Human genome

DNA Structure: The Double Helix

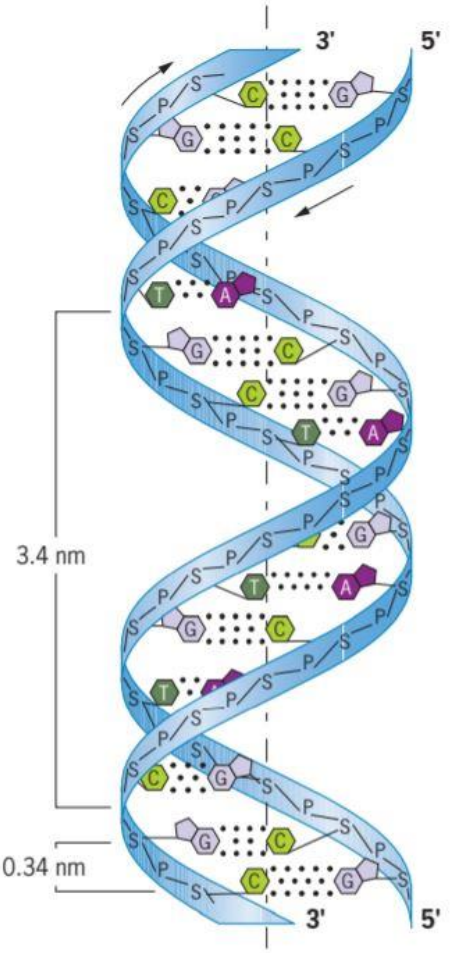
Gregor Mendel, an Austrian monk, concluded in 1865 from genetic crossing experiments that hereditary information is passed from parents to offspring in discrete packets. Later on, these units of heredity were called genes. The biological nature of genes was not revealed until 1944, when it was demonstrated that genetic information is carried by a biological substance with the chemical designation deoxyribonucleic acid (DNA). DNA is a polymer composed of deoxyribonucleotides, each of which typically contains one of four different bases: adenine (A), guanine (G), thymine (T), or cytosine (C). The discovery of the double-helical structure of DNA by James Watson and Francis Crick in 1953, certainly one of the most significant scientific discoveries of the 20th century, provided the first clues on how DNA can be accurately duplicated in order to transmit genomic information from one generation to the next. Watson and Crick's double-helix structure was based on two major kinds of evidence:

1. When Erwin Chargaff and colleagues analyzed the composition of DNA from many different organisms, they found that the concentration of thymine was always equal to the concentration of adenine and the concentration of cytosine was always equal to the concentration of guanine. Their results strongly suggested that thymine and adenine as well as cytosine and guanine were present in DNA in some fixed interrelationship. Their data also showed that the total concentration of pyrimidines (thymine plus cytosine) was always equal to the total concentration of purines (adenine plus guanine).

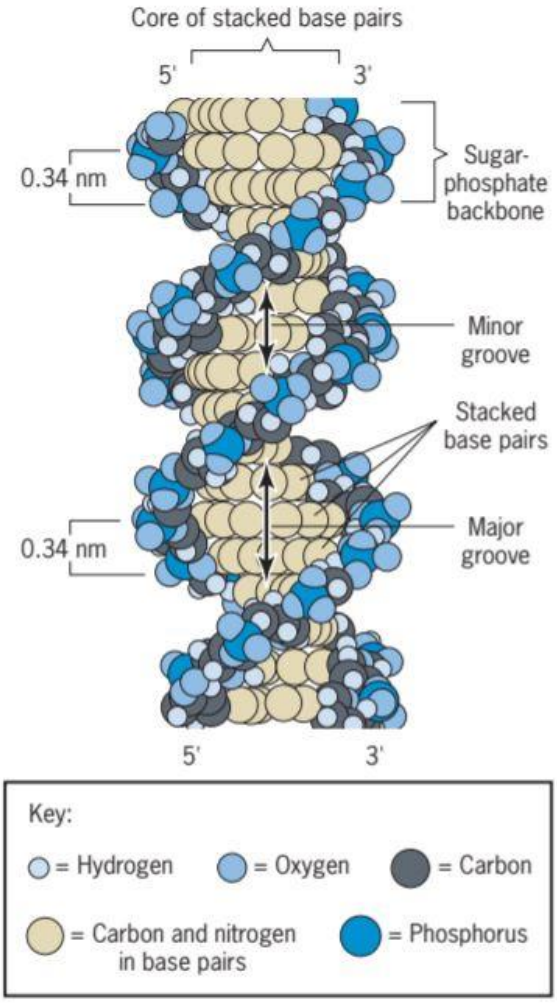
2. When X rays are focused through fibers of purified molecules, the rays are deflected by the atoms of the molecules in specific patterns, called diffraction patterns, which provide information about the organization of the components of the molecules. These X-ray diffraction patterns can be recorded on X-ray-sensitive film just as patterns of light can be recorded with a camera and light-sensitive film. Watson and Crick used X-ray diffraction data on DNA structure provided by Maurice Wilkins, Rosalind Franklin, and their coworkers. These data indicated that DNA was a highly ordered, two-stranded structure with repeating substructures spaced every 0.34 nanometer along the axis of the molecule.

Within the DNA double helix structure, two DNA strands closely interact via hydrogen bonds between their nucleotide bases to form specific inter-strand base pairs: adenine pairs with thymine and guanine pairs with cytosine. Hence, the two strands of the DNA double helix are complementary to each other and can both serve as templates during replication, the process

by which cells copy their genetic information before cell division. Complete replication of a DNA molecule results in two identical DNA duplexes, each consisting of one parental and one newly synthesised DNA strand. The base pairs in DNA are stacked about 0.34 nm apart, with 10 base pairs per turn (360^{o}) of the double helix. The sugar-phosphate backbones of the two complementary strands are antiparallel. Unidirectionally along a DNA double helix, the phosphodiester bonds in one strand go from a 3’ carbon of one nucleotide to a 5’ carbon of the adjacent nucleotide, whereas those in the complementary strand go from a 5’ carbon to a 3’ carbon. This “opposite polarity” of the complementary strands of a DNA double helix plays an important role in DNA replication, transcription, and recombination. The stability of DNA double helices results in part from the large number of hydrogen bonds between the base pairs (even though each hydrogen bond by itself is weak, much weaker than a covalent bond) and in part from the hydrophobic bonding (or stacking forces) between adjacent base pairs. The planar sides of the base pairs are relatively nonpolar and thus tend to be hydrophobic (water-insoluble). Because of this insolubility in water, the hydrophobic core of stacked base pairs contributes considerable stability to DNA molecules present in the aqueous protoplasms of living cells. The major groove, is much wider than the other, the minor groove. The difference between the major groove and the minor groove is important when one examines the interactions between DNA and proteins that regulate gene expression. Some proteins bind to the major groove; others bind to the minor groove. Genetic information is encoded within the nucleotide sequence of the DNA strands. The readout of genetic information occurs through a process called transcription, the synthesis of RNA, a single-stranded nucleic acid consisting of ribonucleotides, from a DNA template. Transcription is catalysed by RNA polymerases and results in a RNA molecule with the identical nucleotide sequence as the coding DNA strand of a gene. This RNA transcript can either have biological activity by itself or, as in the case of protein-coding genes, can be used as a template to synthesise a specific polypeptide chain from different amino acid components.



■ **FIGURE 9.10** Diagram of the double-helix structure of DNA.



■ **FIGURE 9.12** Space-filling diagram of a DNA double helix.

Genome organization

A genome is the sum total of all DNA within an organism. In addition to genomic DNA present in the nucleus, there is also DNA present in the mitochondria, semi-autonomous organelles found in most eukaryotic cells that serve as energy factories. The mitochondrial genome contributes to less than 1% of the total cellular DNA in mammalians. In the nuclei of eukaryotic cells, DNA is organised in a set of chromosomes. Human cells contain 22 pairs of autosomes; one chromosome of each pair is inherited from the mother and the other from the father. Since cells contain two copies of each autosome, they also contain two copies of each gene, the so-called alleles. In addition to the autosomes, there are the sex chromosomes X and Y. Females possess two X chromosomes, one from each parent, whereas males possess an X chromosome inherited from their mother and a Y chromosome inherited from their father.

The DNA within the 23 human chromosomes contains a total of 3.2×10^9 nucleotide bases (3.2 Gb) and, if completely stretched, would be approximately 2 m long. To fit the genomic DNA into the cell nucleus, which is only a few micrometres in diameter, an enormous level of compaction has to be achieved. Chromosomes represent the most compact form of nuclear DNA and are only observed during cell division. In non-dividing cells, so-called interphase cells, the DNA material occupies the cell nucleus as chromatin without distinguishable chromosomes. In chromatin, DNA is organised into nucleosomes; ~200 bp of DNA are wrapped around a protein disc containing a histone protein octamer. Nucleosomes are arranged like beads on a string to form a 10 nm chromatin fibre that can be further compacted to a 30 nm fibre by incorporating a specific linker histone protein, H1. As a 30 nm fibre, a human chromosome would still span the nucleus more than 100 times. The mechanism(s) underlying further condensation of the 30 nm fibre into higher order structures such as chromosomes are not yet fully understood. Two major types of chromatin can be distinguished based on the level of DNA condensation. Euchromatin occupies most of the nucleus and the underlying DNA fibres are much less densely packed as compared to heterochromatin, which exhibits a level of DNA compaction comparable to chromosomes. Of the 3.2 Gb of the human genome, 2.95 Gb or 92% are euchromatic and only 0.35 Gb or 8% are heterochromatic.