rearrangements we discussed in Chapter 8 and for detecting chromosomal abnormalities in cancer cells as was discussed in Chapter 19.

20.5 DNA Sequencing Is the Ultimate Way to Characterize DNA Structure at the Molecular Level

In a sense, a cloned DNA molecule or any DNA, from a single gene to an entire genome, is completely characterized at the molecular level only when its nucleotide sequence is known. The ability to sequence DNA has greatly enhanced our understanding of genome organization and increased our knowledge of gene structure, function, and mechanisms of regulation.

Historically the most commonly used method of DNA sequencing was developed by Fred Sanger and his colleagues and is known as **dideoxynucleotide chain-termination sequencing** or simply **Sanger sequencing**. In this technique, a double-stranded DNA molecule whose sequence is to be determined is converted to single strands that are used as a template for synthesizing a series of complementary strands. The DNA to be sequenced is mixed with a primer that is complementary to the target DNA or vector along with DNA polymerase, and the four deoxyribonucleotide triphosphates (dATP, dCTP, dGTP, and dTTP) are added to each tube.

The key to the Sanger technique is the addition of a small amount of one modified deoxyribonucleotide (Figure 20–15), called a **dideoxynucleotide** (abbreviated ddNTP). Notice that dideoxynucleotides have a 3′ hydrogen instead of a 3′ hydroxyl group. Dideoxynucleotides are called chain-termination nucleotides because they lack the 3′ oxygen required to form a phosphodiester bond with another nucleotide. Thus when ddNTPs are included in a reaction as DNA synthesis takes place, the polymerase occasionally inserts a dideoxynucleotide instead of a deoxyribonucleotide into a growing DNA strand. Since the dideoxynucleotide has no 3′-OH group, it cannot form a 3′ bond with another nucleotide, and DNA synthesis terminates because DNA polymerase cannot add new nucleotides to a ddNTP. The Sanger reaction takes advantage of this key modification.

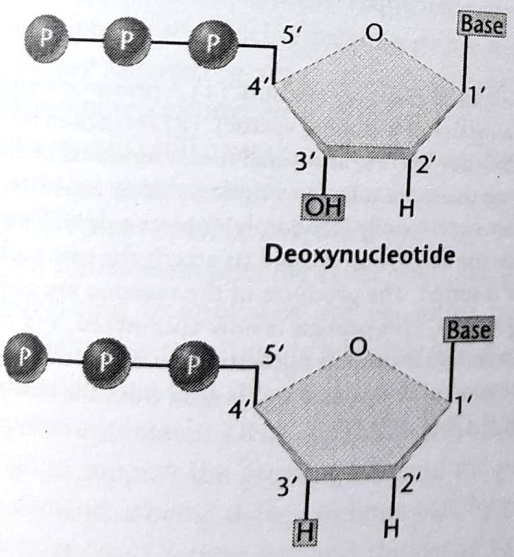

FIGURE 20–15 Deoxynucleotides (top) have an OH group at the 3′ position in the deoxyribose molecule. Dideoxynucleotides (bottom) lack an OH group and have only hydrogen (H) at this position. Dideoxynucleotides can be incorporated into a growing DNA strand, but the lack of a 3′-OH group prevents formation of a phosphodiester bond with another nucleotide, terminating further elongation of the template strand.

For example, in Figure 20–16, notice that the shortest fragment generated is a sequence that has added ddCTP to the 3′ end of the primer and the chain has terminated. Over time as the reaction proceeds eventually there will be a ddNTP inserted at every location in the newly synthesized DNA so that each strand synthesized differs in length by one nucleotide and is terminated by a ddNTP. This allows for separation of these DNA fragments by gel electrophoresis, which can then be used to determine the sequence.

When the Sanger technique was first developed, four separate reaction tubes, each with a different single ddNTP (e.g., ddATP, ddCTP, ddGTP, and ddTTP), were used. These reactions typically used either a radioactively labeled primer or radioactively labeled ddNTP for analysis of the sequence following polyacrylamide gel electrophoresis and autoradiography. This approach involved large polyacrylamide gels in which each reaction was loaded on a separate lane of the gel and ladder-like banding patterns revealed by autoradiography were read to determine the sequence. This original approach could typically read about 800 bases of 100 DNA molecules simultaneously. *Read length*, that is, the amount of sequence that can be generated in a single individual reaction and the total amount of DNA sequence generated in a sequence *run*, which is effectively read length times the number of reactions an instrument can run during a given period of time, have become hot areas for innovation in sequencing technology.

In the past 20 years, modifications of the Sanger technique have led to technologies that now allow sequencing reactions to occur in a single tube in which each of the four ddNTPs is labeled with a different-colored fluorescent dye (Figure 20–16). These reactions were often carried out in PCR-like fashion using cycling reactions that permit greater read and run capabilities. The reaction products are separated through a single, ultrathin-diameter polyacrylamide tube gel called a capillary gel (capillary gel electrophoresis). As DNA fragments move through the gel, they are scanned with a laser. The laser stimulates fluorescent dyes on each DNA fragment, which then emit different wavelengths of

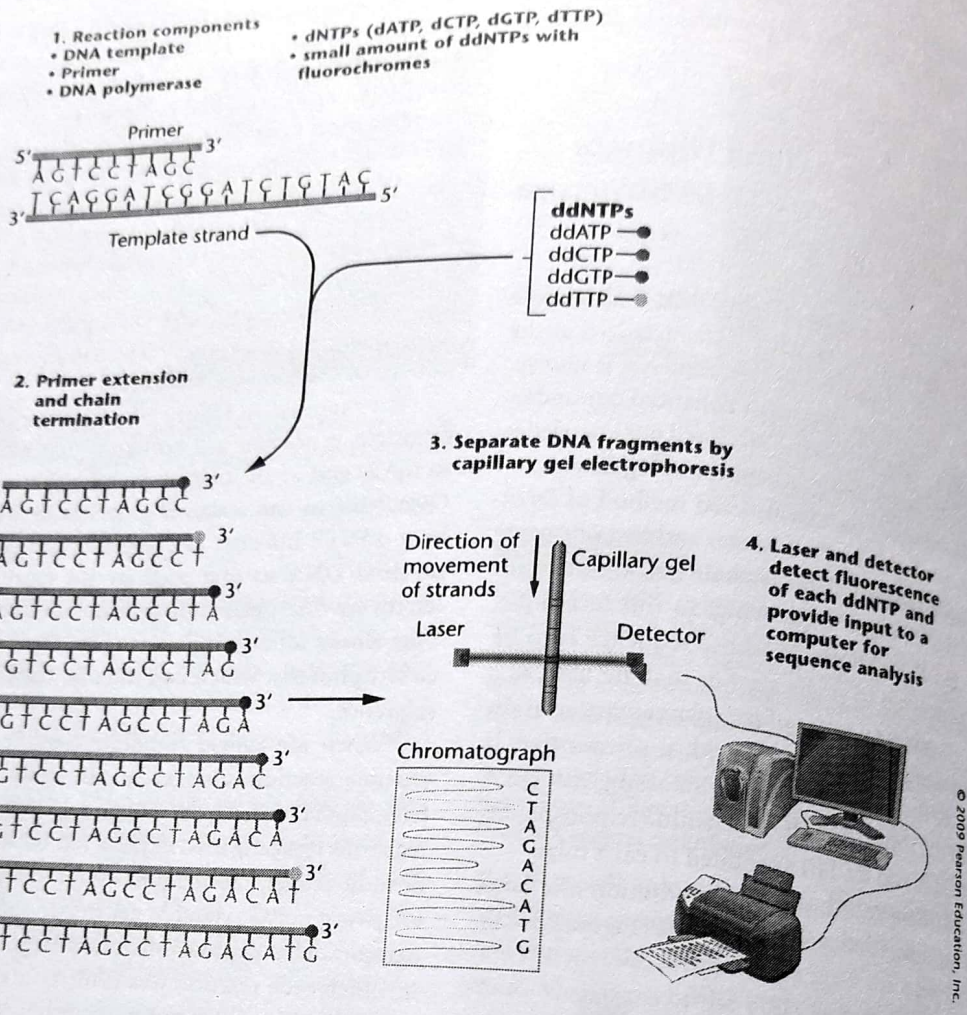

FIGURE 20–16 Computer-automated DNA sequencing using the chain-termination (Sanger) method. (1) A primer is annealed to a sequence adjacent to the DNA being sequenced (usually near the multiple cloning site of a cloning vector). (2) A reaction mixture is added to the primer–template combination. This includes DNA polymerase, the four dNTPs, and small molar amounts of dideoxynucleotides (ddNTPs) labeled with fluorescent dyes. All four ddNTPs are added to the same tube, and during primer extension, all possible lengths of chains are produced. During primer extension, the polymerase occasionally (randomly) inserts a ddNTP instead of a dNTP, terminating the synthesis of the chain because the ddNTP does not have the OH group needed to attach the next nucleotide. Over the course of the reaction, all possible termination sites will have a ddNTP inserted. The products of the reaction are added to a single lane on a capillary gel, and the bands are read by a detector and imaging system. This process is now automated, and robotic machines, such as those used in the Human Genome Project, sequence several hundred thousand nucleotides in a 24-hour period and then store and analyze the data automatically. The sequence is obtained by extension of the primer and is read from the newly synthesized strand, not the template strand. Thus, the sequence obtained begins with 5′-CTAGACATG-3′.

light for each ddNTP. Emitted light is captured by a detector that amplifies and feeds this information into a computer to convert the light patterns into a DNA sequence that is technically called an electropherogram (Figure 20–17). The data are represented as a series of colored peaks, each corresponding to one nucleotide in the sequence.

Since the early 1990s DNA sequencing has largely been performed through computer-automated Sanger-reaction-based technology and is referred to as **computer-automated high-throughput DNA sequencing**. Such systems generate relatively large amounts of sequence DNA. Computer-automated sequences can achieve read lengths of approximately 1000 bp with about 99.999 percent accuracy for about $0.50 per kb. Automated DNA sequencers often contain multiple capillary gels (as many as 96) that are several feet long and can process several thousand bases of sequences so that many of these instruments made it possible to generate over 2 million bp of sequences in a day! These systems became essential for