

## The Structure and Function of DNA

Biologists in the 1940s had difficulty in accepting DNA as the genetic material because of the apparent simplicity of its chemistry. DNA was known to be a long polymer composed of only four types of subunits, which resemble one another chemically. Early in the 1950s, DNA was first examined by x-ray diffraction analysis, a technique for determining the three-dimensional atomic structure of a molecule (discussed in Chapter 8). The early x-ray diffraction results indicated that DNA was composed of two strands of the polymer wound into a helix. The observation that DNA was double-stranded was of crucial significance and provided one of the major clues that led to the Watson-Crick structure of DNA. Only when this model was proposed did DNA's potential for replication and information encoding become apparent. In this section we examine the structure of the DNA molecule and explain in general terms how it is able to store hereditary information.

### A DNA Molecule Consists of Two Complementary Chains of Nucleotides

A DNA molecule consists of two long polynucleotide chains composed of four types of nucleotide subunits. Each of these chains is known as a *DNA chain*, or a *DNA strand*. *Hydrogen bonds* between the base portions of the nucleotides hold the two chains together (Figure 4-3). As we saw in Chapter 2 (Panel 2-6, pp. 120-121), nucleotides are composed of a five-carbon sugar to which are attached one or more phosphate groups and a nitrogen-containing base. In the case of the nucleotides in DNA, the sugar is deoxyribose attached to a single phosphate group (hence the name deoxyribonucleic acid), and the base may be either *adenine (A)*, *cytosine (C)*, *guanine (G)*, or *thymine (T)*. The nucleotides are covalently linked together in a chain through the sugars and phosphates, which thus form a “backbone” of alternating sugar-phosphate-sugar-phosphate (see Figure 4-3). Because only the base differs in each of the four types of subunits, each polynucleotide chain in DNA is analogous to a necklace (the backbone) strung with four types of beads (the four bases A, C, G, and T). These same symbols (A, C, G, and T) are also commonly used to denote the four different nucleotides—that is, the bases with their attached sugar and phosphate groups.



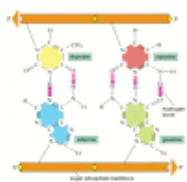
**Figure 4-3**

DNA and its building blocks. DNA is made of four types of nucleotides, which are linked covalently into a polynucleotide chain (a DNA strand) with a sugar-phosphate backbone from which the bases (A, C, G, and T) extend. A DNA molecule is composed of two (more...)

The way in which the nucleotide subunits are lined together gives a DNA strand a chemical polarity. If we think of each sugar as a block with a protruding knob (the 5' phosphate) on one side and a hole (the 3' hydroxyl) on the other (see Figure 4-3), each completed chain, formed by interlocking knobs with holes, will have all of its subunits lined up in the same orientation. Moreover, the two ends of the chain will be easily distinguishable, as one has a hole (the 3'

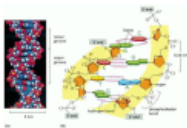
hydroxyl) and the other a knob (the 5' phosphate) at its terminus. This polarity in a DNA chain is indicated by referring to one end as the *3' end* and the other as the *5' end*.

The three-dimensional structure of DNA—the **double helix**—arises from the chemical and structural features of its two polynucleotide chains. Because these two chains are held together by hydrogen bonding between the bases on the different strands, all the bases are on the inside of the double helix, and the sugar-phosphate backbones are on the outside (see Figure 4-3). In each case, a bulkier two-ring base (a purine; see Panel 2-6, pp. 120–121) is paired with a single-ring base (a pyrimidine); A always pairs with T, and G with C (Figure 4-4). This *complementary base-pairing* enables the base pairs to be packed in the energetically most favorable arrangement in the interior of the double helix. In this arrangement, each base pair is of similar width, thus holding the sugar-phosphate backbones an equal distance apart along the DNA molecule. To maximize the efficiency of base-pair packing, the two sugar-phosphate backbones wind around each other to form a double helix, with one complete turn every ten base pairs (Figure 4-5).



**Figure 4-4**

Complementary base pairs in the DNA double helix. The shapes and chemical structure of the bases allow hydrogen bonds to form efficiently only between A and T and between G and C, where atoms that are able to form hydrogen bonds (see Panel 2-3, pp. 114–115) ([more...](#))



**Figure 4-5**

The DNA double helix. (A) A space-filling model of 1.5 turns of the DNA double helix. Each turn of DNA is made up of 10.4 nucleotide pairs and the center-to-center distance between adjacent nucleotide pairs is 3.4 nm. The coiling of the two strands around ([more...](#))

The members of each base pair can fit together within the double helix only if the two strands of the helix are antiparallel—that is, only if the polarity of one strand is oriented opposite to that of the other strand (see Figures 4-3 and 4-4). A consequence of these base-pairing requirements is that each strand of a DNA molecule contains a sequence of nucleotides that is exactly complementary to the nucleotide sequence of its partner strand.

## The Structure of DNA Provides a Mechanism for Heredity

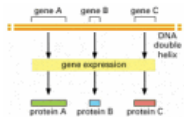
---

Genes carry biological information that must be copied accurately for transmission to the next generation each time a cell divides to form two daughter cells. Two central biological questions arise from these requirements: how can the information for specifying an organism be carried in chemical form, and how is it accurately copied? The discovery of the structure of the DNA double helix was a landmark in twentieth-century biology because it immediately suggested answers to both questions, thereby resolving at the molecular level the problem of heredity. We discuss briefly the answers to these questions in this section, and we shall examine them in more detail in subsequent chapters.

DNA encodes information through the order, or sequence, of the nucleotides along each strand. Each base—A, C, T, or G—can be considered as a letter in a four-letter alphabet that spells out

biological messages in the chemical structure of the DNA. As we saw in Chapter 1, organisms differ from one another because their respective DNA molecules have different nucleotide sequences and, consequently, carry different biological messages. But how is the nucleotide alphabet used to make messages, and what do they spell out?

As discussed above, it was known well before the structure of DNA was determined that genes contain the instructions for producing proteins. The DNA messages must therefore somehow encode proteins (Figure 4-6). This relationship immediately makes the problem easier to understand, because of the chemical character of proteins. As discussed in Chapter 3, the properties of a protein, which are responsible for its biological function, are determined by its three-dimensional structure, and its structure is determined in turn by the linear sequence of the amino acids of which it is composed. The linear sequence of nucleotides in a gene must therefore somehow spell out the linear sequence of amino acids in a protein. The exact correspondence between the four-letter nucleotide alphabet of DNA and the twenty-letter amino acid alphabet of proteins—the genetic code—is not obvious from the DNA structure, and it took over a decade after the discovery of the double helix before it was worked out. In Chapter 6 we describe this code in detail in the course of elaborating the process, known as *gene expression*, through which a cell translates the nucleotide sequence of a gene into the amino acid sequence of a protein.



**Figure 4-6**

The relationship between genetic information carried in DNA and proteins.

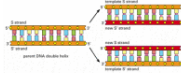
The complete set of information in an organism's DNA is called its genome, and it carries the information for all the proteins the organism will ever synthesize. (The term genome is also used to describe the DNA that carries this information.) The amount of information contained in genomes is staggering: for example, a typical human cell contains 2 meters of DNA. Written out in the four-letter nucleotide alphabet, the nucleotide sequence of a very small human gene occupies a quarter of a page of text (Figure 4-7), while the complete sequence of nucleotides in the human genome would fill more than a thousand books the size of this one. In addition to other critical information, it carries the instructions for about 30,000 distinct proteins.

**Figure 4-7**

The nucleotide sequence of the human  $\beta$ -globin gene. This gene carries the information for the amino acid sequence of one of the two types of subunits of the hemoglobin molecule, which carries oxygen in the blood. A different gene, the  $\alpha$ -globin (more...)

```
CCCTGTTGAGGACACCTAGGTTGGGCA
ATTTACTCTGAGAGACAGAGGCGGAGG
CGAGGCGTGGATAAAGGAGGCGAGAG
CCATCTATGCTTAGACCTGGCTTGTACT
AACTGTTGCTACTAGAACCTAAGAGACA
CCCTGTGTGACCTGACTCTGGGAGAAAT
CTGGCTTTTACTGCCCTTTGGGGAAAGCG
ACCTGATGATGATTGGGTGAGCCCTGG
GCGATTTGATACAAATTACAAAGACAT
TTAAGGAGACAAATGAAACTGGGCGGCG
GGAGACAGAGGACTCTGGGTTCTGAGG
GGCTACTACTCTTCTGCTATGTTGTCAT
TTTCTTACCTTTCATATTGGGTGGCTGG
CCCTGAGCGAGAGTCTTGAATGTCCTT
GGGATCTGGCACCTCTGAGGCTTGTGAA
GGACCCCAAGGTTGAGGCTCAGGAGAG
AAAGTCTGGCTGCTTTTGGGATGAGCTT
GTACTTGGCAAGCTTGAAGACTGGCTT
GGTACCTGACTGAGCTGACTGTGACAG
GTGACCTGGATCTGGAATCTGAGGCTG
AGTCTATGGAGCTTGATGTTCTCTTGC
CCCTCTTCTATGTTTAACTGAGGCGAG
AGGAGGCGAGAACTAGAGGAGGCGTTC
AGAACTGGGAAAGAGCAAAATGATCCAG
GTGTTGAGCCCTGAGGAGGTTTGAATTC
TTTTATTGCGTCTATAGAAATGGTTC
TTTGTGTGATCTTGGCTTGCCTTCTTTC
CTTTCGGATTTCTACTGTGAGGCTGGAA
TGCCTTACATTTGATACAAAAGGAAA
TATCTCTGAGACATTAATGAACTTAAA
AAAGCTTACACGCTGGCTATGAGCTT
ACTTTTGGAGAACTGTGTGTCTGTTC
ABATGATACTCCCTACTTGTATTTTC
TATTTTATTTGATGATGATGATGATAC
AATTTATGCTTAAAGGCTGATGTTTAA
TGGGCTGACATATTGACAAATGAGCT
AATTTCTGATTTGATTTTAAAGATCT
TCTCTCTTAAATACCTTTTGTGAACT
TTATTCTAAACTTCCCTAAATCTTTC
TTTCAAGGCAATATGATCAATATGCTE
GCTCTTTGACGCTTTGAGGATAGAG
TATTAATCTGGTTAGGCAATGCACT
ATTCTGCTATAAATTTCTGATAGAA
ATTATGCTGATTTAAGGTTCTGATG
CTAATAGGCTAGATCAGCTAGCTGTC
TCTTTCATTTTGGTGGAGGAGCTGC
GATTAATCTGAGCTGACATAGCCCTTC
GCTAATGAGTTGATGCTGATCTGCTC
CCGACCCCTGAGACAGCTGCTGCTC
GGCTGGCCCACTTCTGAAAAAATT
GGCTCAGCGCTGAGGCTGCTATGAA
AATGTCAGCTGCTGCTGAGCAAGG
CCGCAAGTACACTGACTGATCTGTGC
TGTGAAATTTCTATTAAGGTTCTTGT
CCCTAGCTGACTACTAAGCTGGGATG
TTGAGAGGCTTTGAGCTGTGATGCTT
CCGACAAAAGACATTTTATGCTGAA
TGAATTTTAAATTTCTGAAATTTT
ACTAAAAGGGAATGGGAGTCAAGGCA
TTTTAACTAAGAAATGATGCTGTTC
AACTTTGAAATCTACTATCTTAAA
CTCATSAAAGAGGTGAGGTGAGCTGAG
CTAATGCACTTGACAGGCTCTGAGTC
CTATGCTTACTGATCTCTGAAAGGAT
CTCTGTAGGCTTTGTTTGGGCTTAAAG
TTTCTAGCTGATTTACTACTGCTT
TCTTTAGCTGCTCTGAAATGCTGTTC
```

At each cell division, the cell must copy its genome to pass it to both daughter cells. The discovery of the structure of DNA also revealed the principle that makes this copying possible: because each strand of DNA contains a sequence of nucleotides that is exactly complementary to the nucleotide sequence of its partner strand, each strand can act as a template, or mold, for the synthesis of a new complementary strand. In other words, if we designate the two DNA strands as S and S', strand S can serve as a template for making a new strand S', while strand S' can serve as a template for making a new strand S (Figure 4-8). Thus, the genetic information in DNA can be accurately copied by the beautifully simple process in which strand S separates from strand S', and each separated strand then serves as a template for the production of a new complementary partner strand that is identical to its former partner.



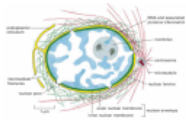
**Figure 4-8**

DNA as a template for its own duplication. As the nucleotide A successfully pairs only with T, and G with C, each strand of DNA can specify the sequence of nucleotides in its complementary strand. In this way, double-helical DNA can be copied precisely. *(more...)*

The ability of each strand of a DNA molecule to act as a template for producing a complementary strand enables a cell to copy, or *replicate*, its genes before passing them on to its descendants. In the next chapter we describe the elegant machinery the cell uses to perform this enormous task.

### In Eucaryotes, DNA Is Enclosed in a Cell Nucleus

Nearly all the DNA in a eucaryotic cell is sequestered in a nucleus, which occupies about 10% of the total cell volume. This compartment is delimited by a *nuclear envelope* formed by two concentric lipid bilayer membranes that are punctured at intervals by large nuclear pores, which transport molecules between the nucleus and the cytosol. The nuclear envelope is directly connected to the extensive membranes of the endoplasmic reticulum. It is mechanically supported by two networks of intermediate filaments: one, called the *nuclear lamina*, forms a thin sheetlike meshwork inside the nucleus, just beneath the inner nuclear membrane; the other surrounds the outer nuclear membrane and is less regularly organized (Figure 4-9).



**Figure 4-9**

A cross-sectional view of a typical cell nucleus. The nuclear envelope consists of two membranes, the outer one being continuous with the endoplasmic reticulum membrane (see also Figure 12-9). The space inside the endoplasmic reticulum (the ER lumen) ([more...](#))

The nuclear envelope allows the many proteins that act on DNA to be concentrated where they are needed in the cell, and, as we see in subsequent chapters, it also keeps nuclear and cytosolic enzymes separate, a feature that is crucial for the proper functioning of eucaryotic cells. Compartmentalization, of which the nucleus is an example, is an important principle of biology; it serves to establish an environment in which biochemical reactions are facilitated by the high concentration of both substrates and the enzymes that act on them.

## Summary

---

Genetic information is carried in the linear sequence of nucleotides in DNA. Each molecule of DNA is a double helix formed from two complementary strands of nucleotides held together by hydrogen bonds between G-C and A-T base pairs. Duplication of the genetic information occurs by the use of one DNA strand as a template for formation of a complementary strand. The genetic information stored in an organism's DNA contains the instructions for all the proteins the organism will ever synthesize. In eucaryotes, DNA is contained in the cell nucleus.

By agreement with the publisher, this book is accessible by the search feature, but cannot be browsed.

Copyright © 2002, Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter; Copyright © 1983, 1989, 1994, Bruce Alberts, Dennis Bray, Julian Lewis, Martin Raff, Keith Roberts, and James D. Watson .

Bookshelf ID: NBK26821