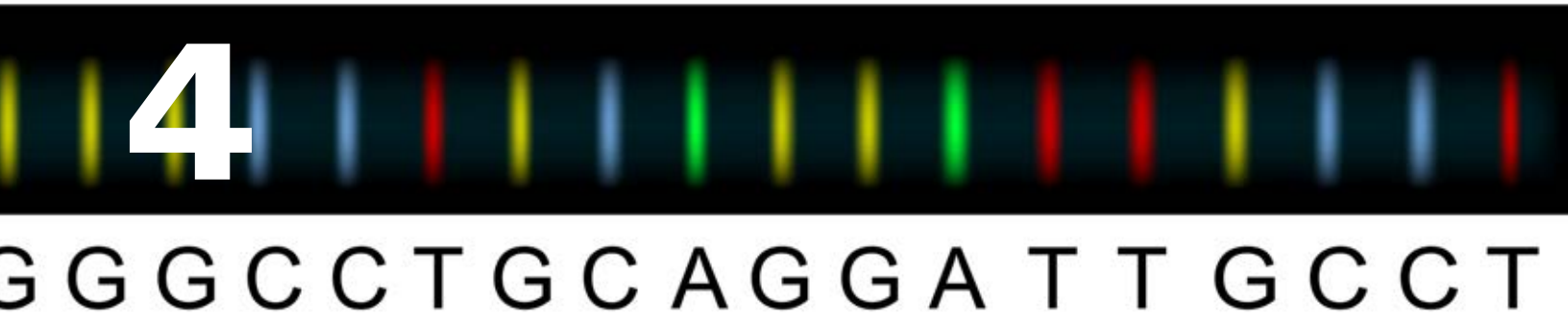


Molecular Biology Technology

CHAPTER OUTLINE

Molecular biologists use physical methods and enzyme-catalyzed reactions to characterize and modify nucleic acids. These techniques have become so widely used that you may have encountered some of them in other life science courses. The first part of this chapter describes methods used to isolate and characterize nucleic acids. These methods also apply to proteins and nucleoproteins. The second part describes techniques that use enzymes to manipulate and sequence DNA in the laboratory. Of course, these enzymes have important biological functions in the cell. But for now our attention is directed toward using the enzymes as laboratory tools to cut, sequence, replicate, or otherwise manipulate DNA in the test tube. The biological functions of the enzymes are considered elsewhere.

4.1 Nucleic Acid Isolation

The method used to isolate DNA depends on the DNA source.

DNA isolation, an essential step in many molecular biology experiments, begins with the release of the DNA, often with other biological molecules such as proteins and RNA, from the cells or viruses of interest. Because the biological sources' structures and compositions vary, the particular method used to release the DNA must be tailored to the cellular or viral DNA source. The method of choice depends on how the DNA is enclosed and the percent of total dry weight that is DNA (which varies from about 1% in complex mammalian cells to about 50% in bacterial viruses). Care must be taken to avoid vigorous stirring or other procedures that produce hydrodynamic shear forces that may break DNA. The following procedures are used to isolate viral, bacteria, plasmid, and yeast DNA:

1. *Viral DNA*. An aqueous suspension of virus particles (virions) is gently mixed with phenol, a reagent that is slightly miscible with water. A small amount of phenol enters the aqueous layer, breaking open the protein coat and denaturing individual protein molecules. (Sometimes chloroform and isoamyl alcohol are added to assist the phenol.) Most of the denatured protein either enters the phenol layer or precipitates at the phenol–water interface. The upper aqueous layer, containing the DNA, is carefully separated from the phenol layer and ethanol is added to the aqueous fraction to precipitate DNA. After collection by centrifugation, DNA is dissolved in an aqueous buffer solution.
2. *Bacterial DNA*. The DNA in a bacterial cell is enclosed in a cell envelope consisting of a cell wall and membrane(s). Phenol cannot, by itself, remove this envelope. The envelope can be removed, however, by first treating a bacterial cell suspension with lysozyme to digest the cell wall and then with a detergent such as sodium dodecyl sulfate [SDS; $CH_3(CH_2)_{11}OSO_3^-Na^+$] to disrupt the cell membrane. After

the cell envelope is removed phenol is added with gentle mixing, and the mixture is allowed to stand so the aqueous and phenol phases separate. The upper aqueous phase is collected and treated with ethanol to precipitate DNA and RNA. The precipitate containing DNA and some contaminating RNA is spooled out with a glass rod and excess alcohol removed. Then the spooled DNA is dissolved in a buffered solution containing RNase to digest contaminating RNA. Chloroform is added to the mixture with gentle mixing, and the mixture is allowed to stand so the phases can separate. The upper aqueous phase is collected and treated with ethanol to precipitate DNA, which is spooled out with a glass rod and dissolved in an aqueous buffer solution.

3. *Plasmid DNA*. Plasmid DNA is released from bacteria by first incubating the bacterial suspension with lysozyme to digest the cell wall. Then an SDS–sodium hydroxide solution is added to the mixture. SDS disrupts the cell membrane, and sodium hydroxide hydrolyzes RNA to form a mixture of 2′- and 3′-nucleoside monophosphates. DNA, which lacks 2′-hydroxyl groups, is denatured but not hydrolyzed. The two strands in a linear duplex separate completely after denaturation. Strands derived from the covalently closed, circular double-stranded plasmid DNA, however, remain intertwined after denaturation. When acid is added to neutralize the extract, intertwined plasmid DNA strands rewind to form the intact plasmid. In contrast, the separated bacterial DNA strands combine to form an insoluble aggregate that is removed, along with denatured protein, by centrifugation. The soluble plasmid DNA is precipitated by adding ethanol, collected by centrifugation, and dissolved in an aqueous buffer solution.
4. *Yeast and fungal DNA*. The polysaccharides that make up yeast and fungal cell walls are resistant to lysozyme. Other enzymes (such as cellulase isolated from snails), however, break down these cell walls. Once cell walls have been disrupted, the DNA isolation procedure is similar to that used for bacteria.

Great care must be taken to protect RNA from degradation during its isolation.

One of the greatest concerns when isolating RNA is that the RNA will be degraded by RNases released during the isolation procedure. This problem can be solved by freezing the cells in liquid nitrogen, transferring the frozen cells to a mortar, and grinding the frozen cells with a pestle. Then the ground cells are suspended in an acidic solution containing guanidinium thiocyanate, sodium acetate, phenol, and chloroform. Under these acidic conditions RNA is present in the upper aqueous phase, whereas most of the DNA and proteins are present in either the lower organic phase or the interphase. The aqueous phase containing the RNA is transferred to a clean tube, and ethanol or isopropanol is added. The RNA, which precipitates out of solution,

is collected by centrifugation, washed with alcohol, and dissolved in an aqueous buffer solution.

4.2 Physical Techniques Used to Study Macromolecules

Different physical techniques are used to study macromolecules.

Several physical techniques are used to study nucleic acids, proteins, and nucleoproteins. Four of these—electron microscopy, sucrose gradient sedimentation, equilibrium density gradient centrifugation, and electrophoresis—are described briefly so experiments can be understood.

Electron microscopy allows us to see nucleic acids and nucleoprotein complexes.

Three electron microscopy techniques are widely used to study nucleic acids and nucleoprotein complexes:

1. *Metal shadowing:* Metal shadowing provides a three-dimensional image of the surface of a biological sample (FIGURE 4.1). The biological sample is dried on a translucent

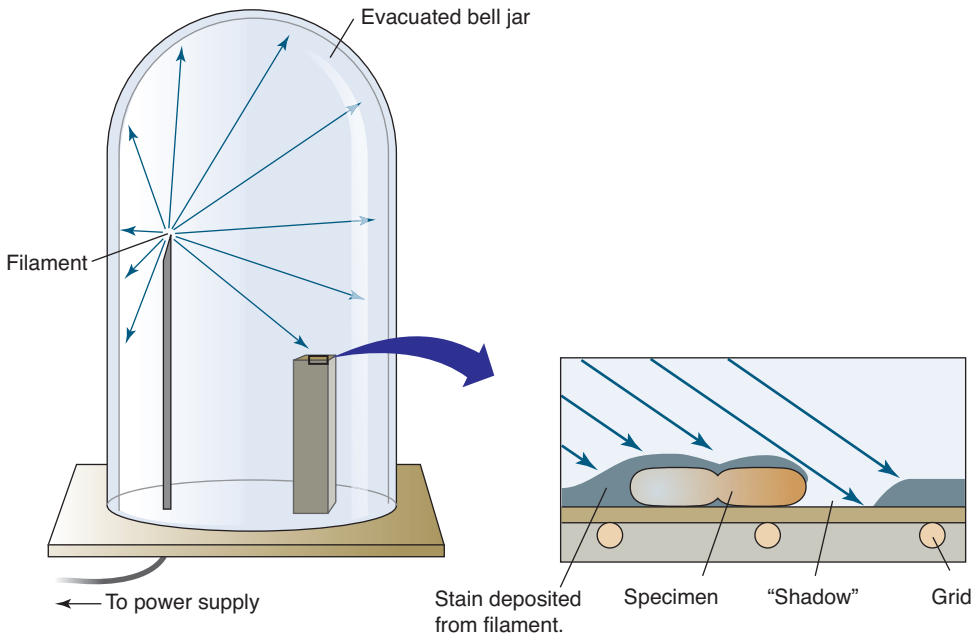


FIGURE 4.1 Metal shadowing. This technique allows the observer to view details on the surface of small particles. The sample is dried on a thin film, which is then placed in a vacuum chamber. A filament of a heavy metal such as tungsten is heated and the metal evaporates, forming a thin metal coat on the sample. The observer uses an electron microscope to view the metal coating, which is a replica of the biological sample. (Adapted from Ingraham, J., and Ingraham, C. 2006. *Introduction to Microbiology* (3rd ed.). Academic Internet Publishers.)

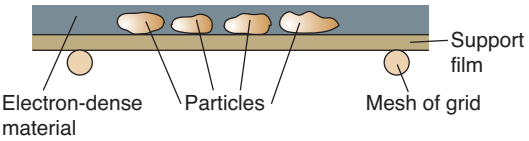
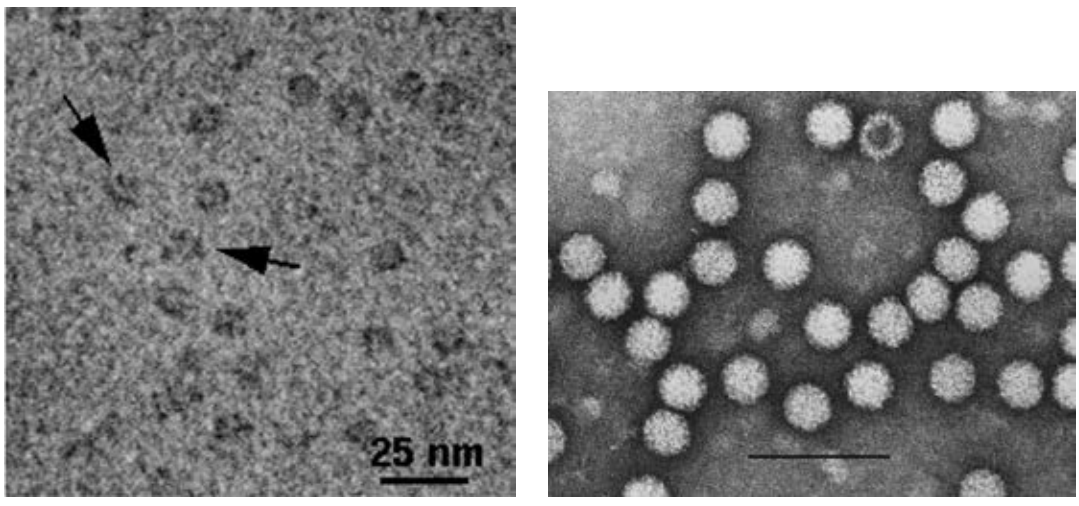


FIGURE 4.2 Negative staining method of visualizing particles by electron microscopy. Four virus particles (and macromolecules) are embedded in a substance that absorbs electrons strongly. As the beam passes through the sample, the fraction of the electrons in the beam that is absorbed depends on the total thickness of the substance; therefore, more electrons will pass through the regions containing each particle and the particle will appear bright against a dark background.

film and then placed in a vacuum chamber, where a filament of a heavy metal such as tungsten is heated. The evaporated metal forms a thin metal coat on the surface of the biological sample. The observer uses an electron microscope to view the metal coating, which is a replica of the biological sample. Metal shadowing is a very useful technique for observing very long molecules when one is primarily interested in their length and linear topology.

2. *Negative staining:* Negative staining provides an alternative means for visualizing macromolecules and virus particles. The macromolecules or virus particles of interest are added to a solution containing a heavy metal salt such as uranyl acetate. The mixture is then deposited as microdroplets onto a supporting grid and allowed to dry. The heavy metal salts form a mold surrounding and covering the macromolecules or virus particles **(FIGURE 4.2)**. The mold surrounding the macromolecules or virus particles is thicker than that covering them. Therefore, when the grid is placed in an electron beam, more electrons pass through the sample than the surrounding region, producing a negative image of the macromolecules or virus particles. **FIGURE 4.3** shows proteins and virus particles visualized by using the negative staining technique.
3. *Cryoelectron microscopy:* Cryoelectron microscopy is used to create three-dimensional constructs of protein and nucleoprotein complexes. This technique involves rapidly freezing a droplet of a buffer solution containing the specific protein or nucleoprotein complexes so that a thin amorphous layer of ice forms that holds the complexes in random orientations. A picture is formed in the electron microscope by using a



(a) (b)

FIGURE 4.3 Electron micrographs obtained by the negative contrast procedure. (a) Image of negatively stained bacteriophage T7 helicase/primase in the presence of dTDP. (b) Images of negatively stained tomato bushy stunt virus particles. Note the surface details, which show the individual protein molecules that form the protein coat of which each particle is composed. (Part a reproduced with kind permission from Springer Science+Business Media: Ohi, M., et al., *Biol. Proc. Online* 6 [2004]: 23–34. Photo courtesy of Thomas Walz, Harvard Medical School. Part b photo courtesy of Robert G. Milne, Plant Virus Institute, National Research Council, Turin, Italy, with permission from Damion Milne.)

radiation dose that is so low the complexes are not damaged. The resulting micrograph shows hundreds of the protein or nucleoprotein complexes lying in different orientations. Computer-assisted image processing converts the projections into three-dimensional constructs such as those shown for the bacterial ribosome and its subunits in **FIGURE 4.4**.

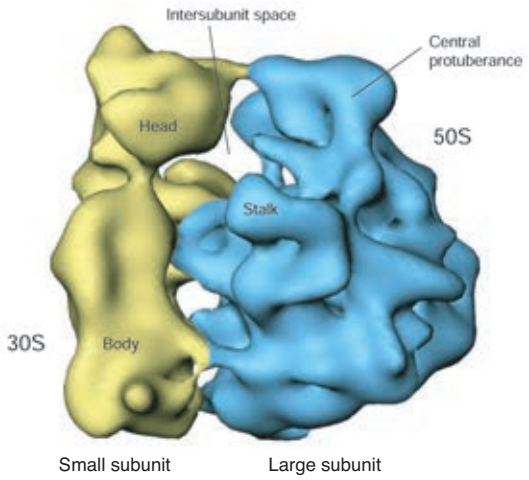


FIGURE 4.4 Cryoelectron microscopy reconstruction of the *Escherichia coli* ribosome. (Reproduced from Frank, J., and Agrawal, R. K. 2001. Bacterial ribosomes. *Encyclopedia of Life Sciences*. Reproduced with permission from John Wiley & Sons, Ltd: Chichester. Photo courtesy of Rajendra K. Agrawal, Wadsworth Center.)

Centrifugation can separate macromolecules and provide information about their size and shape.

Several important properties of macromolecules can be determined from the rates at which they move in a centrifugal field. These studies cannot be performed with an ordinary laboratory centrifuge because it cannot produce a sufficient centrifugal force to sediment molecules with enough velocity to overcome randomized diffusion. Modern ultracentrifuges, however, can generate forces as great as 800,000 × gravity, which is more than enough to cause macromolecules to sediment through a solution.

A particle's sedimentation velocity depends on its mass and shape. If two particles have the same shape, the one with the larger mass sediments more rapidly than one with the smaller mass. If two particles both have the same molecular mass, the one with the more compact shape sediments more rapidly under the influence of a centrifugal force. Shape influences sedimentation rate because the more compact a particle is, the less frictional drag it experiences as it sediments through a solution.

The ratio of sedimentation velocity to centrifugal force is called the sedimentation coefficient (s). That is,

$$s = \text{velocity/centrifugal force}$$

The value of s for a particular molecule is often the same in many different solutions, so an s value is frequently considered to be a constant that characterizes a macromolecule. Furthermore, because the s value depends on molecular mass and shape, changes in the s value, as experimental conditions are varied, can be used to monitor changes in molecular aggregation or conformation. For most macromolecules the s value is between 1×10^{-13} and 100×10^{-13} sec. In honor of Theodor Svedberg, the ultracentrifuge's inventor, 10^{-13} seconds is called one Svedberg, or 1S. The s value of a particle formed by the association of two smaller particles cannot be determined by simply adding the s values of the two smaller particles. For example, a 30S bacterial ribosomal subunit combines with a 50S bacterial ribosomal subunit to form a 70S bacterial ribosome (and *not* an 80S bacterial ribosome).

Two different centrifugation techniques are often used in molecular biology laboratories. The first technique is *sucrose gradient centrifugation*. A centrifuge tube is filled with a sucrose solution so the sucrose concentration increases continuously from the top to the bottom of the centrifuge tube (**FIGURE 4.5**). The protein, nucleic acid, or nucleoprotein solution of interest is carefully layered on top of the sucrose gradient to form a band (or zone). Then the plastic centrifuge tube is placed in a swinging bucket rotor, which is a type of centrifuge head that contains buckets that can swivel on pins. The rotor is placed in the centrifuge. As

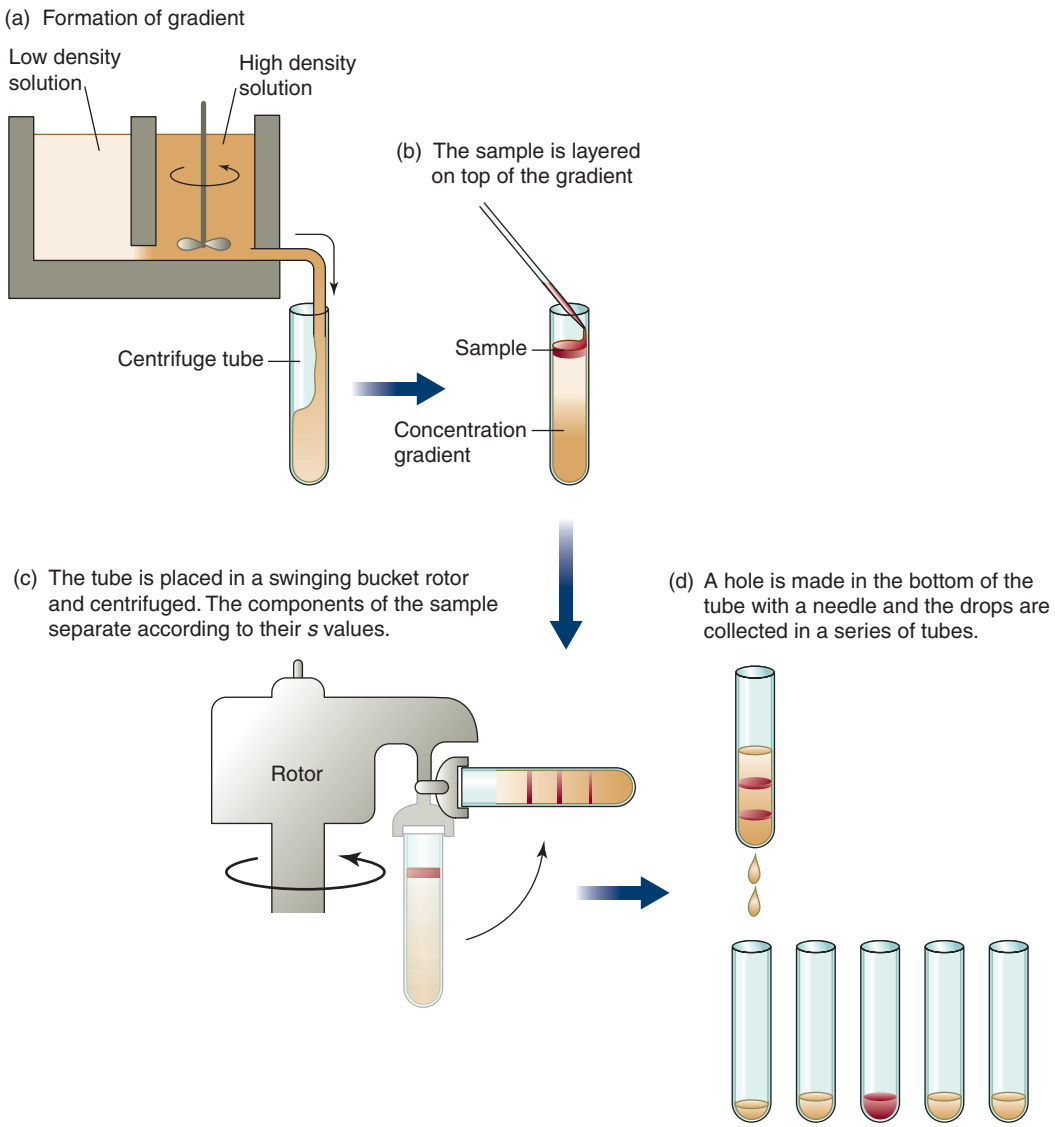


FIGURE 4.5 **Sucrose gradient centrifugation.**

the rotor turns the buckets swivel, causing the centrifuge tube to assume a horizontal position. After centrifugation is complete, the swinging bucket returns to a vertical position and the centrifuge tube inside it is removed. A tiny hole is then punched in the bottom of the centrifuge tube, and drops of the solution are collected in separate tubes. These drops, representing successive layers of solution in the tube, are analyzed to determine the macromolecule's concentrations along the tube.

The second centrifugation technique is *equilibrium density gradient centrifugation*. In this technique, nucleic acids or nucleoproteins are suspended in a CsCl (or Cs_2SO_4) solution at a concentration that is chosen so the solution density is approximately equal to that of the macromolecule. Under the influence of a powerful centrifugal force, Cs^+ and Cl^- (or SO_4^{2-}) ions move toward the bottom of the centrifuge to some extent. They do not accumulate on the bottom of the centrifuge tube, however, because the centrifugal force is not great enough to counteract the tendency for diffusion to maintain a uniform distribution of the ions. After a period

of several hours the ions achieve an equilibrium concentration distribution in which there is a nearly linear concentration gradient and, hence, a nearly linear density gradient in the centrifuge tube.

The density is maximal at the bottom of the tube. As the density gradient forms macromolecules begin to migrate. Those in the upper reaches of the tube move toward the bottom, stopping at the position at which their density equals the solution density. Similarly, macromolecules in the lower part of the tube move upward, stopping at the same position. In this way the macromolecules form a narrow band in the tube. If the solution contains macromolecules having different densities, each macromolecule forms a band at the position in the gradient that matches its own density, and thus the macromolecules can be separated.

This technique's resolution is extraordinary. For example, it can separate single-stranded DNA molecules that differ only in the nitrogen isotopes present in their bases. As shown in **FIGURE 4.6**, single-stranded DNA molecules with the naturally occurring ^{14}N isotope are separated from single-stranded DNA molecules with the heavier ^{15}N isotope. The densities of the ^{14}N and ^{15}N single-stranded DNAs are 1.710 g • cm^{-3} and 1.724 g • cm^{-3}, respectively.

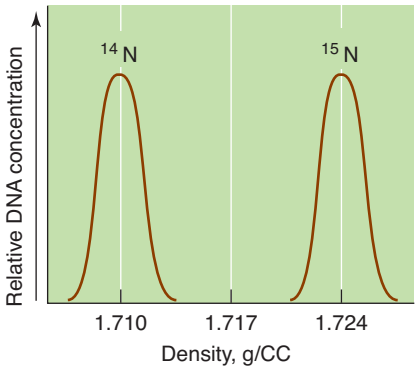


FIGURE 4.6 Demonstration of DNA strand separation by equilibrium density centrifugation in CsCl. ^{14}N indicates single-stranded DNA prepared from bacteria cultured in a medium that contained $[^{14}N]NH_4Cl$. ^{15}N indicates single-stranded DNA prepared from bacteria cultured in a medium that contained $[^{15}N]NH_4Cl$.

Gel electrophoresis separates nucleic acids based on their rate of migration in an electric field.

Gel electrophoresis can be used to separate DNA or RNA molecules according to size (**FIGURE 4.7**). A gel slab, which has small wells at the

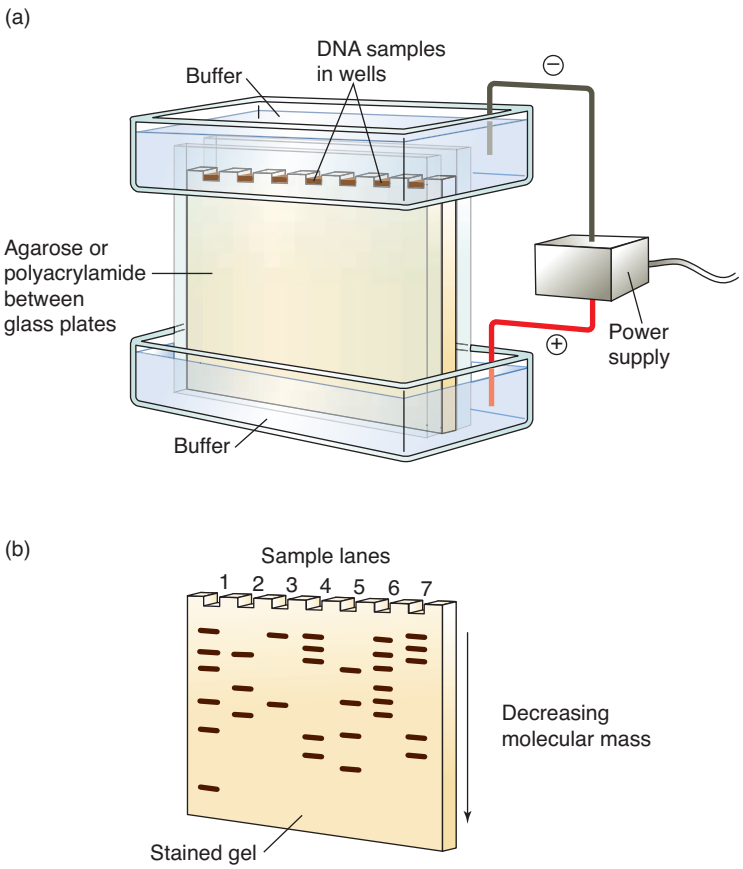


FIGURE 4.7 Gel electrophoresis. (a) Apparatus for gel electrophoresis capable of handling several samples simultaneously. An agarose or polyacrylamide suspension is placed in a mold fitted with glass or plastic plates on each side, and an appropriately shaped mold is placed on top of the gel during hardening to make "wells" for the samples. After the gel has hardened the mold on top is removed and samples are placed in the wells. The power supply is connected to the electrophoresis apparatus and the nucleic acids migrate toward the positive electrode (anode). When electrophoresis is complete the power is turned off and the nucleic acids in the gel are made visible by removing the glass or plastic plates and immersing the gel in a solution containing a reagent that binds to or reacts with the separated molecules. (b) The separated components of a sample appear as bands, which may either be colored or fluoresce. The region of a gel in which the components of one sample can move is called a lane. This gel thus has seven lanes.

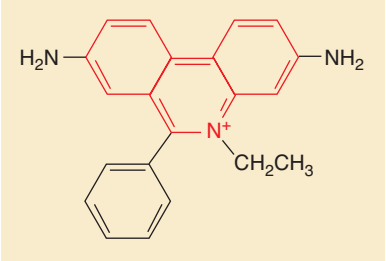


FIGURE 4.8 Ethidium bromide. Ethidium bromide contains a planar phenanthridium ring (shown in red) that inserts between the stacked bases in DNA. DNA absorbs ultraviolet light radiation at 260 nm and transmits it to the bound dye, which re-emits the light at 590 nm in the red-orange region of the spectrum.

top, is prepared on a glass or plastic support. Then the gel and support are immersed in a buffer solution, samples are loaded into the wells, and an electric field is applied across the gel. Negatively charged nucleic acids penetrate and move through the complex network of molecules that comprise the gel and toward the positive electrode or anode. A nucleic acid's migration rate depends on its total charge and shape (that is, its frictional resistance). Its migration rate is also influenced by its molecular mass because surface area, which affects frictional drag, increases with molecular mass. Smaller nucleic acids squeeze through the narrow, tortuous passages in the gel network more easily than larger molecules. Migration rate therefore increases as the molecular mass decreases.

DNA bands can be viewed by staining the gel with ethidium bromide (**FIGURE 4.8**), which forms a DNA • ethidium bromide complex that fluoresces under ultraviolet light. The sensitivity of detection is sufficiently great that samples with as little as 0.1 µg DNA are easily seen. Because ethidium bromide is carcinogenic, many workers prefer to work with commercially available safer substitutes. Conditions can be adjusted so the rate of migration depends on length in nucleotides for single-stranded nucleic acids or base pairs for double-stranded nucleic acids. Gel electrophoresis can also be used to separate topoisomers (**BOX 4.1 IN THE LAB: GEL ELECTROPHORESIS AND TOPOISOMER SEPARATION**).

Polyacrylamide gels can separate single-stranded DNA molecules that differ in size by just one nucleotide provided the strands are between 5 and 750 nucleotides long. Agarose gels have lower resolving power but separate DNA molecules that range in size from 200 base pair (bp) to about 50,000 bp. Empirical studies show the distance moved by DNA (D) during slab gel electrophoresis depends logarithmically on its length in base pairs, obeying the equation,

$$D = a - b \log bp$$

in which a and b are empirically determined constants that depend on the buffer, the gel concentration, and the temperature. This logarithmic relationship allows us to determine a DNA molecule's length in base pairs by comparing the distance it moves with distances moved by DNA standards of known length in base pairs under the same conditions (**FIGURE 4.9**).

Although quite effective in separating DNA molecules by size, slab gel electrophoresis is (1) slow (usually requiring hours to complete a run), (2) difficult to automate, and (3) does not easily provide quantitative data. An alternate means for separating DNA in an applied electric field, called **capillary gel electrophoresis**, does not suffer from these disadvantages (**BOX 4.2 IN THE LAB: CAPILLARY GEL ELECTROPHORESIS**).

SDS-PAGE can be used to determine a polypeptide's molecular mass.

The molecular mass of native proteins cannot be determined by gel electrophoresis because proteins do not have a uniform charge across the entire molecule and proteins come in a variety of shapes

BOX 4.1: IN THE LAB

Gel Electrophoresis and Topoisomer Separation

Gel electrophoresis can separate topoisomers (FIGURE B4.1). Because superhelical molecules are compact, they move through the gel more rapidly than relaxed circles with the same mass. Gel electrophoresis therefore serves as a convenient method for detecting topoisomerase activity. The advantages of this method are that it requires very little DNA, is very sensitive, can be used to analyze several samples at one time, requires relatively inexpensive equipment, and is simple to perform.

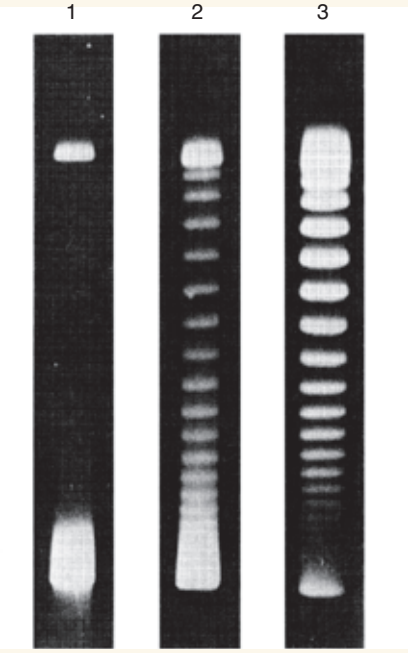


FIGURE B4.1 Agarose gel electrophoresis pattern of covalently closed circular SV40 (simian virus 40) DNA. The DNA was applied to the top of the gel. Lane 1 contains untreated negatively supercoiled native DNA (lower band). Lanes 2 and 3 contain the same DNA that was treated with a type I topoisomerase, which makes a single strand break in only one chain that relaxes negative supercoils by causing successive one unit increases in the linkage number ($\Delta Lk = +1$). The DNA in lanes 2 and 3 were treated for 5 and 30 minutes respectively. (Reproduced from Keller, W. 1975. *Proc Natl Acad Sci USA* 72:2550–2554. Photo courtesy of Walter Keller, University of Basel.)

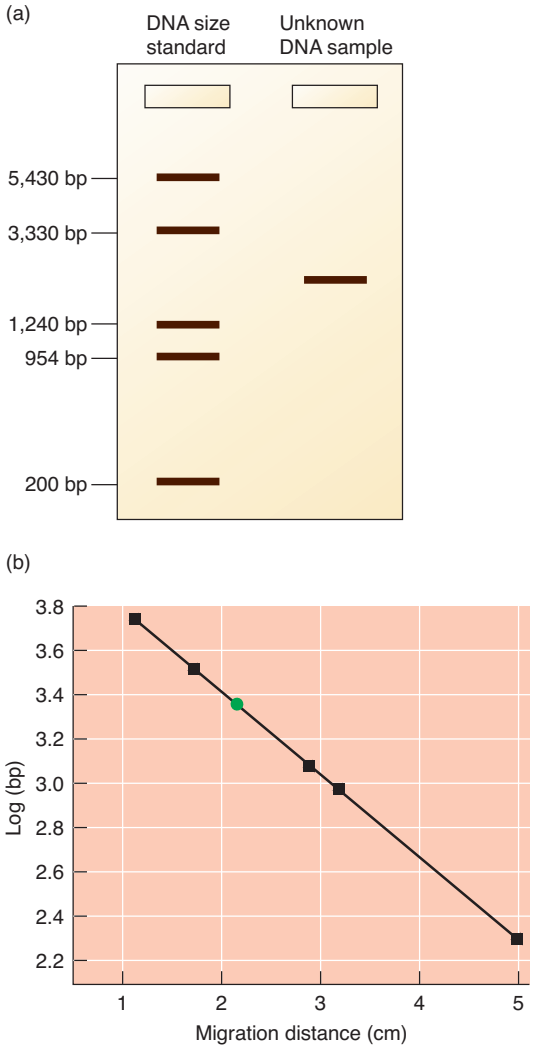


FIGURE 4.9 DNA size determination by gel electrophoresis. (a) A DNA size standard and a DNA of unknown size were placed in the left and right lanes, respectively. (b) A graph showing the migration distance in cm vs. log bp. The points shown in black squares are for the standard DNA and the point shown in a green circle is for the DNA of unknown size. (Modified from illustrations by Michael Blaber, Department of Biomedical Sciences, Florida State University.)

so there is no simple way to predict the migration rate. Proteins could be separated solely on the basis of size if they had a uniform charge and the shape factor could be eliminated. A technique known as **SDS-PAGE (sodium dodecyl sulfate-polyacrylamide gel electrophoresis)** does just that. The detergent SDS and the disulfide bond-breaking agent β-mercaptoethanol ($HSCH_2CH_2OH$) are added to the protein, causing all the polypeptide chains to be denatured, forming rod-shaped structures that are coated with negatively

BOX 4.2: IN THE LAB

Capillary Electrophoresis

Separations in capillary electrophoresis are performed in very thin (20–100 μm) capillary tubes made of quartz, glass, or plastic filled with replaceable high-molecular-mass linear polymers. Because the capillary tubes are so thin, heat is rapidly dissipated. This heat dissipation allows separations in electrical fields (typically 100–500 $V \cdot cm^{-1}$), which are about 10 times higher than those used in slab gel electrophoresis. Capillary electrophoresis thus allows high-resolution separations comparable with those obtained by slab gels in minutes rather than hours. Furthermore, capillary electrophoresis can be automated and provide quantitative data. Commercially available capillary electrophoresis systems can run as many as 96 capillaries at the same time, allowing high-throughput DNA sequence analysis.

charged SDS. The net effect is that, as in the case of DNA, all the proteins migrate toward the anode, the migration rate increases as chain length decreases, and the dependence is logarithmic. SDS-PAGE can therefore be used to determine a polypeptide's molecular mass by comparing its migration rate with the migration rates of polypeptide standards.

4.3 Enzymatic Techniques Used to Manipulate DNA

Nucleases are useful tools in DNA investigations.

Nucleases, enzymes that digest polynucleotides by cleaving the 5′→3′ phosphodiester bonds that link neighboring nucleosides, are valuable tools for studying nucleic acids. Those that digest DNA are designated deoxyribonucleases (DNases), whereas those that digest RNA are designated ribonucleases (RNases). Some nucleases digest both kinds of nucleic acids. Some DNases are specific for double-stranded (ds) DNA, others for single-stranded (ss) DNA, and still others act on both kinds of DNA.

Nucleases also differ in the specificity with which they cut the nucleic acid. Nucleases that act within a strand are called **endonucleases**. Some endonucleases are specific and cleave only between particular bases. Nucleases that act only at the end of a nucleic acid, removing a single nucleotide at a time, are called **exonucleases**. Exonucleases that begin cutting at the 5′- or 3′-ends are designated 5′→3′ or 3′→5′ exonucleases, respectively. Although nucleases have many important biological functions, the present focus is on using them as laboratory tools. Some specific nucleases are described in TABLE 4.1 and others are described when relevant in the text.

TABLE 4.1 Properties of Selected Nucleases

Nuclease	Substrate	Site of Cleavage	Product
Pancreatic RNase	ssRNA in high salt; all RNA in low salt	endonuclease; after C or U	mono- or oligonucleotides with a 3′-P
T1 RNase	ssRNA in high salt; all RNA in low salt	endonuclease; after G	mono- or oligonucleotides with a 3′-P
Pancreatic DNase I	ss or dsDNA	endonuclease	oligonucleotides
Phosphodiesterase venom	RNA or DNA	exonuclease from the 3′ end	5′-P mononucleotides
Phosphodiesterase spleen	RNA or DNA	exonuclease from the 5′ end	3′-P mononucleotides
Micrococcal nuclease	DNA (or RNA) ss or ds (prefers ssDNA to dsDNA; ssRNA to ds RNA)	endonuclease and exonuclease; at AT or AU rich regions, requires Ca^{2+}	mononucleotides and oligonucleotides with 3′-P ends
S1 nuclease	ssDNA (or ssRNA. S1 nuclease is 5× more active on ssDNA than on ssRNA)	endonuclease	5′-P mononucleotides
Exonuclease I *E. coli*	ssDNA	3′→5′ exonuclease	5′-P mononucleotides plus a terminal dinucleotide
Exonuclease III *E. coli*	dsDNA	3′→5′ exonuclease	5′-P mononucleotides
Exonuclease VII *E. coli*	ssDNA	3′→5′ exonuclease and 5′→3′ exonuclease	5′-P mononucleotides

ss = single-stranded; ds = double-stranded.

A specific example, the **S1 endonuclease** isolated from *Aspergillus oryzae*, illustrates how we use nucleases to investigate nucleic acid structure. The S1 endonuclease acts exclusively on single-stranded polynucleotides or on single-stranded regions of double-stranded nucleic acids. It differs from other single-strand–specific enzymes in that the single-stranded region can be as small as one or two bases. Negatively supercoiled DNA contains single-stranded bubbles resulting from localized transient melting. Supercoiled DNA can be cleaved by S1 nuclease because of these regions. In fact, this nuclease can be used to distinguish supercoiled from both nonsupercoiled covalent circles and nicked circular DNA, both of which are resistant to the enzyme. S1 nuclease makes a double-strand break because it acts on both single-stranded branches of the bubble.

Restriction endonucleases cleave within specific nucleotide sequences in DNA.

Molecular biologists recognized they could use nucleases that cut DNA at specific sites to study nucleic acids in the same way protein chemists use trypsin and other specific proteases to cut polypeptide chains. The problem of finding nucleases with sufficient sequence specificity to be

useful was solved when investigators discovered a class of nucleases that cleave bacterial virus DNA molecules that enter a bacterial cell, impeding viral replication. These nucleases, called **restriction endonucleases** because they block or restrict viral replication, act only on DNA with specific recognition sequences and only when the recognition sequences are not modified. Host DNA is protected because it has methyl groups attached to specific bases within the recognition sequence.

Four major types of restriction endonucleases are known. We examine just one of these, the type II restriction endonucleases, which are ideally suited for manipulating and studying DNA because they recognize and cleave specific sequences that are 4 to 8 bp long.

Hamilton O. Smith was the first to isolate and characterize a type II restriction endonuclease. The enzyme, isolated from *Haemophilus influenzae* and called HindII, recognizes the set of sequences,

5′····GTPy↓PuAC-····3′
3′····CAPu↑PyTG-····5′

where Py and Pu represent pyrimidine and purine, respectively. The arrows indicate sites at which each strand is cut.

The discovery of HindII motivated investigators to seek other endonucleases that cut within specific nucleotide sequences. The search has been rewarded. More than 2,500 different type II enzymes have been identified, and several of these have been sequenced and biochemically characterized. In many cases two or more different restriction endonucleases recognize the same sequence. Different restriction endonucleases that recognize the same nucleotide sequence and that cleave it in the same position are called **isoschizomers.** Longer recognition sequences are statistically less likely to appear within a DNA molecule than are shorter ones. Restriction endonucleases that recognize 8-bp sequences therefore make many fewer cuts in a DNA molecule than those that recognize 4-bp sequences.

Type II restriction endonuclease recognition sites are inverted repeat sequences (**FIGURE 4.10**). When rotated 180 degrees about the central point in the plane of the page, the recognition site reads exactly as it did before the rotation. Molecular biologists call a sequence with such a dyad axis of symmetry a palindrome. Type II restriction endonucleases cut each strand within the palindrome sequence one time. Some type II enzymes cut each strand at the axis of symmetry to generate **flush** or **blunt ends** (Figure 4.10a). Others make staggered cuts (cuts that are symmetrically placed around the axis of symmetry) to generate **cohesive** or **sticky ends** (Figure 4.10b). In either case phosphodiester bond cleavage generates 3′-hydroxyl and 5′-phosphate ends. **TABLE 4.2** lists some type II restriction endonucleases along with their recognition sites.

Important insights concerning enzyme–DNA interactions have been gained by examining the way restriction endonucleases act on DNA. To be effective in protecting the host cell from viral attack, the host's restriction endonuclease has to cut the viral DNA before its modification enzyme can methylate the DNA. A restriction endonuclease must therefore be able to reach its target site quickly. It does

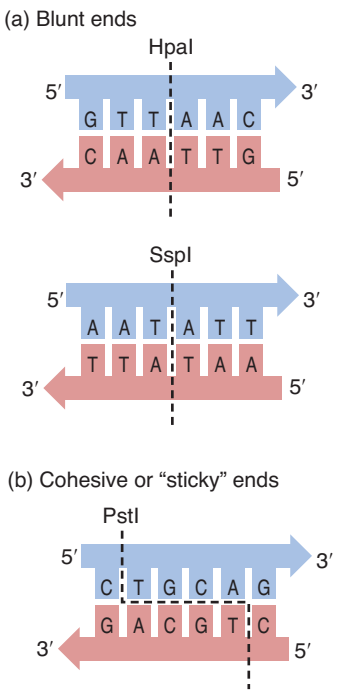


FIGURE 4.10 Restriction endonuclease generated ends. (a) Some restriction endonucleases such as HpaI and SspI cut on the line of symmetry to produce blunt ends. (b) Other restriction endonucleases such as PstI cut on either side of the line of symmetry to produce cohesive ends (also called "sticky" or staggered ends). (Adapted from an illustration by Maria Price Raposa, Carolina Biological Supply Company.)

TABLE 4.2 Sequence Specificity of Some Restriction Endonucleases

Organism of Origin	Restriction Endonuclease	Recognition Sequence
Arthrobacter luteus	AluI	5′…A G↓CT … 3′
Anebaena variabilis	Ava I	5′...C↓PyCGPuG...3′[a]
Bacillus amyloliquefaciens H	Bam HI	5′...G↓GATCC...3′
Bacillus globigii	Bgl HI	5′...A↓GATCT...3′
Escherichia coli RY13	Eco RI	5′...G↓AATTC...3′
Escherichia coli J62 pLG74	Eco RV	5′...GAT↓ATC...3′
Haemophilus aegyptius	Hae II	5′... PuGCGC↓Py...3′
Haemophilus aegyptius	Hae III	5′...GG↓CC...3′
Haemophilus haemolyticus	Hha I	5′...GCG↓C...3′
Haemophilus influenzae Rd	Hind II	5′...GTPy↓PuAC...3′
Haemophilus influenzae Rd	Hind III	5′...A↓AGCTT…3′
Haemophilus parainfluenzae	Hpa I	5′...GTT↓AAC...3′
Haemophilus parainfluenzae	Hpa II	5′...C↓CGG...3′
Kliebsiella pneumoniae	Kpn I	5′...GGTAC↓C...3′
Moraxella bovis	Mbo I	5′... ↓GATC...3′
Nocardia otitidis-caviarum	Not I	5′...GC↓GGCCGC...3′
Providencia stuartii	Pst I	5′...CTGCA↓G...3′
Serratia marcescens	Sma I	5′...CCC↓GGG...3′
Streptomyces stanford	Sst I	5′...GAGCT↓C...3′
Xanthomonas malvacearum	XmaI	5′...C↓CCGGG...3′

[a]Py, pyrimidine; Pu, purine.

so by first binding to the foreign DNA in a nonspecific fashion and then scanning the DNA for recognition sequences. This linear diffusion along the DNA molecule allows the restriction endonuclease to find and cleave all recognition sites rapidly and efficiently. The alternative possibility, a three-dimensional diffusion process in which the endonuclease alternately associates with and dissociates from DNA until it finally binds to the recognition sequence, would be too slow to protect the host from viral replication.

Restriction endonucleases can be used to construct a restriction map of a DNA molecule.

Because a particular type II restriction endonuclease recognizes a unique sequence, it can only make a limited number of cuts in an organism's DNA. For example, an endonuclease that recognizes a 6-bp sequence will cut a typical bacterial DNA molecule, which contains roughly 3×10^6 bp, into a few hundred to a few thousand fragments. Smaller DNA molecules, such as phage or plasmid DNA molecules, may have fewer than 10 sites of cutting (frequently, one or two and, often, none). Because of the specificity just mentioned, a particular restriction endonuclease generates a unique family of fragments from a particular DNA molecule. Another restriction endonuclease will generate a different family of fragments from the same DNA molecule.

The family of fragments generated by a single restriction endonuclease can usually be resolved by agarose gel electrophoresis. A fragment's length is determined by comparing its mobility with the mobility of fragments of known length run concurrently. A **restriction map** shows the restriction endonuclease cutting sites along a DNA molecule. One approach to constructing a restriction map is to cut the DNA sample in three ways: with one restriction endonuclease, with a second restriction endonuclease, and with both enzymes (a double digest). The fragments generated are resolved by gel electrophoresis, which also reveals the number of fragments and their lengths in kilobase pair (kbp). The fragmentation pattern and fragment lengths are used to deduce the order of restriction endonuclease cut sites and to assign intervals between them. Because actual lengths are determined by comparing fragment mobilities with the mobilities of standards of known length, all intervals in the map are additive.

A simple example will help to illustrate this approach (**FIGURE 4.11**). Samples of a 6-kbp linear duplex are cut by BamHI, EcoRV, or both endonucleases into fragments, which are separated by gel electrophoresis as shown in Figure 4.11a. EcoRV cleaves at one site to produce two fragments (2.3 and 3.7 kbp). Inspection of restriction fragments produced by the BamHI and EcoRV mixture reveals that BamHI cuts the 3.7-kbp fragment into two pieces (1.2 and 2.5 kbp) and the 2.3-kbp fragment into two pieces (0.5 and 1.8 kbp). BamHI cleaves at two sites to produce three fragments (4.3 kbp, 1.2 kbp, and 0.5 kbp). A comparison of the fragments produced by BamHI with those

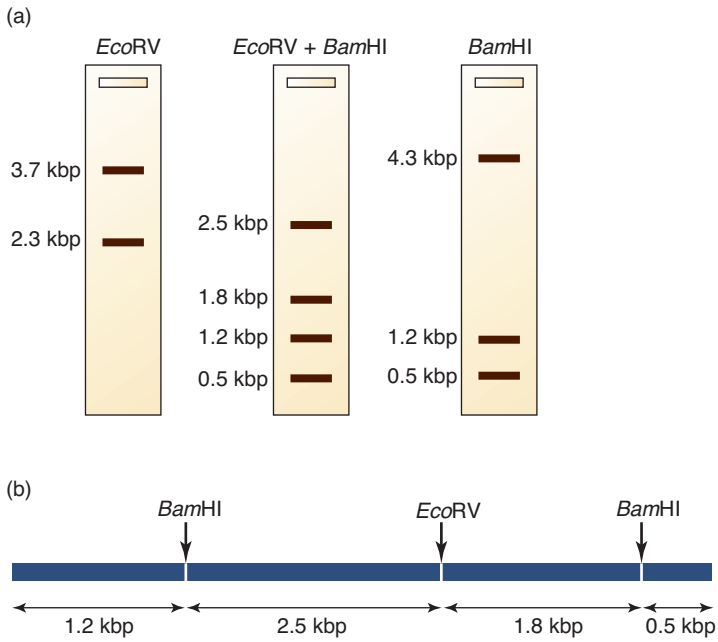


FIGURE 4.11 Restriction map construction. (a) Gel electrophoresis pattern of fragments produced by digesting the same DNA molecule with the restriction endonuclease(s) shown at the top of each lane. Fragment sizes are given in kilobase pairs (kbp). (b) The restriction map deduced from the fragment sizes indicated by gel electrophoresis. Specific restriction site(s) for each enzyme are shown by the vertical arrows.

produced by the double digest indicates that EcoRV cuts the 4.3-kbp fragment into a 2.5-kbp and a 1.8-kbp fragment. This information allows us to determine how the 3.7- and 2.3-kbp fragments produced by EcoRV are joined, yielding the restriction map shown in Figure 4.11b.

Although the principles are simple, the approach does have practical limitations. Some fragments are so small they are difficult to see or they move off the gel. Under these conditions the sum of the fragment sizes will not add up to the size of the original DNA. Two (or more) fragments may be the same size or have such similar sizes the gel does not resolve them. When this occurs one of the bands in the agarose gel will appear to have twice the fluorescence intensity as neighboring bands of similar size in the same gel. Finally, interpretation of results becomes progressively more difficult as the number of bands increase. Results become difficult, if not impossible, to interpret when each restriction endonuclease produces more than seven fragments.

In 1976 Hamilton Smith and Max L. Birnstiel devised a simple method for DNA restriction site mapping. This method reduces the computational complexity by starting with a DNA molecule with a radioactive label at just one end. It exploits the ability of the enzyme polynucleotide kinase to transfer the γ-phosphoryl group from ATP to the free hydroxyl group at the 5′-end of a DNA or RNA molecule. Labeling DNA at one end requires time and effort but can be achieved by the approach shown in **FIGURE 4.12**. The DNA is treated with phosphatase to remove phosphate groups from the 5′-ends, incubated with polynucleotide kinase and [γ-^{32}P]ATP to add labeled phosphate groups

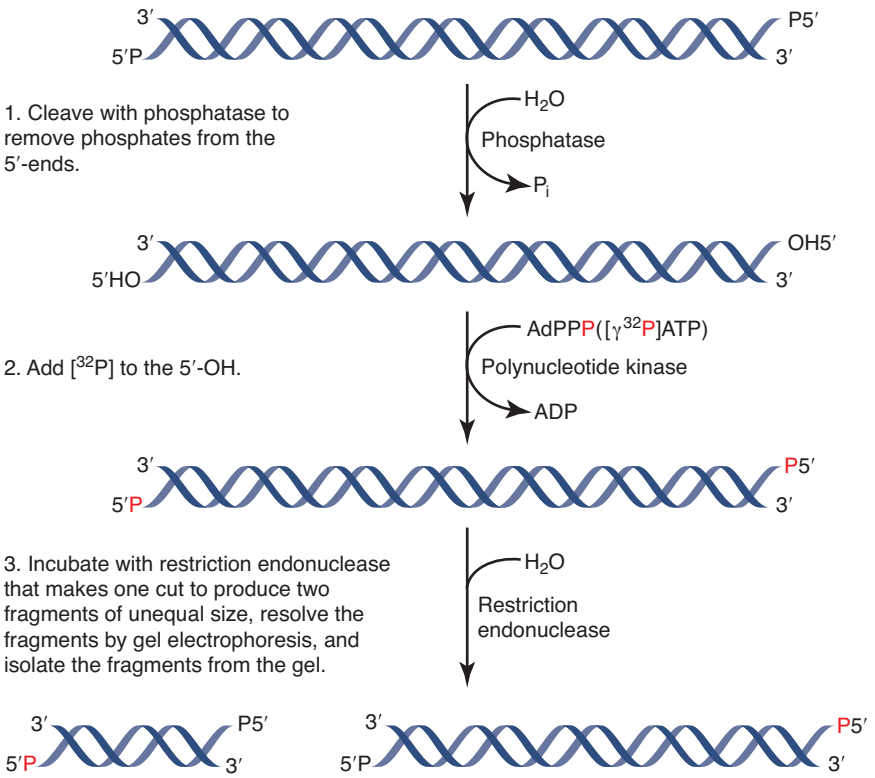


FIGURE 4.12 Method for introducing [^{32}P] at the 5′-end of DNA.

to the 5′-ends, and then cut with a restriction endonuclease to produce two DNA fragments of unequal size (each with a single labeled end) that are purified by gel electrophoresis.

The restriction map of each labeled fragment can then be determined by placing the labeled fragment in a tube, partially digesting it with a new restriction endonuclease, and then resolving the fragments by gel electrophoresis. A photographic film is then placed on top of the gel and stored in the dark. The β particles released by the ^{32}P reduce the silver ions in the photographic film, producing dark bands over each of the fragments. This autoradiographic detection technique reveals patterns that are easy to interpret, with each lane containing bands of fragments that start at the labeled end and extend to a site to be mapped.

The order of the sites is obtained by reading the bands on the gel from the bottom to the top, i.e., from the smallest fragment (site closest to the radioactively labeled end) to the largest fragment (site furthest from the labeled end). Intervals between cutting sites correspond to differences between fragment sizes. Although simple in principle this method has several shortcomings. As fragments get longer, measurements of chain length become less precise. It is also difficult to find conditions that provide the desired level of cleavage. Furthermore, the rate of cutting is affected by the nucleotides around the cleavage site so that some cutting sites will be cleaved to a greater extent than others.

DNA fragments can be inserted into plasmid DNA vectors.

It is often desirable to generate large amounts of a particular DNA fragment. Although one might guess that replication of a DNA fragment could be accomplished by first introducing the fragment into a cell and then using the cell's replication machinery to do the work of replicating the fragment, this approach rarely works. The reason for the high rate of failure is that DNA molecules must have specific replication sequences before a cell's replication machinery can act on them. Most DNA fragments lack these sequences. There is a simple way to circumvent the problem, however. Insert the DNA fragment into a DNA molecule that has the required replication sequence and then introduce the recombinant DNA molecule into the cell.

Plasmids are excellent carriers or **vectors** for DNA fragments. As illustrated in **FIGURE 4.13**, the insertion of the DNA fragment into a plasmid is in principle a simple process. A plasmid bearing an antibiotic resistance gene is cut with a restriction endonuclease to produce cohesive ends that are complementary to the cohesive ends of the DNA fragment. *The cut must be at a site that is outside both the antibiotic resistance gene and the sequence required for plasmid DNA replication.* The DNA fragment is inserted into the plasmid DNA by joining the ends of the DNA fragment with the ends of the plasmid DNA. An enzyme called DNA ligase catalyzes the joining or ligation process. The **recombinant plasmid**, which contains the DNA fragment inserted within the plasmid vector, then is introduced into living bacterial cells that have been made competent to accept plasmid DNA.

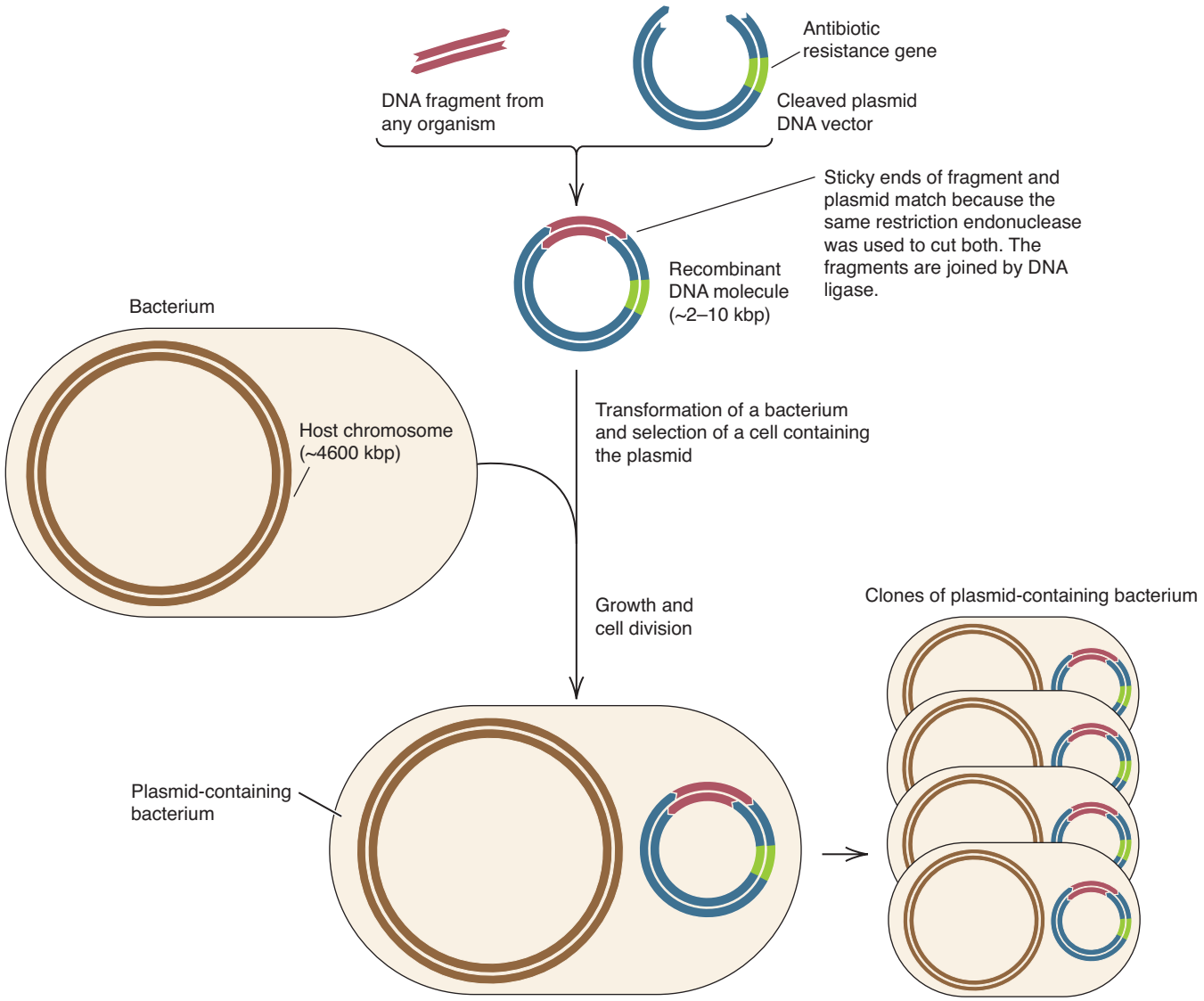


FIGURE 4.13 An example of cloning. A fragment of DNA from any organism is joined to a cleaved plasmid. The recombinant plasmid is then used to transform a bacterial cell, where the recombinant plasmid is replicated and transmitted to the progeny bacteria. The bacterial host chromosome is not drawn to scale. It is typically about 1,000 times larger than the plasmid.

One method for making *Escherichia coli* cells **competent** to take up DNA from the surrounding medium is to incubate the cells with 100 mM calcium chloride solution at 4°C for 30 minutes, centrifuge, resuspend the bacterial pellet in a 75 mM calcium chloride solution containing 15% glycerol, and then freeze 0.2 mL cell samples. The cells are transformed by thawing the samples on ice, adding recombinant DNA, incubating the suspension at 42°C for 90 seconds, and then spreading the suspension on an agar plate containing the selective antibiotic. Each cell containing a recombinant plasmid produces a colony or clone consisting of millions of progeny bacteria with the same recombinant plasmid. For this reason the recombinant plasmid is said to be cloned, and the process is called **cloning**. Other types of autonomously replicating DNA molecules such as viral DNA are also used as cloning vectors.

The Southern blot procedure is used to detect specific DNA fragments.

The availability of restriction endonucleases that cleave DNA into unique and fairly small fragments makes it possible to detect small segments within large DNA molecules by taking advantage of the fact that complementary polynucleotide strands (DNA or RNA) can anneal to form double-stranded molecules. In 1975 Edwin M. Southern invented a technique that allows investigators to hybridize a specific polynucleotide (DNA or RNA) to a large number of particular DNA segments without the necessity of purifying individual DNA fragments. This technique, called the **Southern transfer** or **Southern blot procedure** after its inventor, exploits the fact that very thin nitrocellulose or nylon membranes tightly bind single-stranded DNA fragments. The bases of the bound fragments remain free to form hydrogen bonds with complementary single strands of DNA or RNA.

In the Southern blot procedure (**FIGURE 4.14**) one or more restriction endonucleases completely digest DNA and the resulting fragments are separated according to size by gel electrophoresis. After separation is complete the gel is soaked in a sodium hydroxide solution to denature the DNA. Then the gel is rinsed with distilled water, soaked in buffer solution to adjust the pH, and placed on a sponge that is itself in a reservoir of buffer solution. The top surface of the gel is covered with a nitrocellulose or nylon membrane, which in turn is covered with several layers of dry paper towels, and a heavy weight is placed on top of the paper towels. The paper towels act as a blotter, drawing buffer solution from the reservoir through the various layers by capillary action. As the buffer solution moves up, it carries single-stranded DNA from the gel to the nitrocellulose or nylon membrane. In a variation of this procedure, an electrophoretic process called **electroblotting** transfers denatured DNA. In either case the single-stranded DNA molecules bind to positions on the membrane identical to their positions on the agarose gel, preserving the band pattern.

After transfer is complete the setup is disassembled, and the single-stranded DNA is fixed permanently to the nitrocellulose membrane by baking at 80°C in a vacuum or to the nylon membrane by ultraviolet light–induced cross-linking. Then the membrane is incubated in a solution containing bovine serum albumin, polysucrose, polyvinylpyrolidine, and denatured salmon sperm DNA to eliminate the membrane's inherent ability to bind single-stranded DNA. Denatured DNA that was transferred to the membrane from the electrophoresis gel, however, retains its ability to bind complementary single-stranded DNA or RNA molecules.

DNA sequences of interest bound to the membrane are identified by using a complementary single-stranded DNA or RNA probe with a detectable label. The nitrocellulose or nylon membrane is placed in a buffer solution containing the labeled probe and incubated for several hours at a suitable renaturation temperature to permit the probe to hybridize to its complementary sequence in the DNA bound to the nitrocellulose or nylon membrane. Then the membrane is washed with buffer to remove unbound probe and dried. Autoradiography is used to detect a radioactive probe. DNA fragments complementary to the probe appear as black or stained bands.

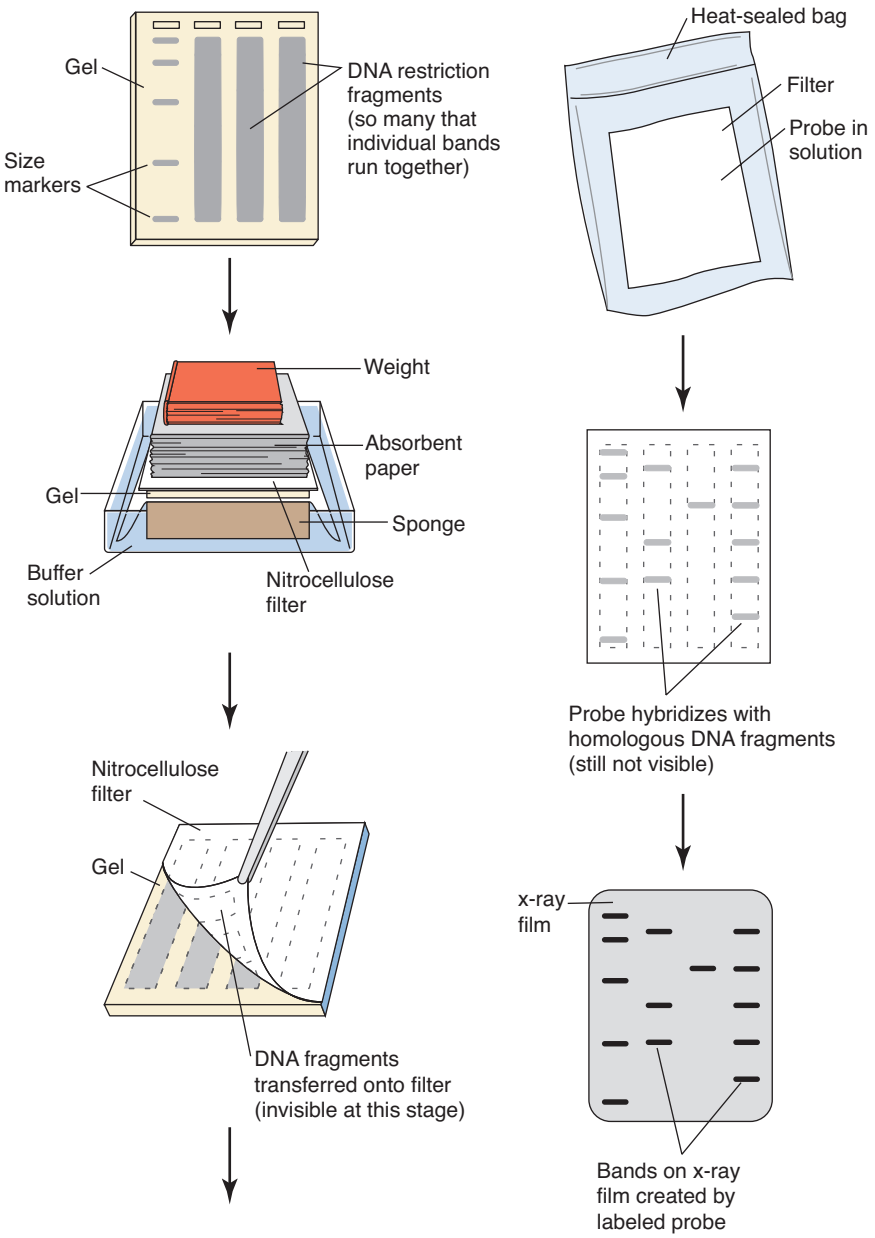


FIGURE 4.14 Southern blot analysis: an experimental method for identifying a specific DNA fragment in a gel. DNA is cleaved by one or more restriction endonucleases and the resulting fragments are separated by gel electrophoresis, the resolved DNA fragments are denatured by soaking the gel in a basic solution and denatured fragments are blotted onto a nitrocellulose filter, the nitrocellulose filter is exposed to a probe with a radioactive or other detectable label, and when a radioactive probe is used, the filter is exposed to photographic film and the film is developed.

FIGURE 4.15 Southern blot of genomic DNA from normal tissue (N) and from a tumor (T) from a patient with breast cancer. Two bands are visible in the normal breast tissue but only the upper band is present in the tumor cells. The faint lower band represents contribution from non-tumor cells in the tumor sample. (Reproduced from Tougas L., et al. 1996. *Clin Invest Med* 19:222–230. Photo courtesy of Serge Jothy, University of Toronto.)

The Southern blot procedure is a very versatile and powerful tool that detects changes in DNA that alter restriction cutting sites or the lengths of segments between restriction cutting sites. It has become an important clinical laboratory tool for identifying genetic problems. For example, the Southern blot in FIGURE 4.15 shows that a band present

in normal breast tissue is missing from tumor cells. Today, investigators usually prefer to use a DNA or RNA probe with a fluorescent tag when performing the Southern blot procedure because they can see fluorescent bands of interest without performing the time-consuming autoradiography step and can work without taking the precautions needed when working with radioisotopes.

Northern and Western blotting are used to detect specific RNA and polypeptide molecules, respectively.

In a variant of Southern blotting, called **Northern blotting** (a play on words), RNA molecules are separated by gel electrophoresis and then transferred to nylon or nitrocellulose membranes. Labeled single-stranded RNA or DNA probes can hybridize with the bound RNA. Still another variant is **Western blotting,** in which proteins are separated by gel electrophoresis and then transferred to a nitrocellullose or polyvinyl membrane. Bound protein is detected with labeled or tagged antibody probes.

DNA polymerase I requires a template-primer.

DNA polymerases, enzymes that catalyze DNA synthesis, are among the most important tools in the molecular biologist's toolkit. Several different DNA polymerases are commonly used in the modern molecular biology laboratory. Polymerase selection is determined by the demands of the experiment. For instance, an investigator may decide to use a heat-resistant DNA polymerase isolated from a hyperthermophilic archaeon when the experiment requires the polymerase to retain full activity after exposure to heat. A great deal of our knowledge about how DNA polymerases work derives from studies of DNA polymerase I, an enzyme that was first detected in *E. coli* extracts by Arthur Kornberg and coworkers in 1956. The Roman numeral I distinguishes Kornberg's enzyme from other DNA polymerases that were subsequently discovered to be present in *E. coli*. DNA polymerase I, like all other DNA polymerases, converts deoxyribonucleoside triphosphates into DNA in the presence of preformed DNA according to the following reaction:

$$\left\{\begin{array}{c}\text{dATP}\\+\\\text{dGTP}\\+\\\text{dCTP}\\+\\\text{dTTP}\end{array}\right\} \underset{\text{DNA polymerase, } Mg^{2+}}{\overset{\text{DNA}}{\rightleftharpoons}} \text{DNA + pyrophospate}$$

The enzyme is assayed by adding it to a mixture that also contains the four deoxyribonucleoside triphosphates, one of which is radioactive, a DNA template-primer (see below), and magnesium ions. The reaction is stopped by adding acid, which also precipitates newly formed radioactive DNA. Then the acid-insoluble DNA is separated from the acid-soluble deoxyribonucleoside triphosphates by centrifugation or filtration and its radioactivity is determined in a liquid

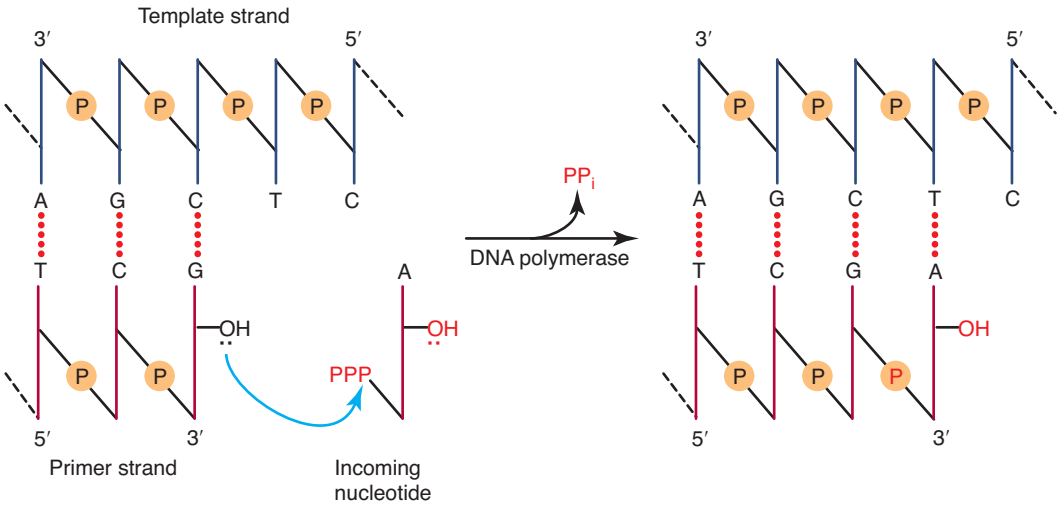


FIGURE 4.16 DNA polymerase I: polymerase function. DNA polymerase I uses the template strand to determine the next nucleotide that it adds to the primer strand.

scintillation counter. All four deoxyribonucleotides must be present for the reaction to take place. The source of the preformed double-stranded DNA does not matter; prokaryotic and eukaryotic DNA work equally well. The two strands of the preformed DNA have different functions at the site of DNA synthesis (**FIGURE 4.16**). One strand, the **primer strand**, is the site of attachment for incoming deoxyribonucleotides. The other strand, the **template strand**, determines the order of attachment to the primer strand according to Watson-Crick base pairing rules. DNA polymerase I has binding sites for the primer strand, the template strand, and the incoming deoxyribonucleotide. The enzyme catalyzes the nucleophilic displacement of a pyrophosphate group, creating a phosphodiester bond between the 3′-end of the growing primer strand and the 5′-end of the incoming deoxyribonucleotide. Hence, the primer strand grows in a 5′→3′ direction, a characteristic of all known DNA and RNA polymerases. Inorganic pyrophosphate hydrolysis drives the reaction to completion inside the cell.

DNA polymerase I has both 3′→5′ and 5′→3′ exonuclease activities.

Early preparations of *E. coli* DNA polymerase I contained exonuclease activity, which Kornberg and coworkers expected to remove by further enzyme purification. Contrary to their expectation, highly purified DNA polymerase I, consisting of a single polypeptide with 928 amino acid residues (molecular mass = 103 kDa), has both 3′→5′ and 5′→3′ exonuclease activities. The discovery that a single enzyme has the ability to both synthesize and degrade DNA was mystifying. Further studies by Kornberg's group, however, showed the two-exonuclease activities, in fact, do make important contributions to DNA synthesis. The 3′→5′ exonuclease, which catalyzes the sequential hydrolysis of one 5′-deoxynucleotide at a time from the 3′-end, has a proofreading function (**FIGURE 4.17**).

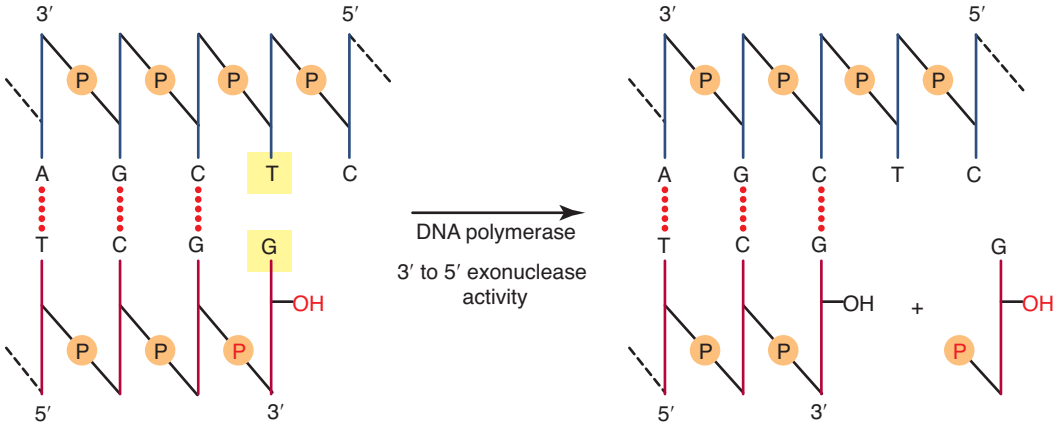


FIGURE 4.17 DNA polymerase I: proofreading function. Although rare, DNA polymerase I does misinsert nucleotides. The 3′→5′ exonuclease activity allows DNA polymerase to remove a misinserted deoxyribonucleotide before the next deoxyribonucleotide is added to the growing primer chain.

Although DNA polymerase I has great specificity, it does make occasional errors either by misinserting a nucleotide or by misaligning the template-primer. Either type of error eventually results in base pair substitution, addition, or deletion. Although errors are rare, occurring at a frequency of about one nucleotide for every 10^4 nucleotides added, failure to correct an error is harmful to the cell. The 3′→5′ exonuclease activity removes the misinsertions before the next deoxyribonucleotide is added to the growing chain. The 5′→3′ exonuclease, which can remove either a mononucleotide or a short oligonucleotide from the 5′-end, plays an editing role in DNA synthesis and, as described below, serves as a useful tool for preparing labeled DNA.

DNA polymerase I is a multifunctional polypeptide (a single polypeptide with more than one catalytic activity) with three domains. Domain 1, which includes the first 325 or so residues at the amino or N-terminus, has 5′→3′ exonuclease activity. Domain 2, consisting of the next 200 or so residues, has 3′→5′ exonuclease activity. Domain 3, which includes the remaining residues at the carboxyl or C-terminus, has polymerase activity. In 1970 Hans Klenow demonstrated that subtilisin (a proteolytic enzyme produced by *Bacillus subtilis*) cleaves DNA polymerase I into two fragments of unequal length (**FIGURE 4.18**). The smaller fragment (molecular mass = 35 kDa) containing domain 1 has 5′→3′ exonuclease activity. The larger one (molecular mass = 68 kDa), known as the **Klenow fragment**, contains domain 2 (3′→5′ exonuclease activity) and domain 3 (polymerase activity).

Thomas Steitz and coworkers determined the three-dimensional structure of the Klenow fragment by x-ray crystallography. The polymerase domain, which is made of three subdomains, resembles a partially opened right hand (**FIGURE 4.19**). The polymerase catalytic site is located in the palm subdomain. As shown in the schematic diagram in **FIGURE 4.20a**, the thumb subdomain appears to contact the minor groove of the primer-template duplex, whereas the finger subdomain binds the uncopied template strand. Three-acidic residues (Asp-705, Asp-882, and Glu-883) form a carboxylate triad (Figure 4.19), which is essential for

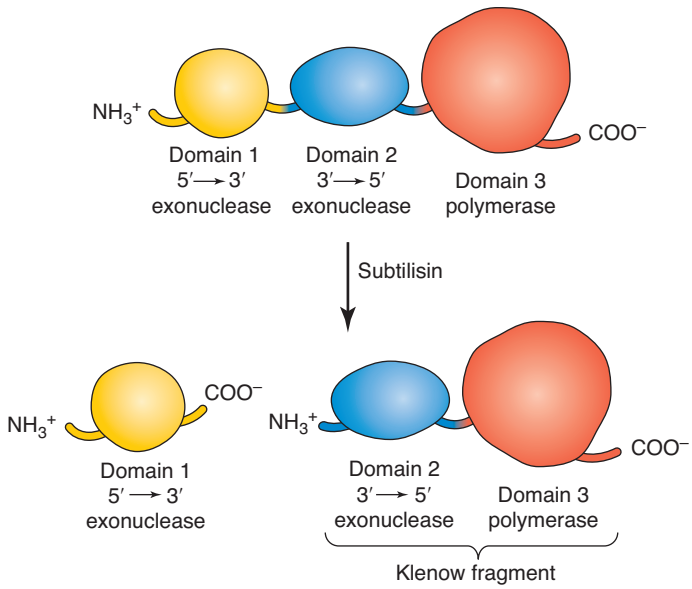


FIGURE 4.18 DNA polymerase I. Generation of a Klenow fragment.

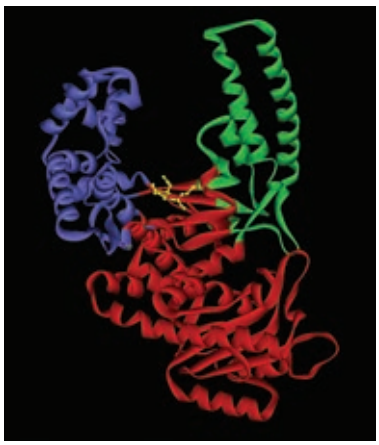


FIGURE 4.19 Ribbon structure of the polymerase domain of the Klenow fragment. The polymerase domain resembles a partially opened right hand. The subdomains corresponding to the thumb, fingers, and palm are colored green, purple, and red, respectively. The carboxylate triad, Asp-705, Asp-882, and Glu-883, at the polymerase active site is shown in yellow. (Structure from the Protein Data Bank ID: 1KFD. Beese, L. S., Friedman, J. M., and Steitz, T. A. 1993. *Biochemistry* 32:14095–14101. Prepared by B. E. Tropp.)

polymerase activity. Two magnesium ions appear to be integral parts of the catalytic site.

Kinetic studies suggest the polymerase stalls when an incorrect nucleotide is attached to the growing end of the primer chain, allowing time for the 3′→5′ exonuclease to remove the incorrect nucleotide (Figure 4.20b). X-ray diffraction studies indicate the polymerase and 3′→5′ exonuclease catalytic sites are separated by about 3.5 nm. Proofreading can occur by an intra- or

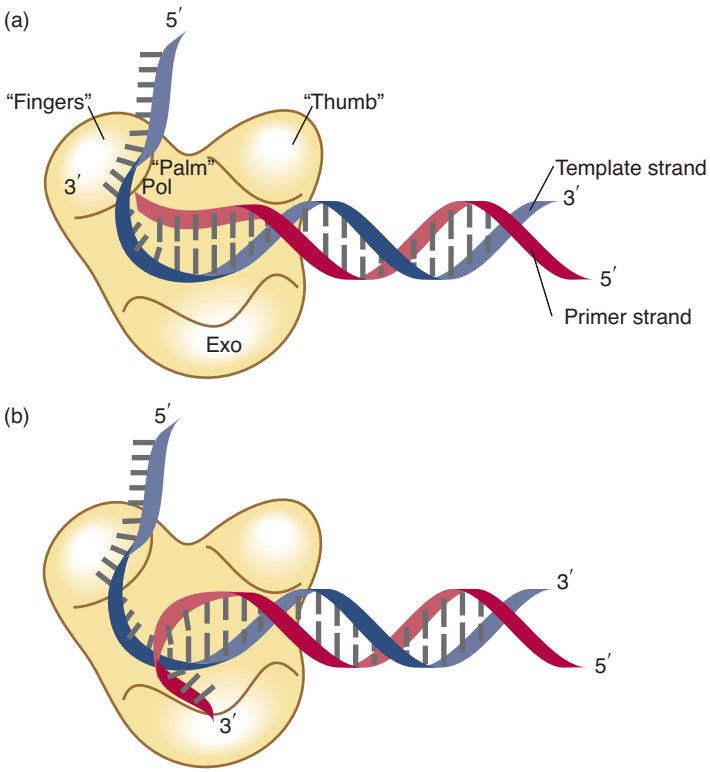


FIGURE 4.20 Schematic of DNA polymerase I in the (a) polymerase mode and (b) proofreading mode. (a) DNA polymerase I in the polymerase mode. The thumb subdomain contacts the minor groove of the template-primer duplex, the finger subdomain binds the uncopied template strand, and the 3′-OH end of the primer strand binds to the polymerase active site. (b) DNA polymerase I in the 3′→5′-exonuclease mode. The thumb subdomain still contacts the minor groove of the primer-template duplex and the finger subdomain still binds the uncopied template strand. However, the 3′-OH end of the primer strand with the misinserted nucleotide binds to the 3′→5′ exonuclease catalytic site active site. (Adapted from Joyce, C. M., and Steitz, T. A. 1995. J *Bacteriol* 177:6321–6329.)

intermolecular mechanism. In the former case the 3′-end moves from the polymerase catalytic site to the 3′→5′ exonuclease catalytic site, whereas in the latter enzyme–DNA complex dissociates and then a different DNA polymerase I removes the mispaired base. A zinc ion and a magnesium ion at the exonuclease catalytic site appear to position a water molecule for a nucleophilic attack on the 5′-phosphate of the mismatched nucleotide.

DNA polymerase I can catalyze nick translation.

As shown in **FIGURE 4.21**, DNA polymerase I has the remarkable ability to synthesize DNA at a single-strand break (nick). This synthesis requires melting the DNA beyond the nick and progressive strand displacement of the 5′-end. When the 5′→3′ exonuclease activity is low or missing (as in the Klenow fragment), DNA synthesis proceeds with strand displacement. When 5′→3′ exonuclease activity is present, however, the 5′-end of the displaced strand is digested and the nick moves along the molecule in the direction of synthesis in a process known as **nick translation**. Using radioactive deoxyribonucleoside triphosphates during nick translation converts unlabeled DNA into radioactive DNA with the same nucleotide sequence. Hence, nick translation provides a convenient means for preparing radioactive

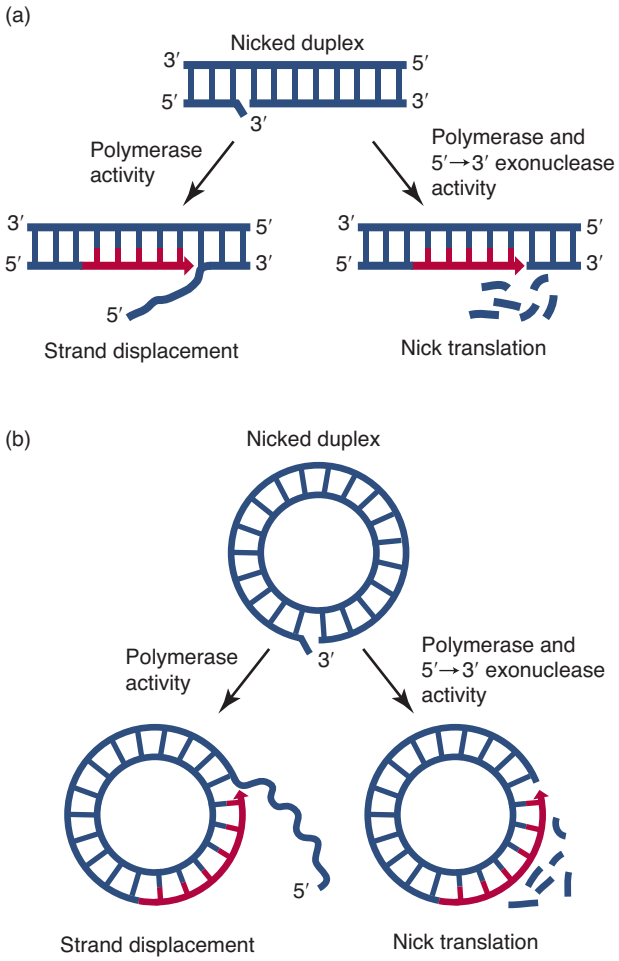


FIGURE 4.21 Strand displacement and nick translation on linear and circular molecules. In nick translation one nucleotide is removed by DNA polymerase I's 5′→3′ exonuclease activity for each nucleotide added by the enzyme's polymerase activity. The growing strand is shown in red. (a) Strand displacement and nick translation on nicked linear DNA duplex. (b) Strand displacement and nick translation on nicked circular DNA duplex.

DNA probes for Southern blotting. The same approach can be used to prepare DNA with fluorescent markers. Although nick translation allows us to synthesize DNA with a radioactive or fluorescent tag, it does not lead to the net increase in total DNA that is present.

The polymerase chain reaction (PCR) is used to amplify DNA.

It seems reasonable to suppose that DNA polymerase I should be able to catalyze the synthesis of many copies of a specific double-stranded DNA fragment. It was not until 1983, however, almost three decades after the discovery of DNA polymerase I, that a satisfactory method was finally devised to do so. The method, which was invented by Kary B. Mullis, called the **polymerase chain reaction (PCR)**, is shown in FIGURE 4.22.

The linear duplex to be amplified is heat denatured. Then an oligonucleotide primer (about 20–30 nucleotides in length) is annealed to each of the denatured single strands. Each primer oligonucleotide sequence is selected so it anneals to a sequence in the outer region of the DNA segment to be amplified (the target sequence). The primers are oriented with their 3′-ends directed toward each other so that chain extension copies the region between them. Then DNA polymerase I is added to extend each primer until it reaches the end of the template. The new strands have defined 5′-ends (the 5′-ends of the oligonucleotide primers). This completes cycle 1 of PCR amplification. The same three-step process of (1) denaturing the linear duplex, (2) annealing the primer, and (3) extending DNA primer with DNA polymerase I is repeated in cycle 2 and each of the subsequent cycles. The strands synthesized during cycle 1 (and each succeeding cycle) serve as templates for the next cycle. Repeated DNA polymerase I additions at each cycle can be avoided by using a thermostable DNA polymerase such as the Pfu or Vent DNA polymerases, which are isolated from the hyperthermophilic archaea *Pyrococcus furiosus* and *Thermococcus litoralis* respectively.

PCR machines, known as **thermal cyclers,** which are programmed to shift their temperature up and down during different stages of the cycle, take advantage of thermostable DNA polymerase to allow PCR to be automated. In theory, amplification is exponential so that after n cycles the amplification yield would be 2^n. A 30-cycle amplification therefore would be expected to produce about 10^9 copies of the target segment. This calculation assumes that each cycle proceeds with a 100% efficiency of amplification. Because the efficiency of amplification is estimated to be 60% to 85%, actual amplification yields are lower than the calculated value. Nevertheless, extraordinary amplifications are possible in just a couple of hours because each cycle requires only 4 to 6 minutes.

Among the many important practical applications of PCR are the following: (1) amplification of a segment of a large DNA molecule for subsequent use in genetic engineering, (2) rapid detection of pathogenic bacteria and viruses, (3) detection of inborn errors of metabolism, and (4) detection of tumors. The popular press has described many situations in which PCR has been used in criminal investigations to amplify DNA from saliva, blood, sperm cells, or even a single hair. The amplified DNA is then characterized by restriction endonuclease digestion followed by gel electrophoresis. Amplified DNA can also be sequenced (see below).

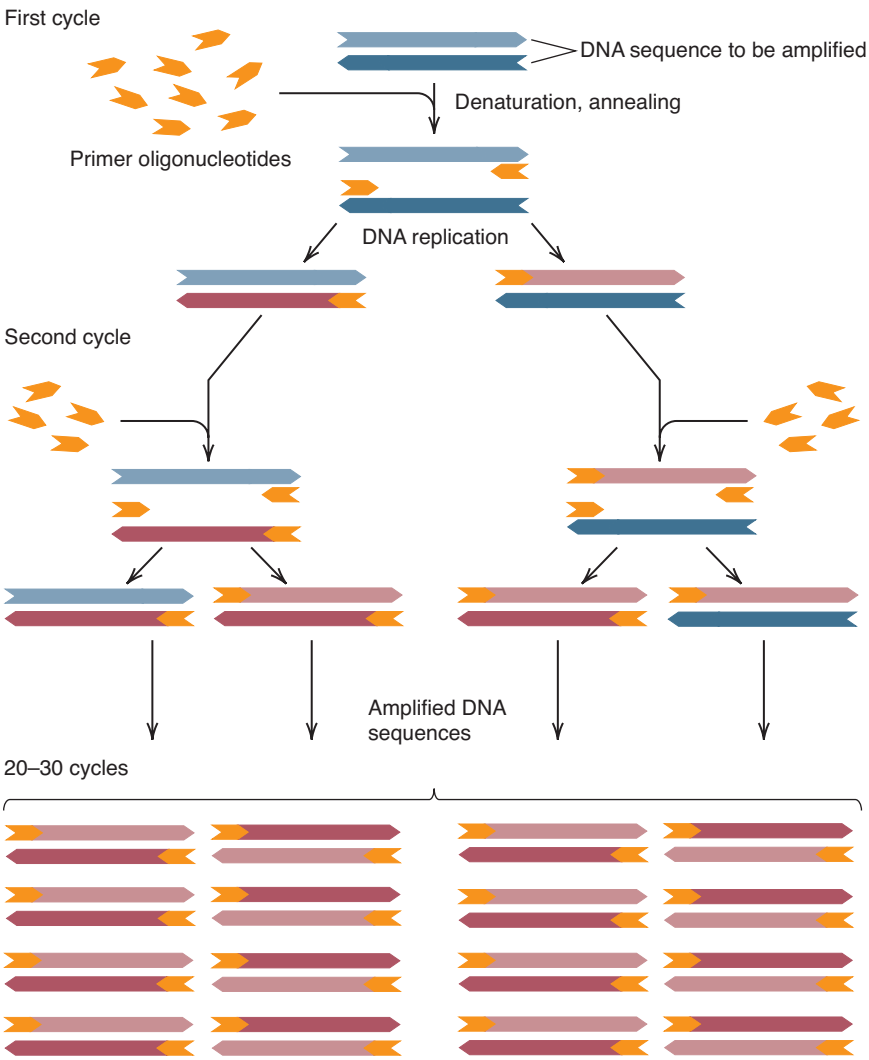


FIGURE 4.22 Schematic for the polymerase chain reaction (PCR), a method for amplifying a specific target DNA sequence. The linear duplex to be amplified (blue) is heat denatured. Then an oligonucleotide primer (orange) is annealed to each of the denatured single strands. DNA polymerase is added to extend each primer. The new strands (red) have defined 5′-ends (the 5′-ends of the oligonucleotide primers). This completes cycle 1 of PCR amplification. The same three-step process of (1) denaturing the linear duplex, (2) annealing the primer, and (3) extending DNA primer with DNA polymerase I is repeated in cycle 2 and each of the subsequent cycles.

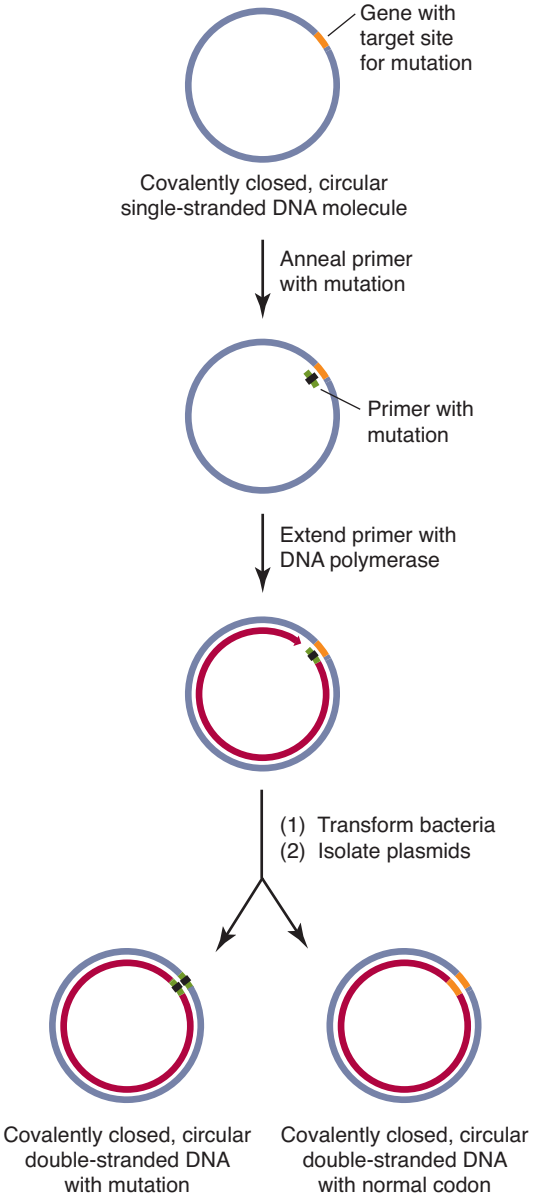


FIGURE 4.23 Schematic for introducing a mutation into a gene by site-directed mutagenesis.

Site-directed mutagenesis can be used to introduce a specific base change within a gene.

Once a gene has been cloned and sequenced (see below) it is possible to introduce specific base changes anywhere within the gene. The technique for introducing the specific base changes, called **site-directed mutagenesis**, first conceived by Michael Smith in the mid-1970s, is a very powerful tool for studying how specific amino acid changes influence protein function. FIGURE 4.23 is a simplified schematic for using site-directed mutagenesis to introduce a specific mutation in a gene. A primer with the desired base change(s) is annealed

to a closed covalent, circular single-stranded DNA molecule. Then DNA polymerase is added to produce a nicked circular strand. The nicked circle bearing the mutation is introduced into competent bacterial cells, which repair the nick in the mutated plasmid, and the cloned mutated plasmid can be isolated from the transformed bacterial strain.

The chain termination method for sequencing DNA uses dideoxynucleotides to interrupt DNA synthesis.

Frederick Sanger recognized that DNA polymerase's ability to add nucleotides to a primer under the direction of a template could be used to sequence DNA. The Sanger method (also called the **chain termination method**) uses small amounts of dideoxynucleoside triphosphates to cause random termination of primer chain extension. Dideoxynucleoside triphosphates are deoxynucleoside triphosphate analogs that have a hydrogen atom attached to C-3′ in place of a hydroxyl group (**FIGURE 4.24**). Once DNA polymerase has transferred a dideoxynucleotide to the 3′-end of a growing primer chain, further chain extension is impossible because the primer chain must have a 3′-hydroxyl group for the next deoxynucleotide to be attached.

The major features of the chain termination method are summarized in **FIGURE 4.25**. Four sequencing reaction mixtures are prepared. The mixtures differ only in the dideoxynucleoside triphosphate that is added. Each kind of dideoxynucleoside triphosphate is labeled with a different color fluorescent dye. All other components, including the single-stranded DNA molecule to be sequenced, DNA polymerase, the oligonucleotide primer, and the four deoxynucleoside triphosphates, are the same in all four reaction mixtures. The primer oligonucleotide, which determines the particular region within the DNA molecule that will be sequenced, is usually synthesized by chemical methods. Primer chain extension continues in a 5′→3′ direction until the process is terminated by random dideoxynucleotide attachment. As a result of this random chain termination each reaction mixture produces a nested population ladder of extended primer sequences in which all extended chains end with the specific dideoxynucleotide that was added to that reaction mixture. Once synthesis is complete the newly synthesized DNA molecules are denatured and the contents of the four different tubes are combined so the fragments can be separated by polyacrylamide gel electrophoresis or capillary electrophoresis. After fragment separation is complete, sequence information can be read directly from the gel.

A major advance in the dideoxynucleotide method, known as **dideoxy-terminator cycle sequencing**, uses a thermal cycler to amplify reaction products, allowing double-stranded DNA molecules to be sequenced. The reaction mixture, containing an oligonucleotide primer complementary to a segment of only one of the two strands, goes through the rounds of denaturation, annealing, and primer extension described for the PCR. Because the reaction mixture is heated during the denaturation part of the cycle, the DNA polymerase used must

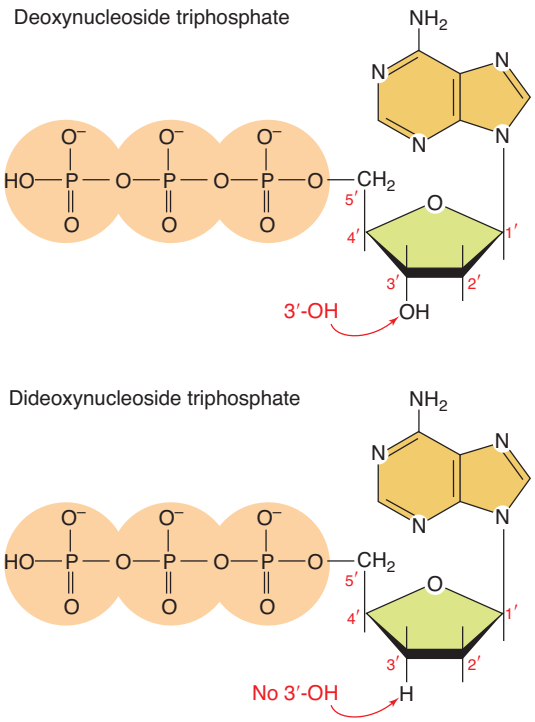


FIGURE 4.24 Comparison of deoxy- and dideoxynucleoside triphosphate structures.

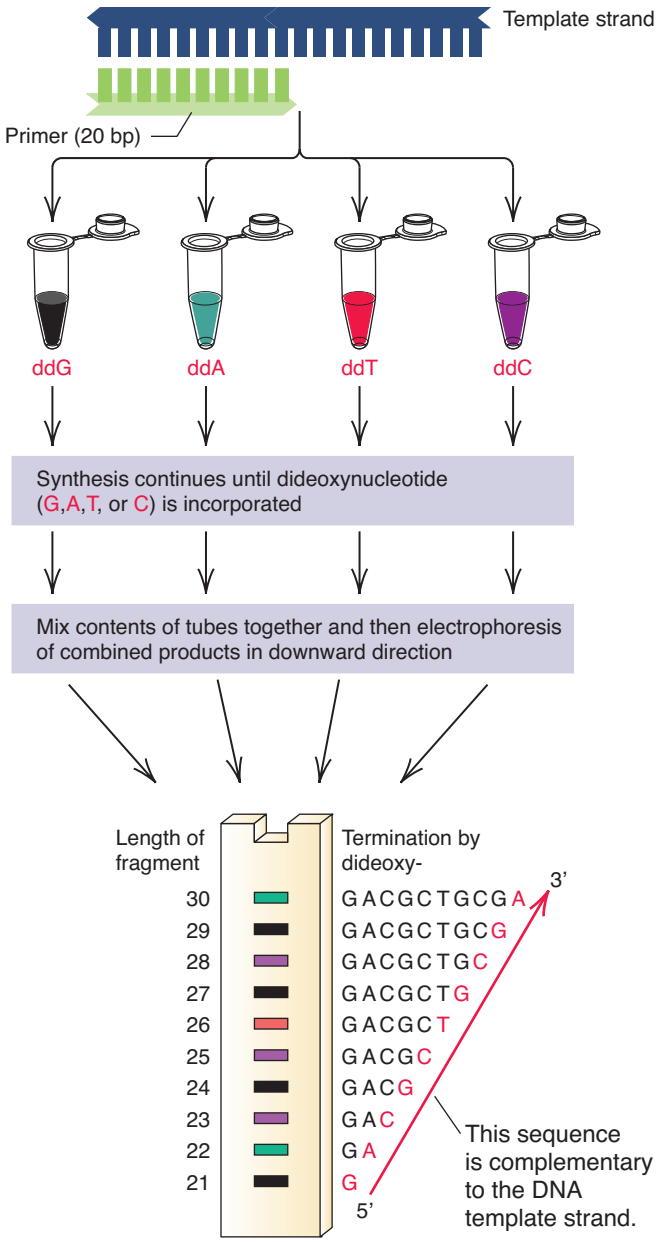


FIGURE 4.25 Chain termination method for sequencing DNA. Lengths of the terminated DNA fragments are shown at the left of the gel. The sequence of the daughter strand is read from the bottom of the gel according to the color of each band as 5′-GACGCTGCGA-3′.

be thermostable. Cycle sequencing amplifies the formation of chain extension products that are terminated by dideoxynucleotides labeled with fluorescent dyes. After the fragments have been separated by size using capillary electrophoresis, a scanner is used to detect the fragments and transfers the information to a computer that stores the nucleotide sequence data. Because the product is amplified, dideoxy-terminator cycle sequencing requires less DNA than the standard technique. Furthermore, both strands of a double-stranded DNA can be sequenced, diminishing the chance of an error.

DNA sequences can be stitched together by using information obtained from a restriction map.

A DNA restriction map can help to sequence long DNA molecules. DNA is cut with the restriction endonuclease(s) used to produce the map. Then the resulting fragments are separated by gel electrophoresis, and each is sequenced by the chain termination method. The restriction map provides the fragment order, which allows the sequences of the small fragments to be stitched together to provide the overall sequence. Although this approach seems fairly straightforward, it has drawbacks. First, the time required (and degree of difficulty) for restriction map construction increases with DNA length. Second, some restriction fragments are too long to permit sequencing without further cutting and others are so short they are lost during purification by gel electrophoresis.

Shotgun sequencing is used to sequence long DNA molecules.

The preferred approach to sequence long DNA molecules is **shotgun sequencing**, a technique devised by Sanger and coworkers in the early 1980s. The approach begins with random DNA cleavage to produce small fragments. Cleavage can be conveniently achieved by subjecting DNA in solution to the shear forces generated by passing the solution through a narrow-gauge syringe. Breakage sites are different for each DNA molecule, producing a heterogeneous population of small fragments. The term "shotgun" refers to the random shredding process. As a result of this random shredding, any given nucleotide in the original large DNA ends up in different small overlapping fragments. Many of the small fragments have 3′- or 5′-overhangs that must be converted to blunt ends before the next sequencing step is possible. This conversion is accomplished by incubating the DNA fragments with the four deoxyribonucleoside triphosphates and a DNA polymerase that fills in 3′-recessed ends and removes protruding 3′-ends to produce blunt ended fragments (FIGURE 4.26). Blunt-end fragments are incubated with polynucleotide kinase and ATP to ensure that 5′-ends are phosphorylated (not shown) and inserted into a cloning vector with a universal primer sequence at its insertion site. A DNA ligase is used that works on blunt ends. Resulting recombinant plasmids are introduced into competent bacteria by transformation. Cloned DNA is isolated, and the short fragments are sequenced using the universal primer. Sequencing reads are stored in a computer that has a sequence assembly software program that searches for overlaps and uses the overlaps to assemble the sequencing reads into a long contiguous sequence.

In 1995 J. Craig Venter, Hamilton Smith, and coworkers used a variant of shotgun sequencing called **whole genome shotgun sequencing** to determine the complete 1.83×10^6 bp genome sequence of *Haemophilus influenzae Rd*, a nonpathogenic variant of a bacterial strain that can cause ear infections and meningitis. This was the first time the genome of a free-living organism had been sequenced. As the term "whole genome shotgun sequencing" suggests, the whole bacterial genome is broken randomly into fragments that are cloned and

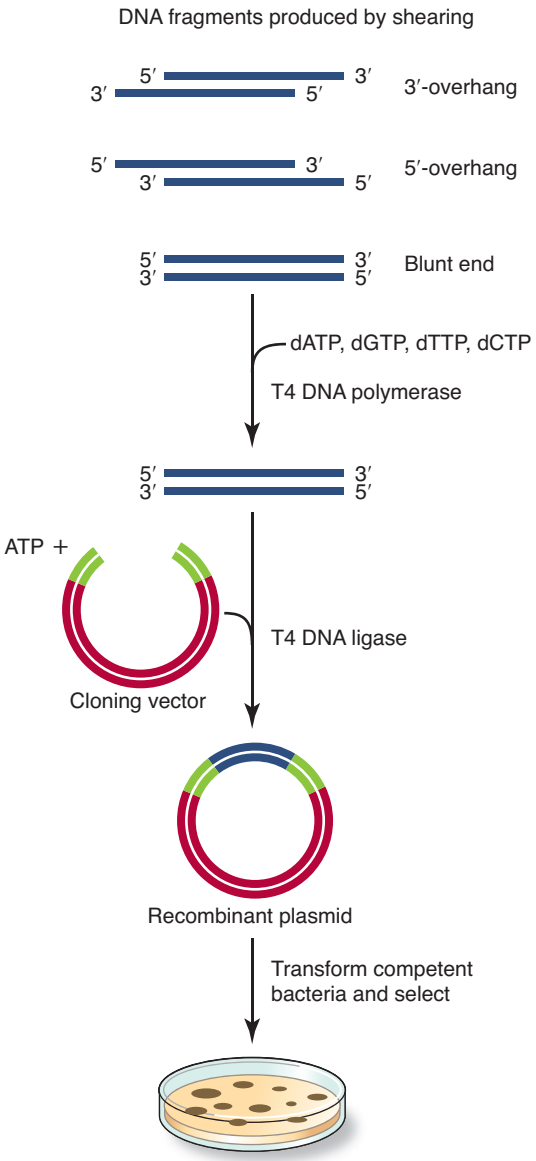


FIGURE 4.26 Recombinant plasmid construction for shotgun sequencing. Fragments formed by mechanical shearing (blue) are incubated with the four deoxyribonucleoside triphosphates and bacteriophage T4 DNA polymerase to fill 3′-recessed ends and remove protruding 3′-ends. Blunt end fragments are incubated with polynucleotide kinase and ATP to ensure that 5′-ends are phosphorylated (not shown) and inserted into a cloning vector (green) with a universal primer sequence (red) at its insertion site. Resulting recombinant plasmids are introduced into competent bacteria by transformation.

sequenced. Approximately 20,000 clones were sequenced to ensure that every nucleotide in the genome was included in the final sequence. Based on the sequence information, the *Haemophilus influenzae Rd* genome is predicted to contain 1,743 genes. Five years later Venter and a very large group of coworkers used whole genome shotgun sequencing to determine the sequence of nearly 67% of the approximately 1.8×10^8 bp genome of the fruit fly *Drosophila melanogaster*, including all genetically active sequences.

A new generation of DNA sequencing techniques is now being used for whole genome shotgun sequencing.

The chain termination method, which has been the predominant DNA sequencing technique for the past 30 years, is now being challenged by a new generation of sequencing techniques. This new generation of sequencing techniques allows whole genome shotgun sequencing to be performed much more quickly than was previously possible and at a much lower cost because they are massively parallel (sequencing thousands of DNA strands at one time) and do not require *in vivo* cloning or gel electrophoresis.

The first new generation sequencer was introduced by 454 Life Science, a member of the Roche Group, in 2005. Two years later an improved version, the Genome Sequencer FLX™ (GS FLX), went on the market. This instrument performs about 400,000 parallel sequencing reads with an average read length of 250 bp. It supplies about 10^8 bp of sequence information in 8 hours, enabling a single investigator to sequence an entire bacterial genome in a few days. Not surprisingly, many investigators have started to use this instrument (**BOX 4.3 IN THE LAB: NEW GENERATION DNA SEQUENCERS**).

The human genome sequence provides considerable new information.

The possibility of sequencing the approximately 3 billion base pairs in a single set of human chromosomes (the haploid human genome) was first formally proposed at a meeting sponsored by the U.S. Department of Energy in 1985. The scientific community had a mixed reaction to the proposal. Some investigators argued that sequence information, which could only be obtained through a large-scale effort, was essential for medical diagnosis and treatment. Others argued that the $3 billion needed for the project would be better spent if used to support investigator-initiated projects designed to address specific medical problems. Investigators were also concerned that DNA techniques, sequencing methods, data handling, and data storage then available were not up to the task of sequencing the human genome. Those who supported the project thought that technological advances would produce the necessary tools and that the project should support efforts to make such advances.

The Human Genome Project was officially launched in the United States in 1990 as a $3 billion, 15-year effort sponsored by the National Institutes of Health and Department of Energy. Participating

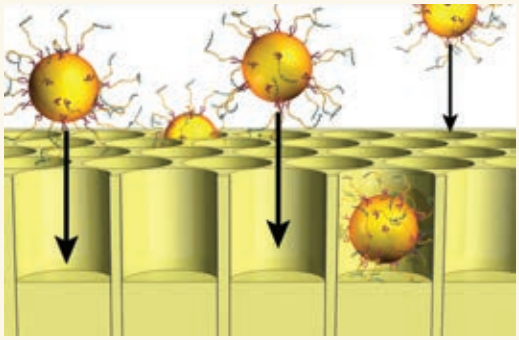
BOX 4.3: IN THE LAB

New Generation DNA Sequencers

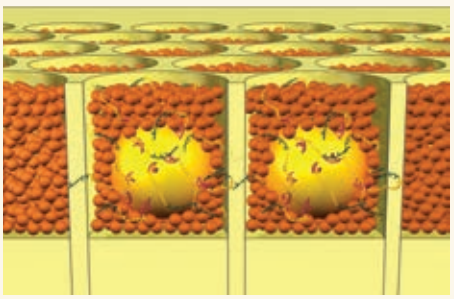
A new generation of DNA sequencers permits rapid and accurate DNA sequence information to be obtained without the need for electrophoresis or *in vivo* cloning. The basic ideas behind these massively parallel sequencing instruments are as follows:

1. Recombinant DNA technology is used to prepare thousands of tiny Sepharose® beads, each with millions of copies of a single-stranded DNA fragment attached to it. The single-stranded DNA fragments attached to any given Sepharose® bead are identical to one another but differ from the fragments attached to other Sepharose® beads. The Sepharose® bead population is large enough to ensure that it contains all the genomic information within an organism. Although the DNA fragments attached to different Sepharose® beads have different sequences, all fragments are constructed to have the same short segment at their 3′-end.
2. The single strands attached to the Sepharose® beads are annealed to a short primer complementary to their 3′-end. Then a slurry containing the Sepharose® beads, DNA polymerase, and enzyme beads with bound ATP sulfurylase and luciferase (see below) is loaded onto a 7 cm × 7.5 cm fiber-optic slide that has 1.6 million wells (**FIGURE B4.3a**). Each well is so small (44 μm diameter) that only one Sepharose® bead (33 μm diameter) can fit inside. Enzyme beads are quite small and therefore many such beads can fit around the Sepharose® bead with DNA segments. After loading, the fiber-optic slide is placed in the GS FLX sequencer. Nucleotide solutions flow across the fiber-optic slide sequentially in a fixed order (**FIGURE B4.3b**).

(a) Loading capture beads onto the fiber-optic slide



Load emzyme beads (ATP sulfurylase and luciferase)



(b) Schematic showing the location of the fiber-optic slide in the GS FLX instrument

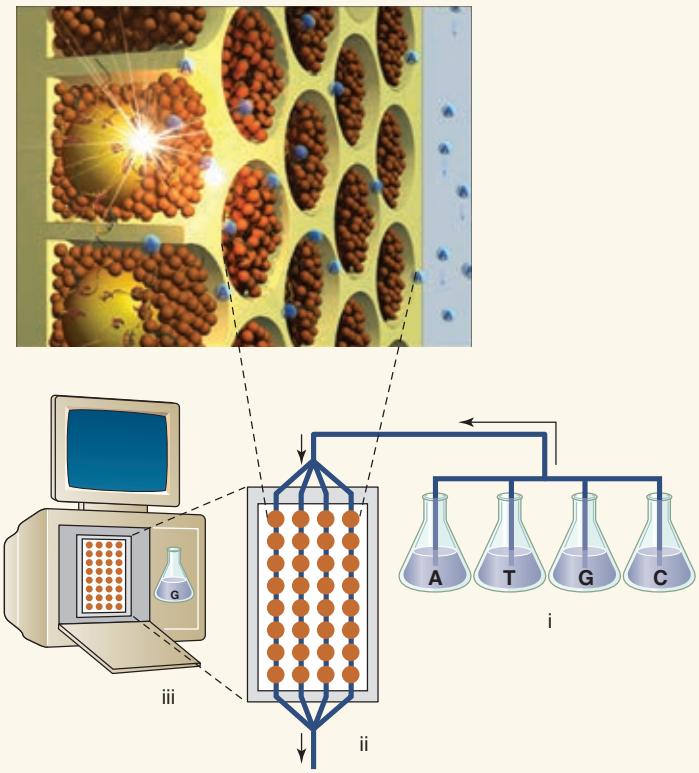


FIGURE B4.3 The fiber-optic slide. (a) Loading beads onto wells. (b) Schematic showing the location of the fiber-optic slide in the GS FLX instrument. The GS FLX instrument has three major subsystems: (i) a flow cell that includes the fiber-optic slide with about one million wells, (ii) a camera to image the fiber-optic slide, and (iii) a computer containing proprietary software needed to collect and analyze the data and assemble the sequence. (Parts a and b [top panel] courtesy of 454 Sequencing. © 2010 Roche Diagnostics. Part b [bottom panel] adapted from 454 Life Sciences, How Is Genome Sequencing Done? Roche Diagnostics, 2010.)

(continues)

BOX 4.3: IN THE LAB

New Generation DNA Sequencers (*Continued*)

The sequencing process is best understood by focusing on chemical events that take place in a single well. When the nucleotide that flows into a well is complementary to the next base in the template strand, the nucleotide is added to the primer strand and pyrophosphate is released. If the nucleotide is not complementary, then no reaction takes place. Because the order in which nucleotides flow into a cell is known, sequencing can be performed by correlating pyrophosphate release with this known order of nucleotides that flow into the well.

The problem of finding a quick and accurate way to monitor pyrophosphate release was solved by using three coupled enzyme reactions to monitor pyrophosphate (**FIGURE B4.3c**). The three reactions are as follows: (1) DNA polymerase releases inorganic

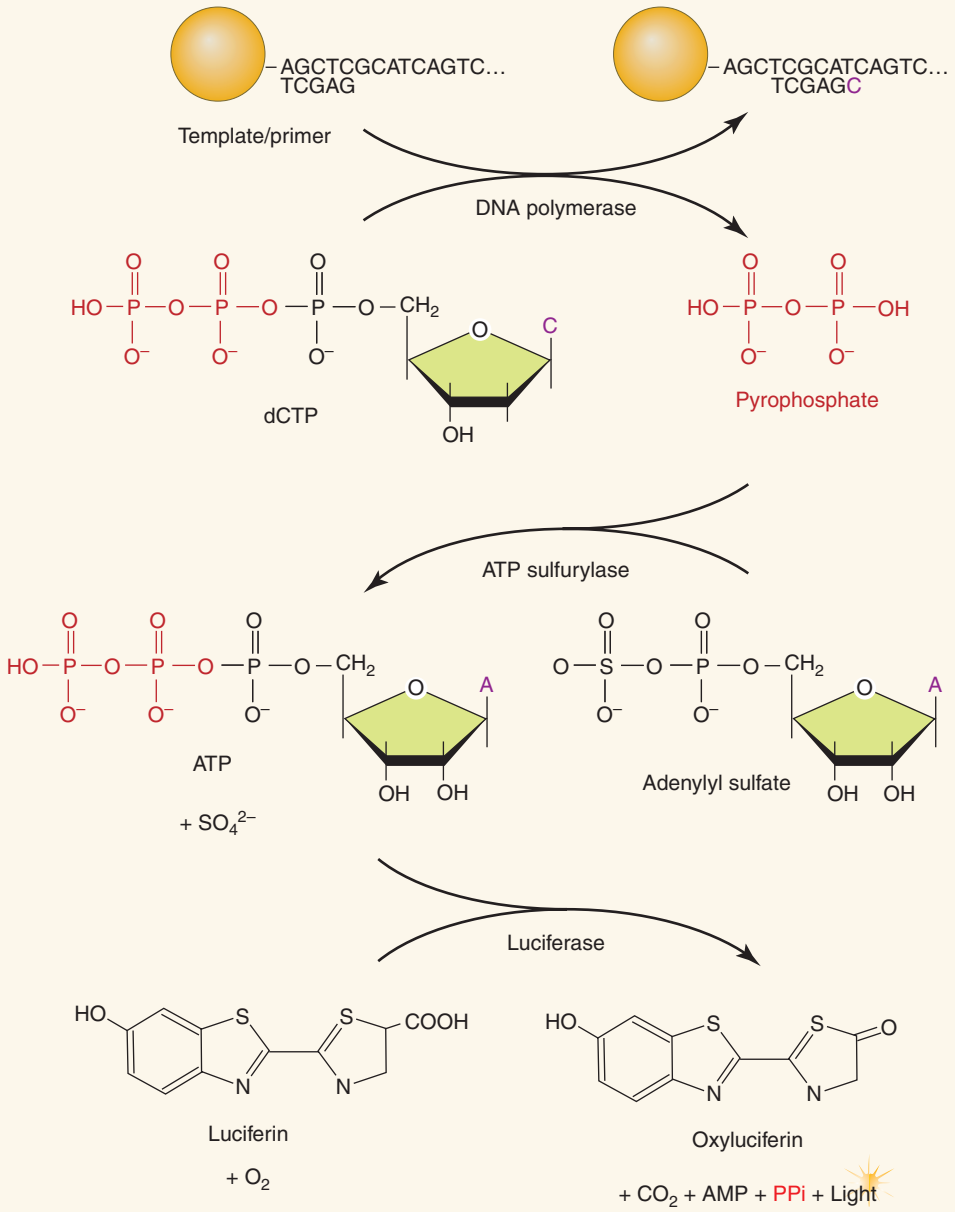


FIGURE B4.3c **Pyrosequencing reactions.**

pyrophosphate as it adds each deoxynucleotide to the primer, (2) ATP sulfurylase (ATP: sulfate adenylyltransferase) catalyzes adenylyl group transfer from adenylyl sulfate to the released pyrophosphate to form ATP, and (3) the firefly enzyme luciferase catalyzes an ATP-dependent chemiluminescent reaction in which luciferin conversion to oxyluciferin is accompanied by light emission.

Although the enzyme assay is very fast, reproducible, and quite sensitive, it does have one drawback. dATP is a weak substrate for luciferase, causing background light emission. This difficulty is circumvented by replacing dATP with an analog that is a substrate for DNA polymerase but not for luciferase. The analog, dATPαS [deoxyadenosine 5′-(*O*-1-thiotriphosphate)], differs from dATP in only one atom; a nonbonding oxygen atom in the α-phosphate group is replaced by a sulfur atom (**FIGURE B4.3d**).

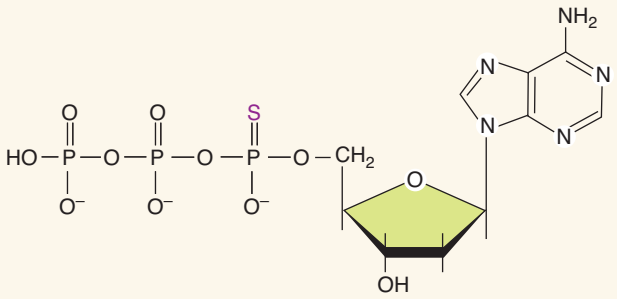


FIGURE B4.3d dATPαS [deoxyadenosine 5′-(*O*-1-thiotriphosphate)].

The light signal generated in a well each time a nucleotide adds to the primer chain is recorded by a camera. This sequencing technique, called pyrosequencing, has a single read accuracy that is greater than 99.5%. When errors do occur, they usually result from the presence of a homopolymer region in the template DNA. In a limited range (up to about 10 repeated nucleotides), the light signal strength is proportional to the number of nucleotides. The camera can record light emission events as they occur in each of the 1.6 million wells on a single plate and transmit the information to a computer for analysis.

The first application of massively parallel DNA sequencing to human genome sequencing was completed in 2008. Dr. James D. Watson, the co-discoverer of the double helical structure of DNA, provided a blood sample to a small team of investigators, who then determined his DNA sequence in less than 4 months at a cost of $1.5 million. New generation sequencing still has a way to go before it can be used in clinical laboratories. The race is now on to find a technique that will determine an individual's DNA sequence for under $1,000 because DNA sequencing then might become a practical method for diagnosing disease and providing a rational basis for treatment. So-called new-new generation techniques offer considerable promise for reaching the $1,000 goal.

In 2008 Stephan C. Schuster and coworkers sequenced nuclear DNA extracted from hair shafts collected from permafrost remains of the extinct woolly mammoth (*Mammuthus primigenius*). They used hair because it allowed them to remove bacteria and other contaminants without damaging the keratin-encased endogenous DNA. Based on C-14 dating the hair sample was calculated to be about 18,500 years old. Although fragmented and damaged, the ancient DNA could be sequenced using the GS FLX sequencer. Schuster and coworkers determined the sequence for 3.3 billion bases out of an estimated total of about 4 billion bases. Based on their data they estimated that the extinct wooly mammoth and the African elephant are 99.78% identical at the amino acid level. The estimated divergence rate between mammoth and African elephant is half of that between human and chimpanzee.

The DNA of a second extinct species was sequenced at about the same time. Svante Pääbo and coworkers used the GS FLX sequencer to determine the complete sequence of mitochondrial DNA from a bone sample of a Neanderthal individual who lived about 38,000 years ago. Based on the sequence information it appears separation between Neanderthal and present-day human populations took place between 270,000 and 440,000 years ago.

Clearly, DNA sequencing's impact extends well beyond molecular biology. It is providing important new knowledge in all fields of biology and promises to continue doing so at an accelerating rate in the future.

laboratories in the United States, designated genome sequencing centers, were soon joined by sequencing centers in China, France, Germany, Japan, and the United Kingdom. The Human Genome Project completed initial sequencing in 2001. Further efforts closed gaps and improved accuracy. Approximately 2.85 billion of the 3 billion nucleotides in the human genome have been sequenced with an error rate that is estimated to be less than 1 bp in every 10,000 bp. The sequenced DNA includes virtually all active or potentially active genes. Work continues to fill in the remaining gaps in the human genome.

The human genome sequence has provided a great deal of important information. First, the human genome contains about 20,000 to 25,000 protein coding genes, a number that is much smaller than anticipated. Second, DNA that codes for the protein-coding genes accounts for only about 2% of the human genome. Third, gene density varies among chromosomes, with some chromosome being more gene-rich than others. Fourth, several thousand non–protein-coding genes code for RNA molecules. Fifth, gene-dense DNA regions within a chromosome usually have a high G-C content. Sixth, about half the protein-coding genes specify proteins of unknown function. Finally, the human genome sequence is almost identical (99.9%) in all people.

The finished human genome sequence serves as an invaluable reference that is now being used to learn how our genes work, influence metabolic processes, and are linked to diseases. Although the human genome sequence provides considerable valuable information, it is in some ways like a road map that does not show the names of the roads or the towns and cities they pass through. The human genome requires **annotation**. That is, protein-coding genes and their regulatory signals need to be identified, introns and exons indicated, and regions that code for regulatory RNA molecules specified. The annotated human genome has the potential to provide molecular biologists with a factual base that is every bit as important as the periodic table is for chemists.

Questions and Problems

1. Briefly define, describe, or illustrate each of the following.
   a. SDS-Page
   b. Nuclease
   c. Endonuclease
   d. Exonuclease
   e. S1 endonuclease
   f. Restriction endonuclease
   g. Isoschizomers
   h. Blunt ends
   i. Cohesive ends
   j. Restriction map
   k. Recombinant plasmid
   l. Competent bacteria
   m. Southern transfer technique
   n. Northern blotting
   o. Western blotting
   p. DNA polymerase I
   q. Chain termination method
   r. Klenow fragment
   s. Nick translation
   u. Site-directed mutagenesis
   v. Shotgun sequencing
   w. Whole genome shotgun sequencing
   x. Pyrosequencing
2. Why does supercoiled plasmid DNA sediment more rapidly during sucrose gradient centrifugation than a linear duplex with the same number of base pairs?
3. How can you separate linear double-stranded DNA that is uniformly labeled with N-14 from linear double-stranded DNA with the same sequence that is uniformly labeled with N-15?

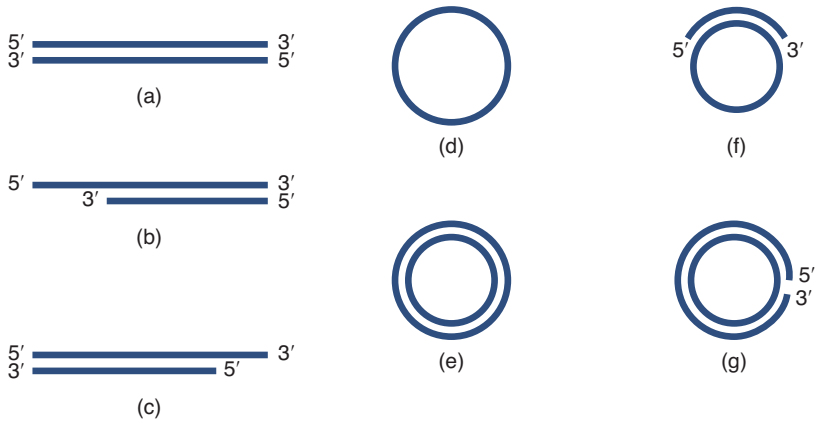
4. An enzyme has been isolated that migrates as a single band during gel electrophoresis. The same enzyme migrates as two bands of unequal size during SDS-PAGE. How do you explain these results? What other information about the protein can you obtain from SDS-PAGE?
5. How do endonucleases differ from exonucleases?
6. A DNA fragment contains the sequence

   5′-AAGATCGGATCCGAATTCTTTGCA-3′
   3′-TTCTAGCCTAGGCTTAAGAAACGT-5′

   Based on the information presented in Table 4.2, predict the sequence of the fragments that would be produced if the original fragment were incubated with
   a. BamHI
   b. BglII
   c. EcoRI
   d. EcoRV
   e. MboI
7. A 12.0-kb linear double-stranded DNA molecule is cut with BamHI or BglII and then with both enzymes. The results are as follows.

Restriction endonuclease	Restriction fragments (kb)
BamHI	7.5 and 4.5
BglII	6.0, 2.4, and 3.6
BamHI and BglII	6.0, 3.6, 1.5, and 0.9

   Construct a restriction map of the 12.0-kb DNA molecule from the data in the table.
8. EcoRV makes a single cut in each strand of a DNA molecule but produces only one fragment. How is this possible?
9. An investigator isolates a plasmid from *E. coli*. When the plasmid is analyzed by agarose gel electrophoresis, four bands are observed. The plasmid is reported to have a single BamHI restriction site. When the plasmid is incubated with BamHI and analyzed by agarose gel electrophoresis a single band is observed. Explain these results.
10. Describe an approach that you might use to determine the sequence of a cloned 15-kbp DNA fragment.
11. Describe a method you could use to label DNA with 5′-[^{32}P]phosphate at each end. How can you extend this procedure to isolate two fragments that are each labeled with 5′-[^{32}P]phosphate at one end?
12. Based on the fact that DNA polymerase I requires both a template and primer, predict how DNA polymerase I would act on each of the following DNA molecules:



13. A student has isolated plasmids from *E. coli* and would like to make them radioactive. What enzyme(s) and radioactive precursor(s) will be required to label the plasmids? Draw a diagram to illustrate the process.
14. What is a dideoxynucleotide and how is it used to sequence DNA?

15. The gel below was generated when a DNA strand was sequenced using the chain termination method.

The fluorescent dideoxynucleoside triphosphates used had the following colors: ddATP, green; ddTTP, red; ddGTP, blue; and ddCTP, magenta. Determine the nucleotide sequence of the DNA nucleotide chain synthesized from the primer and indicate the 5′and 3′ ends.

16. The sequence of the entire *E. coli* chromosome is known. How can you use this information and a thermostable DNA polymerase to prepare a DNA fragment that contains a specific 1-kbp chromosomal sequence?
17. You suspect an *E. coli* mutant is missing a DNA segment normally located between two BamHI restriction sites. Describe how the Southern blot procedure could be used to test your hypothesis.
18. Why are fluorescent dideoxynucleoside triphosphates required for chain termination sequencing but not for pyrosequencing?
19. A single-stranded circular DNA has been isolated that contains a gene you would like to alter. How can you change a single nucleotide in the gene?
20. A plasmid has a single BamHI site and a penicillin-resistance gene. The BamHI site is not in the penicillin-resistance gene or a site required for plasmid replication. *E. coli* gene X is located between two BamHI restriction sites, but there are no BamHI restriction sites in the gene. Illustrate how you can use this information to insert gene X into the plasmid and clone gene X.

Suggested Reading

Nucleic Acid Isolation

Bowien, B., and Dürre, P. (eds). 2004. *Nucleic Acids Isolation Methods*. Stevenson Ranch, CA: American Scientific.

Chomczynski, P. 2004. Single-step method of total RNA isolation by guanidine-phenol extraction. *Encyclopedia of Life Sciences*. pp. 1–6. Hoboken, NJ: John Wiley & Sons.

Chomczynski, P., and Mackey, K. 2001. RNA: Methods for preparation. *Encyclopedia of Life Sciences*. pp. 1–2. Hoboken, NJ: John Wiley & Sons.

Rapley, R. (ed). 2000. *The Nucleic Acids Protocol Book*. Totowa, NJ: Humana Press.

Roe, B. A., Crabtree, J. S., and Khan A., 1996. *DNA Isolation and Sequencing*. Hoboken, NJ: John Wiley & Sons.

Physical Techniques Used to Study Macromolecules

Birnie, G. D., and Rickwood, D. (eds). 1978. *Centrifugal Separations in Molecular and Cell Biology*. Tolly, UK: Butterworths.

Bozzola, J. J., and Russell, L. D. 1998. *Electron Microscopy: Principles and Techniques for Biologists* (2nd ed). Sudbury MA: Jones & Bartlett.

Dubochet, J., and Stahlberg, H. 2001. Electron cryomicroscopy. *Encyclopedia of Life Sciences*. pp. 1–5. Hoboken, NJ: John Wiley & Sons.

Graham, J. M., and Rickwood, D. 2001. *Biological Centrifugation*. London, UK: BIOS Scientific Publishers.

Hayat, M. A. 2002. *Principles and Techniques of Electron Microscopy: Biological Applications* (4th ed). Cambridge, UK: Cambridge University Press.

Jones, P. 1996. *Gel Electrophoresis: Nucleic Acids: Essential Techniques*. New York: John Wiley & Sons.

Ludtke, S. J., and Chiu, W. 2001. Electron cryomicroscopy and three-dimensional computer reconstruction of biological molecules. *Encyclopedia of Life Sciences*. pp. 1–5. Hoboken, NJ: John Wiley & Sons.

Price C. A. (ed.) 1982. *Centrifugation in Density Gradients*. San Diego, CA: Academic Press.

Righetti, P. G. 2005. Capillary electrophoresis. *Encyclopedia of Life Sciences*. pp. 1–4. Hoboken, NJ: John Wiley & Sons.

Southern, E. M. 2002. Pulsed-field gel electrophoresis of DNA. *Encyclopedia of Life Sciences*. pp. 1–7. Hoboken, NJ: John Wiley & Sons.

Tao, Y., and Zhang, W. 2000. Recent developments in cryo-electron microscopy reconstruction of single particles. *Curr Opin Struct Biol* 10:616–622.

Unger, V. M. 2001. Electron cryomicroscopy methods. *Curr Opin Struct Biol* 11: 548–554.

Völkl, A. 2002. Ultracentrifugation. *Encyclopedia of Life Sciences*. pp. 1–7. London: Nature.

Westermeier, R. 2005. *Electrophoresis in Practice: A Guide to Methods and Applications of DNA and Protein Separations* (4th ed). New York: John Wiley & Sons.

Enzymatic Techniques Used to Study DNA

Aggarwal, A. K. 1995. Structure and function of restriction endonucleases. *Curr Opin Struct Biol* 5:11–19.

Ansorge, W. J. 2009. Next generation sequencing techniques. *New Biotechnol* 25:195–203.

Bentley, D. R., et al. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456:53–59.

Bickle, T. A. 1993. The ATP-dependent restriction enzymes. Linn S. M., Lloyd, R. S., and Roberts, R. J. (eds). *Nucleases*. pp. 89–109. Woodbury, NY: Cold Spring Harbor Laboratory Press.

Bickle, T. A., and Krüger, D. H. 1993. Biology of DNA restriction. *Microbiol Rev* 57:434–450.

Bowen, D. J. 2005. *In vitro mutagenesis. Encyclopedia of Life Sciences*. pp. 1–6. Hoboken, NJ: John Wiley & Sons.

Chandrasegaran, S. 2001. Restriction enzymes. *Encyclopedia of Life Sciences*. pp. 1–7. London: Nature.

Droege, M., and Hill, B. 2008. The Genome Sequencer FLX™ System—Longer reads, more applications, straight forward bioinformatics and more complete data sets. *J Biotechnol* 136:3–10.

Dunham, I. 2005. Genome sequencing. *Encyclopedia of Life Sciences*. pp. 1–6. Hoboken, NJ: John Wiley & Sons.

Edelheit, O., Hanukoglu, A., and Hanukoglu, A. 2009. Simple and efficient site-directed mutagenesis using two single-primer reactions in parallel to generate mutants for protein structure-function studies. *BMC Biotechnol* 9:61.

França, L. T. C., Carrilho, E., and Kist, T. B. L. 2002. A review of DNA sequencing techniques. *Q Rev Biophys* 35:169–200.

Friedberg, E. C. 2006. The eureka enzyme: the discovery of DNA polymerase. *Nat Rev Mol Cell Biol* 7:143–147.

Green, E. D. 2001. Strategies for the systematic sequencing of complex genomes. *Nat Rev Genet* 2:573–583.

Hardin, S. H. 2001. DNA sequencing. *Encyclopedia of Life Sciences*. pp. 1–6. London: Nature.

Hubbard, T. J. P. 2005. Human genome: draft sequence. *Encyclopedia of Life Sciences*. pp. 1–5. Hoboken, NJ: John Wiley & Sons.

Hutchison III, C. A. 2007. DNA sequencing: bench to bedside and beyond. *Nucleic Acids Res* 35:6227–6237.

Joyce, C. M. 2004. DNA polymerase I, bacterial. *Encycl Biol Chem* 1:720–725.

Kehrer-Sawatzki, H. 2008. Sequencing the human genome: Novel insights into its structure and function. *Encyclopedia of Life Sciences*. pp. 1–9. Hoboken, NJ: John Wiley & Sons.

Lander, E. S., et al. 2001. Initial sequencing and analysis of the human genome. *Nature* 409:860–921.

Mani, M., Kandavelou, K., and Chandrasegaran, S. 2007. Restriction enzymes. *Encyclopedia of Life Sciences*. pp. 1–8. Hoboken, NJ: John Wiley & Sons.

Mardis, E. R. 2008. Next-generation DNA sequencing methods. *Annu Rev Genom Hum Genet* 9:387–402.

Margulies, M., et al. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437:376–380.

Meyers, B. C., Scalabrin, S., and Morgante, M. 2004. Mapping and sequencing complex genomes: Let's get physical! *Nat Rev Genet* 5:578–588.

Miller, W., et al. 2008. Sequencing the nuclear genome of the extinct woolly mammoth. *Nature* 456:387–390.

Mukhopadhyay, R. 2009. DNA sequencers: the next generation. *Anal Chem* 81:1736–1740.

Mullis, K. B. 1990. The unusual origin of the polymerase chain reaction. *Sci Am* 262:56–65.

Nathans, D. 1992. Restriction endonucleases, simian virus 40, and the new genetics. Linsted, J. (ed). *Nobel Lectures, Chemistry 1971–1980*. pp. 498–517. Hackensack, NJ: World Scientific Publishing.

Nyrén, P. 2007. The history of pyrosequencing®. *Methods Mol Biol* 373:1–13.

Pettersson, E., Lundeberg, J., and Ahmadian, A. 2009. Generations of sequencing technologies. *Genomics* 93:105–111.

Roberts R. J., and Halford, S. E. 1993. Type II restriction enzymes. Linn S. M., Lloyd, R. S., and Roberts, R. J. (eds). *Nucleases*. pp. 35–88. Woodbury, NY: Cold Spring Harbor Laboratory Press.

Rothberg, J. M., and Leamon, J. H. 2008. The development and impact of 454 sequencing. *Nat Biotechnol* 26:1117–1124.

Sanger, F. 1992. Determination of nucleotide sequences in DNA. Linsted, J. (ed). *Nobel Lectures, Chemistry* 1971–1980. pp. 431–447. Hackensack, NJ: World Scientific Publishing.

Sanger, F., Nicklen, S., and Coulson, A. R. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 74:5463–5467.

Shendure, J., and Ji, H. 2008. Next-generation DNA sequencing. *Nat Biotechnol* 26:1135–1145.

Smith, H. O. 1992. Nucleotide sequence specificity of restriction endonucleases. Linsted, J. (ed). *Nobel Lectures, Chemistry 1971–1980*. pp. 523–541. Hackensack, NJ: World Scientific Publishing.

Smith, M. 1997. Synthetic DNA and biology. In Malstrom, B. G. (ed). *Nobel Lectures, Physiology or Medicine* 1991–1995. Hackensack, NJ: World Scientific Publishing.

Tűmmler, B., and Mekus, F. 2005. Restriction mapping. *Encyclopedia of Life Sciences*. pp. 1–3. Hoboken, NJ: John Wiley & Sons.

Weber, J. L., and Myers, E. W. 1997. Human whole-genome shotgun sequencing. *Genome Res* 7:401–409.

Wheeler, D. A., et al. 2008. The complete genome of an individual by massively parallel DNA sequencing. *Nature* 452:872–876.

Wilton, S. 2002. Dideoxy sequencing of DNA. *Encyclopedia of Life Sciences*. pp. 1–16. London: Nature.