12.7: DNA and the Human Genome

In Chapter 7, Section 7.6, *deoxyribonucleic acid*, *DNA*, was discussed and it was noted that this macromolecule stores and passes on genetic information that organisms need to reproduce and synthesize proteins. Recall that DNA is composed of repeating units called *nucleotides* each consisting of a molecule of the sugar 2-deoxy-β-D-ribofuranose, a phosphate ion, and one of the four nitrogen-containing bases, adenine, cytosine, guanine, and, thymine (conventionally represented by the letters A, C, G, and T, respectively). DNA is one of two *nucleic acids*, the other one of which is *ribonucleic acid*, *RNA*. Like DNA, RNA consists of repeating nucleotides but the sugar in RNA is β-D-ribofuranose and it contains uracil instead of thymine in its bases. The structural formulas of segments of DNA and RNA are shown in Figure 7.7.

The structure of DNA is a key aspect of its function, and its elucidation by Watson and Crickin 1953 was a scientific insight that set off a revolution in biology that is going on to this day. The huge DNA molecules consist of two strands counter wound with each other and held together by hydrogen bonds; a representation of this structure is shown in Figure 7.8. In this structure, the hydrogen bonds connecting complementary bases on the two strands are represented by dashed lines. Because of their structures that make hydrogen bonding possible, adenine on one strand is always hydrogen bonded to thymine on the opposite strand and guanine to cytosine. During cell division the two strands of DNA unwind and each generates a complementary strand for the DNA of each new cell.

In organisms with eukaryotic cells, DNA is divided into units associated with protein molecules called **chromosomes**. The number of these varies with the organism; humans have 23pairs of chromosomes, a total of 46. The strands of DNA in chromosomes, in turn, are divided into sequences of nucleotides, each distinguished by the nitrogen-containing base in it. These sequences of nucleotides give directions for the synthesis of a specific kind of protein or polypeptide. (Proteins are the biological molecules that make up much of the structure of cells and that perform most of the key functions of living organisms. Polypeptide is a general term for polymers of amino acids; proteins are the relatively long-chain polypeptides.) These specific groups of nucleotides, each of which has a specific function, are called **genes**. When a particular protein is made, DNA produces a nucleic acid segment designated *m*RNA, which goes out into the cell and causes the protein to be formed through a process called **transcription** and **translation** (the gene is said to be **expressed**).

As the units that give the directions for protein synthesis, genes are obviously of the utmost importance in living organisms. As discussed in Section 12.8, genes can now be transferred between different kinds of organisms and will direct the synthesis of the protein for which they are designed in the recipient organism. It is now known that a number of human diseases are the result of defective genes, and there is a genetic tendency toward getting other kinds of diseases. For example, certain gene characteristics are involved in susceptibility to breast cancer.

Because of the known relationship of gene characteristics to disease, the decision was made in the mid-1980s to map all the genes in the human body. This collective body of genes is called the **human genome** and the project to map it is called the Human Genome Project. The original impetus for this project in the U. S. arose because of interest in the damage to human DNA by radiation, such as that from nuclear weapons. But, from the beginning, it was recognized that the project had enormous commercial potential, especially in the pharmaceutical industry, and could be very valuable in human health.

The announcement in 2001 that mapping of the human genome was complete promised great progress in biology, especially in medicine. Genes function by directing the synthesis of specific proteins and the action of most pharmaceuticals is to alter the activities of proteins, the drug's target. In some cases proteins are made more active and in others their activity is diminished. Knowledge of the human genome enables a better understanding of protein activity and should facilitate the development of more specific drugs, something that has developed only slowly. For example, the gene responsible for cystic fibrosis was discovered in 1989 by examination of family histories of the disease and only in 2010 were two drugs designed to combat that disease in clinical trials.

Genome Sequencing and Green Chemistry

The Human Genome Project and related genome sequencing of other organisms have a number of implications for green chemistry. One of the key goals of green chemistry is to use chemicals that have maximum effectiveness for their stated purpose with minimum side effects. This certainly applies to pharmaceuticals in which a knowledge of the human genome may enable development of drugs that do exactly what they are supposed to do without affecting non target systems. This means that drugs can be made very efficiently with little waste material.

Some of the most important effects of DNA sequencing as it relates to green chemistry has to do with a wide variety of organisms other than humans. With an exact knowledge of DNA and the genes that it contains, it is possible to deal with organisms on a highly scientific basis in areas such as pest control and the biosynthesis of raw materials. An accurate map of the genetic makeup of insects, for example, should result in the synthesis of precisely targeted insecticides which kill target pests without affecting other organisms. Such insecticides should be effective at very low doses, thus minimizing the amount of insecticide that has to be synthesized and applied, consistent with the goals of green chemistry.

An exact knowledge of the genomes of organisms is extremely helpful in the practice of genetic engineering in which genes are transferred between species to enable production of desired proteins and to give organisms desirable characteristics, such as pest resistance. A number of medically useful proteins and polypeptides are now produced by genetically engineered microorganisms, most commonly genetically modified *Escherichia coli* bacteria. Perhaps the greatest success with this technology has been the biosynthesis of human insulin, a lack of which causes diabetes in humans. Two genes are required to make this relatively short polypeptide which consists of only 51 amino acids. Other medically useful substances produced by genetically engineered organisms include human growth hormone, tissue plasmogen activator that dissolves blood clots formed in heart attacks and strokes, and various vaccine proteins to inoculate against diseases such as meningitis, hepatitis B, and influenza. Genetic engineering is discussed in more detail in Section 12.8.
