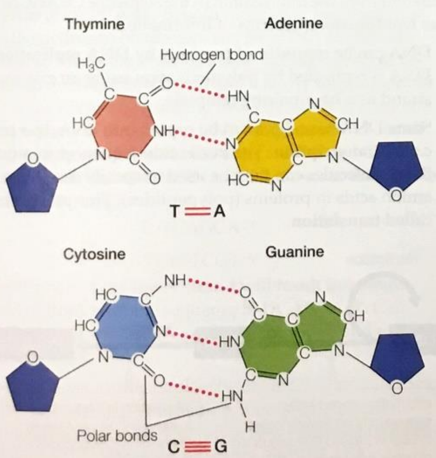


FIGURE 3.2 Linking Nucleotides Together Growth of a nucleic acid (RNA in this figure) from its monomers occurs in the 5′ (phosphate) to 3′ (hydroxyl) direction.

The key to understanding the structure and function of both DNA and RNA is the principle of **complementary base pairing**. In DNA, adenine and thymine always pair (A-T), and cytosine and guanine always pair (C-G):

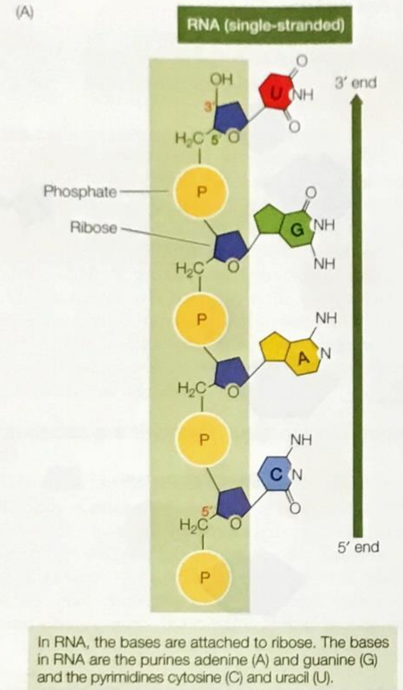


In RNA, the base pairs are A-U and C-G. Base pairs are held together primarily by hydrogen bonds. As you can see, there are polar C═O and N—H covalent bonds in the nucleotide bases (see Concept 2.2 for a discussion of polar covalent bonds). Hydrogen bonds form between the partial negative charge (δ^-) on an oxygen or nitrogen atom of one base, and the partial positive charge (δ^+) on a hydrogen atom of another base. Complementary base pairing occurs because the arrangements of polar bonds in the nucleotide bases favor the pairing of bases as they occur (C with G, and A with U or T).

Individual hydrogen bonds are relatively weak, but there are so many of them in DNA and RNA that collectively they provide a considerable force of attraction. However, this attraction is not as strong as that provided by multiple covalent bonds. This means that base pairs are relatively easy to separate with a modest input of energy. As you will see in Chapters 9 and 10, the breaking and making of hydrogen bonds in nucleic acids is vital to their roles in living systems. Let's now look in a little more detail at the structures of RNA and DNA.

RNA Usually, RNA is single-stranded (**FIGURE 3.3A**). However, many single-stranded RNA molecules fold up into three-dimensional structures, because of hydrogen bonding between nucleotides in separate portions of the molecules (**FIGURE 3.3B**). An RNA strand can also fold back on itself to form a double-stranded helix. This results in a three-dimensional surface for the bonding and recognition of other molecules. It is important to realize that this folding occurs by complementary base pairing, and the structure is thus determined by the particular order of bases in the RNA molecule.

DNA Usually, DNA is double-stranded; that is, it consists of two separate polynucleotide strands of the same length



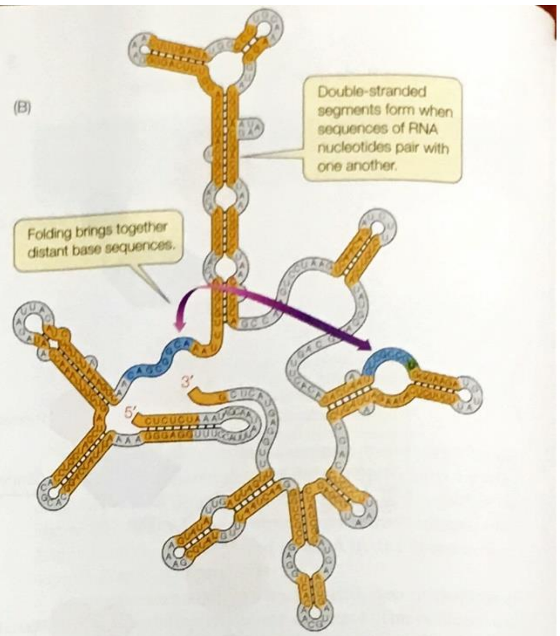
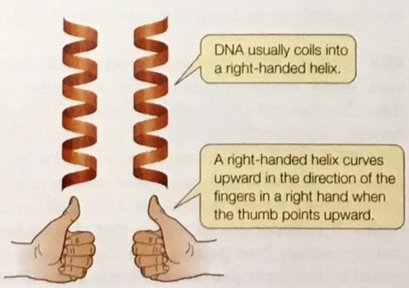


FIGURE 3.3 RNA (A) RNA is usually a single strand. (B) When a single-stranded RNA folds back on itself, hydrogen bonds between complementary sequences can stabilize it into a three-dimensional shape with distinct surface characteristics.

(**FIGURE 3.4A**). The two polynucleotide strands are antiparallel: they run in opposite directions so that their 5′ ends are at opposite ends of the double-stranded molecule. In contrast to RNA's diversity in three-dimensional structure, DNA is remarkably uniform. The A-T and G-C base pairs are about the same size (each is a purine paired with a pyrimidine), and the two polynucleotide strands form a "ladder" that twists into a double helix (**FIGURE 3.4B**). The sugar–phosphate groups form the sides of the ladder, and the bases with their hydrogen bonds form the rungs on the inside. The double helix is almost always right-handed:

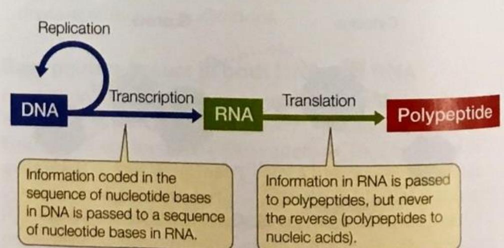


Go to ACTIVITY 3.2 DNA Structure
PoL2e.com/ac3.2

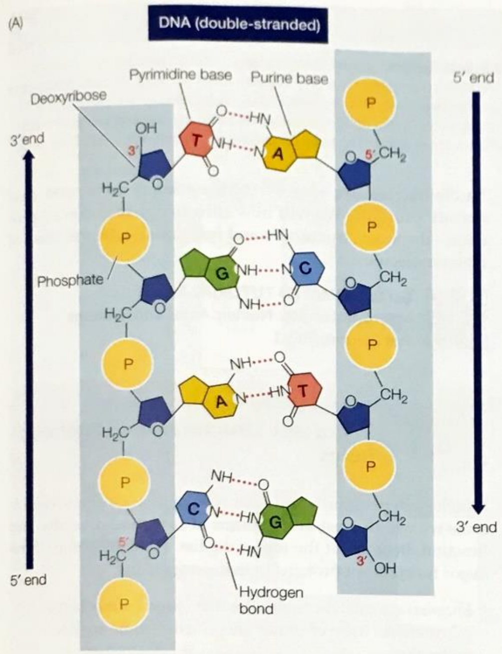
DNA carries information and is expressed through RNA

DNA is a purely informational molecule. The information is encoded in the sequence of bases carried in its strands. For example, the information encoded in the sequence TCAGCA is different from the information in the sequence CCAGCA. DNA has two functions in terms of information:

- DNA can be reproduced precisely by **DNA replication**. DNA is replicated by polymerization using an existing strand as a base-pairing template.
- Some DNA sequences can be copied into RNA, in a process called **transcription**. The nucleotide sequences in most RNA molecules can then be used to specify sequences of amino acids in proteins (polypeptides). This process is called **translation**.



The details of these important processes are described in Chapters 9 and 10, but it is important to realize several things at this point:



In DNA, the bases are attached to deoxyribose, and the base thymine (T) is found instead of uracil. Hydrogen bonds between purines and pyrimidines hold the two strands of DNA together.

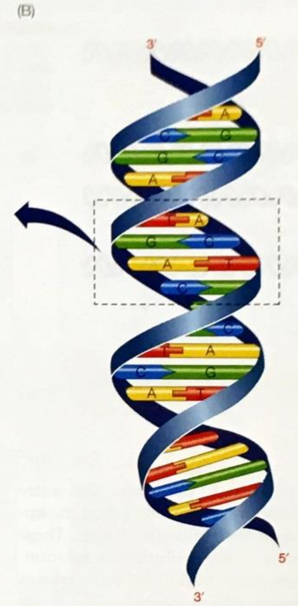


FIGURE 3.4 DNA (A) DNA usually consists of two strands running in opposite directions that are held together by base pairing between purines and pyrimidines opposite one another on the two strands. (B) The two antiparallel strands in a DNA molecule are twisted into a double helix.

- *DNA replication and transcription depend on the base pairing properties of nucleic acids.* In both replication and transcription, the hydrogen bonds between two DNA strands are broken, so that complementary base pairing can occur between an existing DNA strand and a newly forming strand of DNA or RNA. The resulting new DNA or RNA strand is *complementary to* the existing DNA template strand. Recall that the hydrogen-bonded base pairs are A-T and G-C in DNA and A-U and G-C in RNA. Now, consider this double-stranded DNA region:

  5′-TCAGCA-3′

  3′-AGTCGT-5′

  Transcription of the lower strand will result in a single strand of RNA with the sequence 5′-UCAGCA-3′. Can you figure out what RNA sequence the top strand would produce?
- *DNA replication usually involves the entire DNA molecule.* Since DNA holds essential information, it must be replicated completely so that each new cell or new organism receives a complete set of DNA from its parent (**FIGURE 3.5A**).
- *Gene expression is the transcription and translation of specific DNA sequences.* Sequences of DNA that encode specific proteins and are transcribed into RNA are called **genes** (**FIGURE 3.5B**). The complete set of DNA in a living organism is called its **genome**. However, not all of the information in the genome is needed at all times and in all tissues. For example, in humans, the gene that encodes the major protein in hair (keratin) is expressed only in skin cells. The genetic information in the keratin-encoding gene is transcribed into RNA and then translated into the protein keratin. In other tissues such as the muscles, the keratin gene is not transcribed, but other genes are—for example, the genes that encode proteins present in muscles but not in skin.

The DNA base sequence reveals evolutionary relationships

Because DNA carries hereditary information from one generation to the next, a theoretical series of DNA molecules stretches back through the lineage of every organism to the beginning of biological evolution on Earth, about 3.8 billion years ago. The genomes of organisms gradually accumulate changes in their DNA base sequences over evolutionary time. Therefore closely related living species should have more similar base sequences than species that are more distantly related.

Over the past two decades there have been remarkable developments in technologies for determining the order of nucleotides in DNA molecules (DNA sequencing), and in computer technologies to analyze these sequences. These

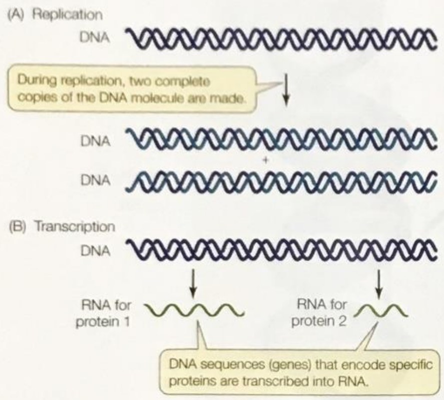


FIGURE 3.5 DNA Replication and Transcription DNA is completely replicated during cell reproduction (A), but it is only partially transcribed (B). In transcription, the DNA code is copied to RNA. The sequence of the latter determines the amino acid sequence of a protein. Transcription of the genes for many different proteins is activated at different times and, in multicellular organisms, in different cells of the body.

advances have enabled scientists to determine the entire DNA base sequences of whole organisms, including the human genome, which contains about 3 billion base pairs. These studies have confirmed many of the evolutionary relationships that were inferred from more traditional comparisons of body structure, biochemistry, and physiology. Traditional comparisons had indicated that the closest living relative of humans (*Homo sapiens*) is the chimpanzee (genus *Pan*). In fact, the chimpanzee genome shares nearly 99 percent of its DNA base sequence with the human genome. Increasingly, scientists turn to DNA analyses to figure out evolutionary relationships when other comparisons are not possible or are not conclusive. For example, DNA studies revealed a close relationship between starlings and mockingbirds that was not expected on the basis of their anatomy or behavior.

LINK

For more on the use of DNA sequences to reconstruct the evolutionary history of life, see **Concept 16.2**

CHECKpoint CONCEPT 3.1

- ✓ List the key differences between DNA and RNA and between purines and pyrimidines.
- ✓ What are the differences between DNA replication and transcription?
- ✓ If one strand of a DNA molecule has the sequence 5′-TTCCGGAT-3′, what is the sequence of the other strand of DNA? If RNA is transcribed from the 5′-TTCCG-GAT-3′ strand, what would be its sequence? And if RNA is transcribed from the other DNA strand, what would be its sequence? (Note that it is conventional to write these sequences with the 5′ end on the left.)
- ✓ How can DNA molecules be so diverse when they appear to be structurally similar?

Nucleic acids are largely informational molecules that encode proteins. We will now turn to a discussion of proteins—the most structurally and functionally diverse class of macromolecules.

Go to ANIMATED TUTORIAL 3.1
Macromolecules: Nucleic Acids and Proteins
PoL2e.com/at3.1

CONCEPT 3.2 Proteins Are Polymers with Important Structural and Metabolic Roles

Proteins are the fourth and final type of biological macromolecule we will discuss, and in terms of structural diversity and function, they are at the top of the list. Here are some of the major functions of proteins in living organisms:

- *Enzymes* are catalytic molecules that speed up biochemical reactions. Most enzymes are proteins (some are RNA molecules).
- *Defensive proteins* such as antibodies recognize and respond to substances or particles that invade the organism from the environment.
- *Hormonal and regulatory proteins* such as insulin control physiological processes.
- *Receptor proteins* receive and respond to molecular signals from inside and outside the organism.
- *Storage proteins* store chemical building blocks—amino acids—for later use.
- *Structural proteins* such as collagen provide physical stability and enable movement.
- *Transport proteins* such as hemoglobin carry substances within the organism.
- *Genetic regulatory proteins* (transcription factors) regulate when, how, and to what extent a gene is expressed.

Clearly, the biochemistry of proteins warrants our attention!

Amino acids are the building blocks of proteins

As we noted in Chapter 2, **proteins** are polymers made up of monomers called **amino acids**. As their name suggests, the amino acids all contain two functional groups: the nitrogen-containing amino group and the (acidic) carboxyl group.