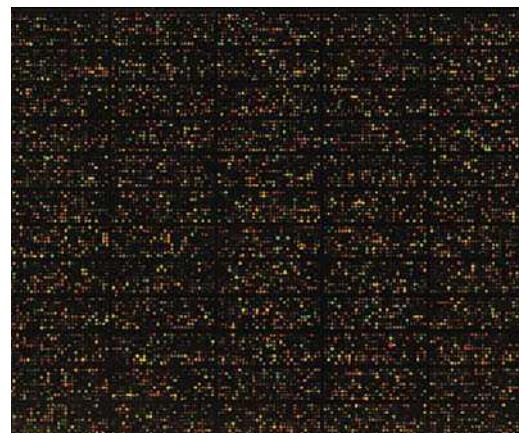


Exploring Genes and Genomes



Processes such as the development from a caterpillar into a butterfly entail dramatic changes in patterns of gene expression. The expression levels of thousands of genes can be monitored through the use of DNA arrays. At the right, a DNA microarray reveals the expression levels of more than 12,000 human genes; the brightness and color of each spot indicates a change in the expression level of the corresponding gene. [(Left) Cathy Keifer/istockphoto.com. (Right) Agilent Technologies.]

Since its emergence in the 1970s, recombinant DNA technology has revolutionized biochemistry. The genetic endowment of organisms can now be precisely changed in designed ways. Recombinant DNA technology is the fruit of several decades of basic research on DNA, RNA, and viruses. It depends, first, on having enzymes that can cut, join, and replicate DNA and those that can reverse transcribe RNA. Restriction enzymes cut very long DNA molecules into specific fragments that can be manipulated; DNA ligases join the fragments together. Many kinds of restriction enzymes are available. By applying this assortment cleverly, researchers can treat DNA sequences as modules that can be moved at will from one DNA molecule to another. Thus, recombinant DNA technology is based on the use of enzymes that act on nucleic acids as substrates.

A second foundation is the base-pairing language that allows complementary sequences to recognize and bind to each other. Hybridization with complementary DNA (cDNA) or RNA probes is a sensitive means of detecting specific nucleotide sequences. In recombinant DNA technology, base-pairing is used to construct new combinations of DNA as well as to detect and amplify particular sequences.

Third, powerful methods have been developed for determining the sequence of nucleotides in DNA. These methods have been harnessed to sequence complete genomes: first, small genomes from viruses; then, larger

OUTLINE

- 5.1 The Exploration of Genes Relies on Key Tools
- 5.2 Recombinant DNA Technology Has Revolutionized All Aspects of Biology
- 5.3 Complete Genomes Have Been Sequenced and Analyzed
- 5.4 Eukaryotic Genes Can Be Quantitated and Manipulated with Considerable Precision

genomes from bacteria; and, finally, eukaryotic genomes, including the 3-billion-base-pair human genome. Scientists are just beginning to exploit the enormous information content of these genome sequences.

Finally, recombinant DNA technology critically depends on our ability to deliver foreign DNA into host organisms. For example, DNA fragments can be inserted into plasmids, where they can be replicated within a short period of time in their bacterial hosts. In addition, viruses efficiently deliver their own DNA (or RNA) into hosts, subverting them either to replicate the viral genome and produce viral proteins or to incorporate viral DNA into the host genome.

These new methods have wide-ranging benefits across a broad spectrum of disciplines, including biotechnology, agriculture, and medicine. Among these benefits is the dramatic expansion of our understanding of human disease. Throughout this chapter, a specific disorder, amyotrophic lateral sclerosis (ALS), will be used to illustrate the effect that recombinant DNA technology has had on our knowledge of disease mechanisms. ALS was first described clinically in 1869 by the French neurologist Jean-Martin Charcot as a fatal neurodegenerative disease of progressive weakening and atrophy of voluntary muscles. ALS is commonly referred to as Lou Gehrig's Disease, for the baseball legend whose career and life were prematurely cut short as a result of this devastating disease. For many years, little progress had been made in the study of the mechanisms underlying ALS. As we shall see, significant advances have been made with the use of research tools facilitated by recombinant DNA technology.

5.1 The Exploration of Genes Relies on Key Tools

The rapid progress in biotechnology—indeed its very existence—is a result of a few key techniques.

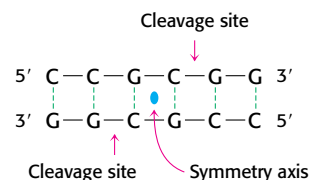
1. *Restriction-enzyme analysis.* Restriction enzymes are precise molecular scalpels that allow an investigator to manipulate DNA segments.
2. *Blotting techniques.* Southern and northern blots are used to separate and identify DNA and RNA sequences, respectively. The western blot, which uses antibodies to characterize proteins, was described in Chapter 3.
3. *DNA sequencing.* The precise nucleotide sequence of a molecule of DNA can be determined. Sequencing has yielded a wealth of information concerning gene architecture, the control of gene expression, and protein structure.
4. *Solid-phase synthesis of nucleic acids.* Precise sequences of nucleic acids can be synthesized de novo and used to identify or amplify other nucleic acids.
5. *The polymerase chain reaction (PCR).* The polymerase chain reaction leads to a billionfold amplification of a segment of DNA. One molecule of DNA can be amplified to quantities that permit characterization and manipulation. This powerful technique can be used to detect pathogens and genetic diseases, determine the source of a hair left at the scene of a crime, and resurrect genes from the fossils of extinct organisms.

A final set of techniques relies on the computer, without which, it would be impossible to catalog, access, and characterize the abundant information generated by the methods outlined above. Such uses of the computer will be presented in Chapter 6.

Restriction enzymes split DNA into specific fragments

Restriction enzymes, also called *restriction endonucleases*, recognize specific base sequences in double-helical DNA and cleave both strands of that duplex at specific places. To biochemists, these exquisitely precise scalpels are marvelous gifts of nature. They are indispensable for analyzing chromosome structure, sequencing very long DNA molecules, isolating genes, and creating new DNA molecules that can be cloned.

Restriction enzymes are found in a wide variety of prokaryotes. Their biological role is to cleave foreign DNA molecules, providing the host organism with a primitive immune system. Many restriction enzymes recognize specific sequences of four to eight base pairs and hydrolyze a phosphodiester bond in each strand in this region. A striking characteristic of these cleavage sites is that they almost always possess *twofold rotational symmetry*. In other words, the recognized sequence is *palindromic*, or an inverted repeat, and the cleavage sites are symmetrically positioned. For example, the sequence recognized by a restriction enzyme from *Streptomyces achromogenes* is



In each strand, the enzyme cleaves the C–G phosphodiester bond on the 3' side of the symmetry axis. As we shall see in Chapter 9, this symmetry corresponds to that of the structures of the restriction enzymes themselves.

Several hundred restriction enzymes have been purified and characterized. Their names consist of a three-letter abbreviation for the host organism (e.g., *Eco* for *Escherichia coli*, *Hin* for *Haemophilus influenzae*, *Hae* for *Haemophilus aegyptius*) followed by a strain designation (if needed) and a roman numeral (to distinguish multiple enzymes from the same strain). The specificities of several of these enzymes are shown in Figure 5.1.

Restriction enzymes are used to cleave DNA molecules into specific fragments that are more readily analyzed and manipulated than the entire parent molecule. For example, the 5.1-kb circular duplex DNA of the tumor-producing SV40 virus is cleaved at one site by *EcoRI*, at four sites by *HpaI*, and at 11 sites by *HindIII*. A piece of DNA, called a restriction fragment, produced by the action of one restriction enzyme can be specifically cleaved into smaller fragments by another restriction enzyme. The pattern of such fragments can serve as a *fingerprint* of a DNA molecule, as will be considered shortly. Indeed, complex chromosomes containing hundreds of millions of base pairs can be mapped by using a series of restriction enzymes.

Restriction fragments can be separated by gel electrophoresis and visualized

In Chapter 3, we considered the use of gel electrophoresis to separate protein molecules (Section 3.1). Because the phosphodiester backbone of DNA is highly negatively charged, this technique is also suitable for the separation of nucleic acid fragments. Among the many applications of DNA electrophoresis are the detection of mutations that affect restriction fragment size (such as insertions and deletions) and the isolation, purification, and quantitation of a specific DNA fragment.

For most gels, the shorter the DNA fragment, the farther the migration. Polyacrylamide gels are used to separate, by size, fragments containing as

Palindrome

A word, sentence, or verse that reads the same from right to left as it does from left to right.

Radar

Senile felines

Do geese see God?

Roma tibi subito motibus ibit amor

Derived from the Greek *palindromos*, "running back again."

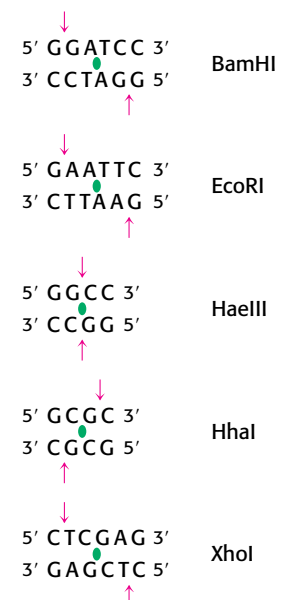


FIGURE 5.1 Specificities of some restriction endonucleases. The sequences that are recognized by these enzymes contain a twofold axis of symmetry. The two strands in these regions are related by a 180-degree rotation about the axis marked by the green symbol. The cleavage sites are denoted by red arrows. The abbreviated name of each restriction enzyme is given at the right of the sequence that it recognizes. Note that the cuts may be staggered or even.

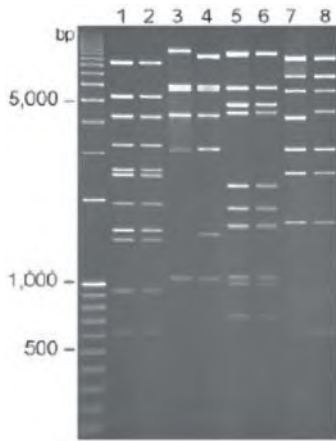


FIGURE 5.2 Gel-electrophoresis pattern of a restriction digest. This gel shows the fragments produced by cleaving DNA from two viral strains (odd- vs. even-numbered lanes) with each of four restriction enzymes. These fragments were made fluorescent by staining the gel with ethidium bromide. [Data from Carr et al., *Emerging Infectious Diseases*, www.cdc.gov/eid, 17(8), August 2011.]

many as 1000 base pairs, whereas more-porous agarose gels are used to resolve mixtures of larger fragments (as large as 20 kb). An important feature of these gels is their high resolving power. In certain kinds of gels, fragments differing in length by just one nucleotide out of several hundred can be distinguished. Bands or spots of radioactive DNA in gels can be visualized by autoradiography. Alternatively, a gel can be stained with a dye such as ethidium bromide, which fluoresces an intense orange under irradiation with ultraviolet light when bound to a double-helical DNA molecule (Figure 5.2). A band containing only 10 ng of DNA can be readily seen.

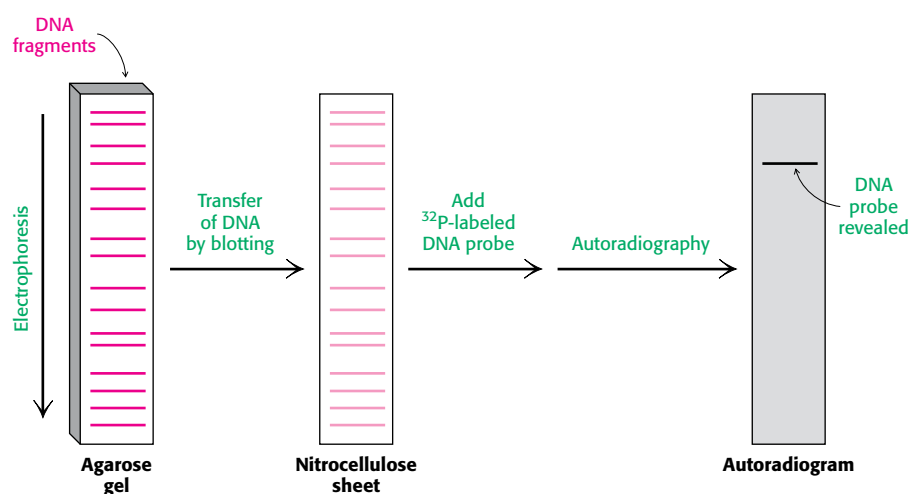
It is often necessary to determine if a particular base sequence is represented in a given DNA sample. For example, one may wish to confirm the presence of a specific mutation in genomic DNA isolated from patients known to be at risk for a particular disease. This specific sequence can be identified by hybridizing it with a labeled complementary DNA strand (Figure 5.3). A mixture of restriction fragments is separated by electrophoresis through an agarose gel, denatured to form single-stranded DNA, and transferred to a nitrocellulose sheet. The positions of the DNA fragments in the gel are preserved during the transfer. The nitrocellulose is then exposed to a ^{32}P -labeled or fluorescently tagged DNA probe, a short stretch of single-stranded DNA which contains a known base sequence. The probe hybridizes with a restriction fragment having a complementary sequence, and autoradiography or fluorescence imaging then reveals the position of the restriction-fragment-probe duplex. A particular fragment amid a million others can be readily identified in this way. This powerful technique is named *Southern blotting*, for its inventor Edwin Southern.

In a similar manner, RNA molecules of a specific sequence can also be readily identified. After separation by gel electrophoresis and transfer to nitrocellulose, specific sequences can be detected by DNA probes. This analogous technique for the analysis of RNA has been whimsically termed *northern blotting*. A further play on words accounts for the term *western blotting*, which refers to a technique for detecting a particular protein by staining with specific antibody (Section 3.3).

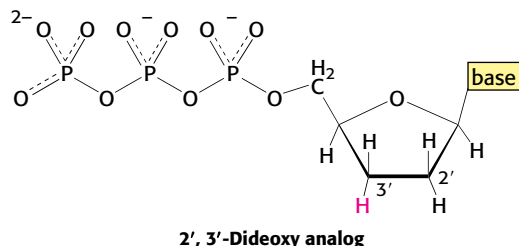
DNA can be sequenced by controlled termination of replication

The analysis of DNA structure and its role in gene expression have been markedly facilitated by the development of powerful techniques for the *sequencing* of DNA molecules. One of the first and most widely-used techniques for DNA sequencing is *controlled termination of replication*, also referred to as the Sanger dideoxy method for its pioneer, Frederick Sanger. The key to this approach is

FIGURE 5.3 Southern blotting. A DNA fragment containing a specific sequence can be identified by separating a mixture of fragments by electrophoresis, transferring them to nitrocellulose, and hybridizing with a ^{32}P -labeled probe complementary to the sequence. The fragment containing the sequence is then visualized by autoradiography.



the generation of DNA fragments whose length is determined by the last base in the sequence (Figure 5.4). In the current application of this method, a DNA polymerase is used to make the complement of a particular sequence within a single-stranded DNA molecule. The synthesis is primed by a chemically synthesized fragment that is complementary to a part of the sequence known from other studies. In addition to the four deoxyribonucleoside triphosphates, the reaction mixture contains a small amount of the 2',3'-*dideoxy analog* of each nucleotide, each carrying a different fluorescent label attached to the base (e.g., a green emitter for termination at A and a red one for termination at T).



The incorporation of this analog blocks further growth of the new chain because it lacks the 3'-hydroxyl terminus needed to form the next phosphodiester bond. The concentration of the dideoxy analog is low enough that chain termination will take place only occasionally. The polymerase will insert the correct nucleotide sometimes and the dideoxy analog other times, stopping the reaction. For instance, if the dideoxy analog of dATP is present, fragments of various lengths are produced, but all will be terminated by the dideoxy analog. Importantly, this dideoxy analog of dATP will be inserted only where a T was located in the DNA being sequenced. Thus, the fragments of different length will correspond to the positions of T.

The resulting fragments are separated by a technique known as *capillary electrophoresis*, in which the mixture is passed through a very narrow tube containing a gel matrix at high voltage to achieve efficient separation within a short time. As the DNA fragments emerge from the capillary, they are detected by their fluorescence; the sequence of their colors directly gives the base sequence. Sequences of as many as 1000 bases can be determined in this way. Indeed, automated Sanger sequencing machines can read more than 1 million bases per day.

DNA probes and genes can be synthesized by automated solid-phase methods

DNA strands, like polypeptides (Section 3.4), can be synthesized by the sequential addition of activated monomers to a growing chain that is linked to a solid support. The activated monomers are protected *deoxyribonucleoside 3'-phosphoramidites*. In step 1, the 3'-phosphorus atom of this incoming unit becomes joined to the 5'-oxygen atom of the growing chain to form a *phosphite triester* (Figure 5.5). The 5'-OH group of the activated monomer is unreactive because it is blocked by a dimethoxytrityl (DMT) protecting group, and the 3'-phosphoryl oxygen atom is rendered unreactive by attachment of the β -cyanoethyl (β CE) group. Likewise, amino groups on the purine and pyrimidine bases are blocked.

Coupling is carried out under anhydrous conditions because water reacts with phosphoramidites. In step 2, the phosphite triester (in which P is trivalent) is oxidized by iodine to form a *phosphotriester* (in which P is pentavalent). In step 3, the DMT protecting group on the 5'-OH group of the growing chain is removed by the addition of dichloroacetic acid, which

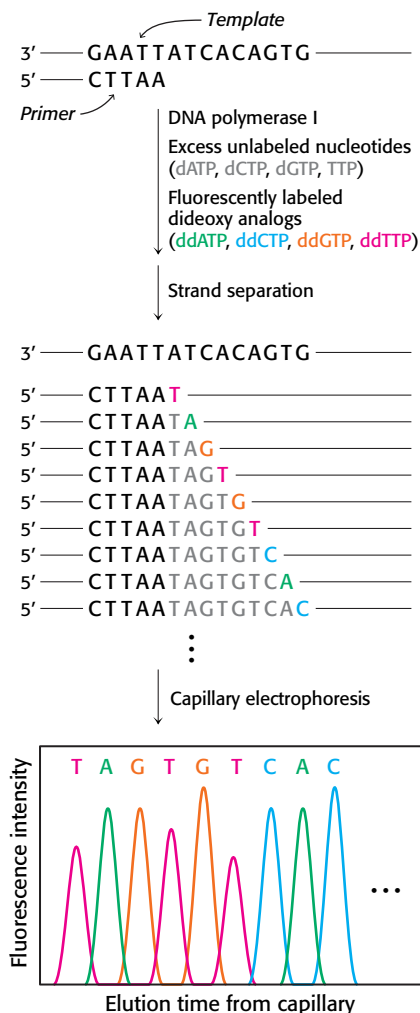


FIGURE 5.4 Fluorescence detection of oligonucleotide fragments produced by the dideoxy method. A sequencing reaction is performed with four chain-terminating dideoxy nucleotides, each labeled with a tag that fluoresces at a different wavelength. The color of each fragment indicates the identity of the last base in the chain. The fragments are separated by size using capillary electrophoresis and the fluorescence at each of the four wavelengths indicates the sequence of the complement of the original DNA template.

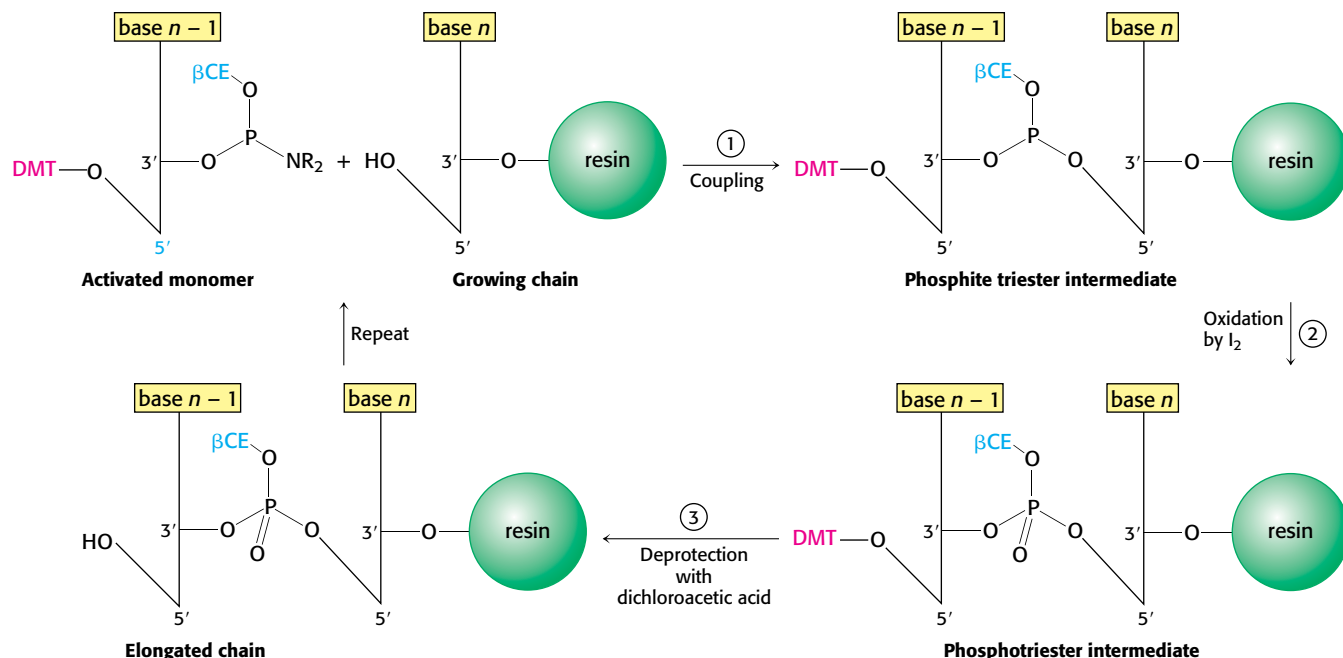
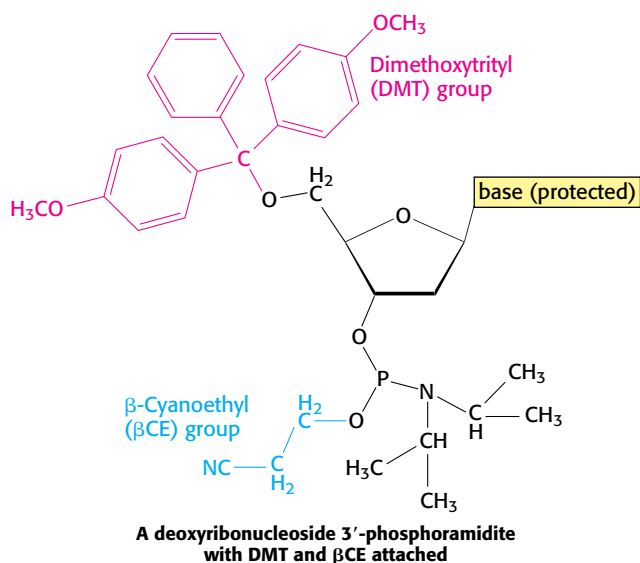


FIGURE 5.5 Solid-phase synthesis of a DNA chain by the phosphite triester method. The activated monomer added to the growing chain is a deoxyribonucleoside 3'-phosphoramidite containing a dimethoxytrityl (DMT) protecting group on its 5'-oxygen atom, a β -cyanoethyl (BCE) protecting group on its 3'-phosphoryl oxygen atom, and a protecting group on the base.



leaves other protecting groups intact. The DNA chain is now elongated by one unit and ready for another cycle of addition. Each cycle takes only about 10 minutes and usually elongates more than 99% of the chains.

This solid-phase approach is ideal for the synthesis of DNA, as it is for polypeptides, because the desired product stays on the insoluble support until the final release step. All the reactions take place in a single vessel, and excess soluble reagents can be added to drive reactions to completion. At the end of each step, soluble reagents and by-products are washed away from the resin that bears the growing chains. At the end of the synthesis, NH_3 is added to remove all protecting groups and release the oligonucleotide from the solid support. Because elongation is never 100% complete, the new DNA chains are of diverse lengths—the desired chain is the longest one. The sample can be purified by high-performance liquid chromatography or by electrophoresis on polyacrylamide gels. DNA chains of as many as 100 nucleotides can be readily synthesized by this automated method.

The ability to rapidly synthesize DNA chains of any selected sequence opens many experimental avenues. For example, a synthesized oligonucleotide labeled at one end with ^{32}P or a fluorescent tag can be used to search for a complementary sequence in a very long DNA molecule or even in a genome consisting of many chromosomes. The use of labeled oligonucleotides as DNA probes is powerful and general. For example, a DNA probe that can base-pair to a known complementary sequence in a chromosome can serve as the starting point of an exploration of adjacent uncharted DNA. Such a probe can be used as a *primer* to initiate the replication of neighboring DNA by DNA polymerase. An exciting application of the solid-phase approach is the synthesis of new tailor-made genes. New proteins with novel

properties can now be produced in abundance by the expression of synthetic genes. Finally, the synthetic scheme heretofore described can be slightly modified for the solid-phase synthesis of RNA oligonucleotides, which can be very powerful reagents for the degradation of specific mRNA molecules in living cells by a technique known as RNA interference (Section 5.4).

Selected DNA sequences can be greatly amplified by the polymerase chain reaction

In 1984, Kary Mullis devised an ingenious method called the *polymerase chain reaction* (PCR) for amplifying specific DNA sequences. Consider a DNA duplex consisting of a target sequence surrounded by nontarget DNA. Millions of copies of the target sequences can be readily obtained by PCR if the sequences flanking the target are known. PCR is carried out by adding the following components to a solution containing the target sequence: (1) a pair of primers that hybridize with the flanking sequences of the target, (2) all four deoxyribonucleoside triphosphates (dNTPs), and (3) a heat-stable DNA polymerase. A PCR cycle consists of three steps (Figure 5.6).

1. *Strand separation.* The two strands of the parent DNA molecule are separated by heating the solution to 95°C for 15 s.
2. *Hybridization of primers.* The solution is then abruptly cooled to 54°C to allow each primer to hybridize to a DNA strand. One primer hybridizes to the 3' end of the target on one strand, and the other primer hybridizes to the 3' end on the complementary target strand. Parent DNA duplexes do not form, because the primers are present in large excess. Primers are typically from 20 to 30 nucleotides long.
3. *DNA synthesis.* The solution is then heated to 72°C, the optimal temperature for heat-stable polymerases. One such enzyme is *Taq* DNA polymerase, which is derived from *Thermus aquaticus*, a thermophilic bacterium that lives in hot springs. The polymerase elongates both primers in the direction of the target sequence because DNA synthesis is in the 5'-to-3' direction. DNA synthesis takes place on both strands but extends beyond the target sequence.

These three steps—strand separation, hybridization of primers, and DNA synthesis—constitute one cycle of the PCR amplification and can be carried out repetitively just by changing the temperature of the reaction mixture. The thermostability of the polymerase makes it feasible to carry out PCR in a closed container; no reagents are added after the first cycle. At the completion of the second cycle, four duplexes containing the target sequence have been generated (Figure 5.7). Of the eight DNA strands comprising these duplexes, two short strands constitute only the target sequence—the sequence including and bounded by the primers. Subsequent cycles will amplify the target sequence exponentially. Ideally, after n cycles, the desired sequence is amplified 2^n -fold. The amplification is a millionfold after 20 cycles and a billionfold after 30 cycles, which can be carried out in less than an hour.

Several features of this remarkable method for amplifying DNA are noteworthy. First, the sequence of the target need not be known. All that is required is knowledge of the flanking sequences so that complementary primers can be synthesized. Second, the target can be much larger than the primers. Targets larger than 10 kb have been amplified by PCR. Third, primers do not have to be perfectly matched to flanking sequences to amplify targets. With the use of primers derived from a gene of known sequence, it is possible to search for variations on the theme. In this way, families of genes are being discovered by PCR. Fourth, PCR is highly specific because of the

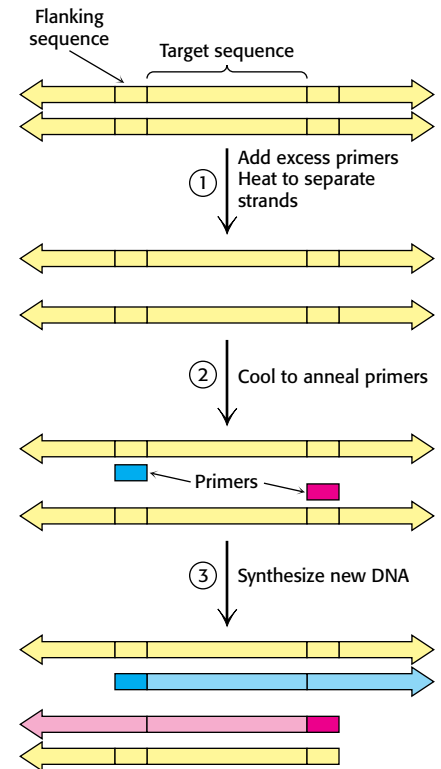
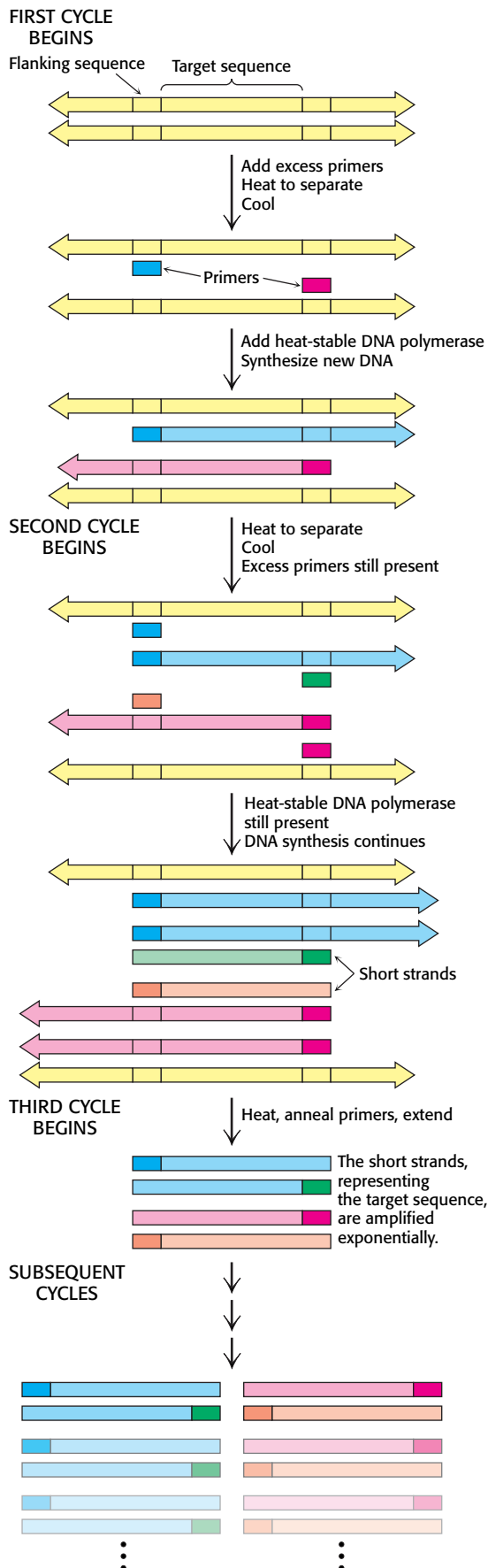


FIGURE 5.6 The first cycle in the polymerase chain reaction (PCR). A cycle consists of three steps: DNA double strand separation, the hybridization of primers, and the extension of primers by DNA synthesis.



stringency of hybridization at relatively high temperature. *Stringency* is the required closeness of the match between primer and target, which can be controlled by temperature and salt. At high temperatures, only the DNA between hybridized primers is amplified. A gene constituting less than a millionth of the total DNA of a higher organism is accessible by PCR. Fifth, PCR is exquisitely sensitive. A single DNA molecule can be amplified and detected.

PCR is a powerful technique in medical diagnostics, forensics, and studies of molecular evolution



PCR can provide valuable diagnostic information in medicine. Bacteria and viruses can be readily detected with the use of specific primers. For example, PCR can reveal the presence of small amounts of DNA from the human immunodeficiency virus (HIV) in persons who have not yet mounted an immune response to this pathogen. In these patients, assays designed to detect antibodies against the virus would yield a false negative test result. Finding *Mycobacterium tuberculosis* bacilli in tissue specimens is slow and laborious. With PCR, as few as 10 tubercle bacilli per million human cells can be readily detected. PCR is a promising method for the early detection of certain cancers. This technique can identify mutations of certain growth-control genes, such as the *ras* genes (Chapter 14). The capacity to greatly amplify selected regions of DNA can also be highly informative in monitoring cancer chemotherapy. Tests using PCR can detect when cancerous cells have been eliminated and treatment can be stopped; they can also detect a relapse and the need to immediately resume treatment. PCR is ideal for detecting leukemias caused by chromosomal rearrangements.

In addition, PCR has made an impact on forensics and legal medicine. An individual DNA profile is highly distinctive because many genetic loci are highly variable within a population. For example, variations at one specific location determine a person's HLA type (human leukocyte antigen type; Section 34.5); organ transplants are rejected when the HLA types of the donor and recipient are not sufficiently matched. PCR amplification of multiple genes is being used to establish biological parentage in disputed paternity and immigration cases. Analyses of blood stains and semen samples by PCR have implicated guilt or innocence in numerous assault and rape cases (Figure 5.8). The root of a single shed hair found at a crime scene contains enough DNA for typing by PCR.

DNA is a remarkably stable molecule, particularly when shielded from air, light, and water. Under such circumstances, large fragments of DNA can remain intact for thousands of years or longer. PCR provides an ideal method for amplifying such ancient DNA molecules so that they can be detected and characterized (Section 6.5). PCR can also be used to amplify DNA from microorganisms that have not yet been isolated and cultured. As will be discussed in Chapter 6, sequences from these PCR products

FIGURE 5.7 Multiple cycles of the polymerase chain reaction. The two short strands produced at the end of the third cycle (along with longer stands not shown) represent the target sequence. Subsequent cycles will amplify the target sequence exponentially and the parent sequence arithmetically.

can be sources of considerable insight into evolutionary relationships between organisms.

The tools for recombinant DNA technology have been used to identify disease-causing mutations

Let us consider how the techniques just described have been utilized in concert to study ALS, introduced at the beginning of this chapter. Five percent of all patients suffering from ALS have family members who also have been diagnosed with the disease. A heritable disease pattern is indicative of a strong genetic component of disease causation. To identify these disease-causing genetic alterations, researchers identify *polymorphisms* (instances of genetic variation) within an affected family that correlate with the emergence of disease. These polymorphisms may themselves cause disease or be closely linked to the responsible genetic alteration. *Restriction-fragment-length polymorphisms* (RFLPs) are polymorphisms within restriction sites that change the sizes of DNA fragments produced by the appropriate restriction enzyme. Using restriction digests and Southern blots of the DNA from members of ALS-affected families, researchers identified RFLPs that were found preferentially in those family members diagnosed with the disease. For some of these families, strong evidence was obtained for the disease-causing mutation within a specific region of chromosome 21.

After the probable location of one disease-causing gene had been identified, this same research group compared the locations of the ALS-associated RFLPs with the known sequence of chromosome 21. They noted that this chromosomal locus contains the *SOD1* gene, which encodes the Cu/Zn superoxide dismutase protein SOD1, an enzyme important for the protection of cells against oxidative damage (Section 18.3). PCR amplification of regions of the *SOD1* gene from the DNA of affected family members, followed by Sanger dideoxy sequencing of the targeted fragment, enabled the identification of 11 disease-causing mutations from 13 different families. This work was pivotal for focusing further inquiry into the roles that superoxide dismutase and its corresponding mutant forms play in the pathology of some forms of ALS.

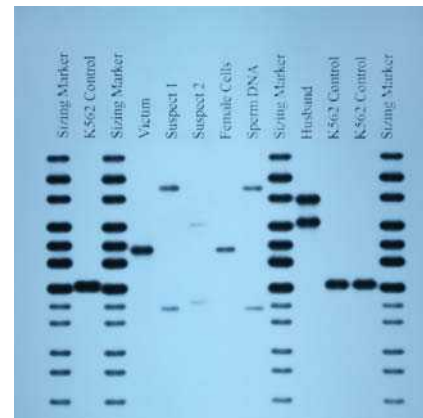


FIGURE 5.8 DNA and forensics. DNA isolated from sperm obtained during the examination of a rape victim was amplified by PCR, then compared with DNA from the victim and three potential suspects—the victim's husband and two additional individuals—using gel electrophoresis and autoradiography. Sperm DNA matched the pattern of Suspect 1, but not that of Suspect 2 or the victim's husband. Sizing marker and K562 lanes refer to control DNA samples. [Martin Shields/Science Source.]

5.2 Recombinant DNA Technology Has Revolutionized All Aspects of Biology

The development of recombinant DNA technology has taken biology from an exclusively analytical science to a synthetic one. New combinations of unrelated genes can be constructed in the laboratory by applying recombinant DNA techniques. These novel combinations can be cloned—amplified many-fold—by introducing them into suitable cells, where they are replicated by the DNA-synthesizing machinery of the host. The inserted genes are often transcribed and translated in their new setting. What is most striking is that the genetic endowment of the host can be permanently altered in a designed way.

Restriction enzymes and DNA ligase are key tools in forming recombinant DNA molecules

Let us begin by seeing how novel DNA molecules can be constructed in the laboratory. An essential tool for the manipulation of recombinant DNA is a *vector*, a DNA molecule that can replicate autonomously in an appropriate host organism. Vectors are designed to enable the rapid,

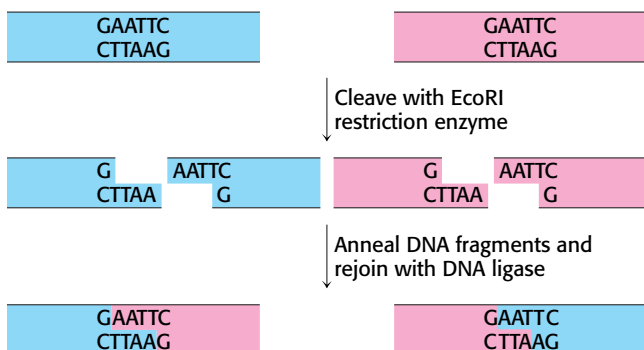


FIGURE 5.9 Joining of DNA molecules by the cohesive-end method. Two DNA molecules, cleaved with a common restriction enzyme such as EcoRI, can be ligated to form recombinant molecules.

covalent insertion of DNA fragments of interest. *Plasmids* (naturally occurring circles of DNA that act as accessory chromosomes in bacteria) and bacteriophage lambda (λ phage), a virus, are choice vectors for cloning in *E. coli*. The vector can be prepared for accepting a new DNA fragment by cleaving it at a single specific site with a restriction enzyme. For example, the plasmid pSC101, a 9.9-kb double-helical circular DNA molecule, is split at a unique site by the EcoRI restriction enzyme. The staggered cuts made by this enzyme produce *complementary single-stranded ends*, which have specific affinity for each other and hence are known as *cohesive* or *sticky ends*. Any DNA fragment can be inserted into this plasmid if it has the same cohesive ends. Such a fragment can be extracted from a larger piece of DNA by using the same restriction enzyme as was used to open the plasmid DNA (Figure 5.9).

The single-stranded ends of the fragment are then complementary to those of the cut plasmid. The DNA fragment and the cut plasmid can be annealed and then joined by *DNA ligase*, which catalyzes the formation of a phosphodiester bond at a break in a DNA chain. DNA ligase requires a free 3'-hydroxyl group and a 5'-phosphoryl group. Furthermore, the chains joined by ligase must be in a double helix. An energy source such as ATP or NAD^+ is required for the joining reaction, as will be discussed in Chapter 28.

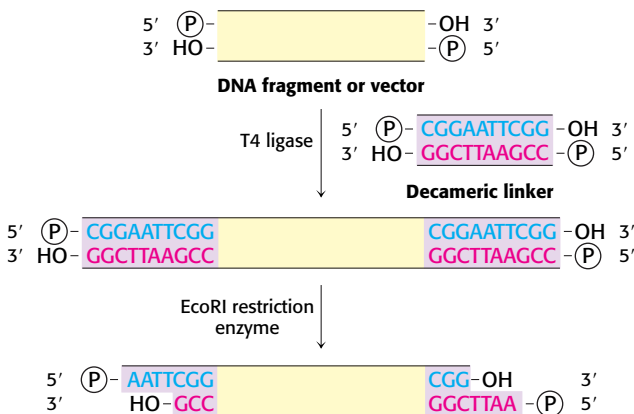


FIGURE 5.10 Formation of cohesive ends. Cohesive ends can be formed by the addition and cleavage of a chemically synthesized linker.

What if the target DNA is not naturally flanked by the appropriate restriction sites? How is the fragment cut and annealed to the vector? The cohesive-end method for joining DNA molecules can still be used in these cases by adding a *short, chemically synthesized DNA linker* that can be cleaved by restriction enzymes. First, the linker is covalently joined to the ends of a DNA fragment. For example, the 5' ends of a decameric linker and a DNA molecule are phosphorylated by polynucleotide kinase and then joined by the ligase from T4 phage (Figure 5.10). This ligase can form a covalent bond between blunt-ended double-helical DNA molecules. Cohesive ends are produced when these terminal extensions are cut by an appropriate restriction enzyme. Thus, *cohesive ends corresponding to a particular restriction enzyme can be added to virtually any DNA molecule*. We see here the fruits of combining enzymatic and synthetic chemical approaches in crafting new DNA molecules.

Plasmids and λ phage are choice vectors for DNA cloning in bacteria

Many plasmids and bacteriophages have been ingeniously modified by researchers to enhance the delivery of recombinant DNA molecules into bacteria and to facilitate the genetic selection of bacteria harboring these vectors. As already mentioned, plasmids are circular double-stranded DNA molecules that occur naturally in some bacteria. They range in size from two to several hundred kilobases. Plasmids carry genes for the inactivation of antibiotics, the production of toxins, and the breakdown of natural products. These accessory chromosomes can replicate independently of the host chromosome. In contrast with the host genome, they are dispensable under certain conditions. A bacterial cell may have no plasmids at all or it may house as many as 20 copies of a naturally-occurring plasmid.

Many plasmids have been optimized for a particular experimental task. Some engineered plasmids, for example, can achieve nearly a thousand copies per bacterial cell. One class of plasmids, known as *cloning vectors*, is particularly suitable for the facile insertion and replication of a collection of DNA fragments. These vectors often feature a *polylinker* region that includes many unique restriction sites within its sequence. This polylinker can be cleaved with many different restriction enzymes or combinations of enzymes, providing great versatility in the DNA fragments that can be inserted. In addition, these plasmids contain *reporter genes*, which encode rapidly-detectable markers such as antibiotic-resistance enzymes or fluorescent proteins. Creative placement of these reporter genes within these plasmids enables the rapid identification of those vectors that harbor the desired DNA insert. For example, in the cloning vector pUC18, insertion of DNA in the polylinker region (Figure 5.11) disrupts the *lacZ α* gene, an effect called *insertional inactivation*. This gene encodes an essential fragment of the protein β -galactosidase, an enzyme which naturally cleaves the milk sugar galactose (Section 11.2). β -Galactosidase also cleaves the synthetic substrate X-gal, releasing a blue dye. Bacterial cells containing a DNA insert at the polylinker will no longer produce the dye in the presence of X-gal, and are readily identified by their white color (Figure 5.12).

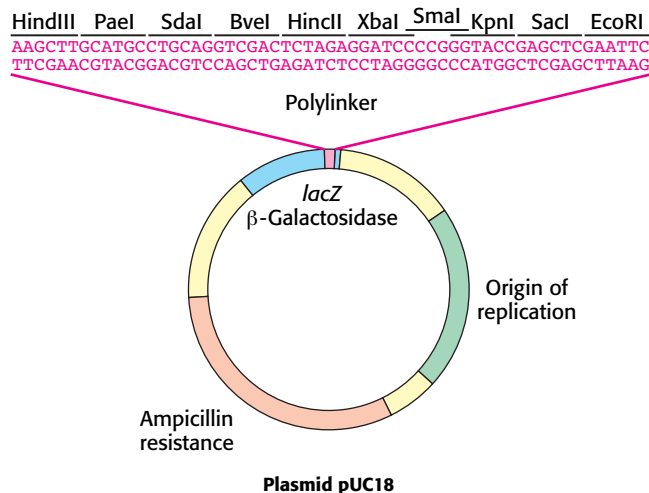


FIGURE 5.11 A polylinker in the plasmid pUC18. The plasmid pUC18 includes a polylinker within an essential fragment of the β -galactosidase gene (often called the *lacZ* gene). Insertion of a DNA fragment into one of the many restriction sites within this polylinker can be detected by the absence of β -galactosidase activity.

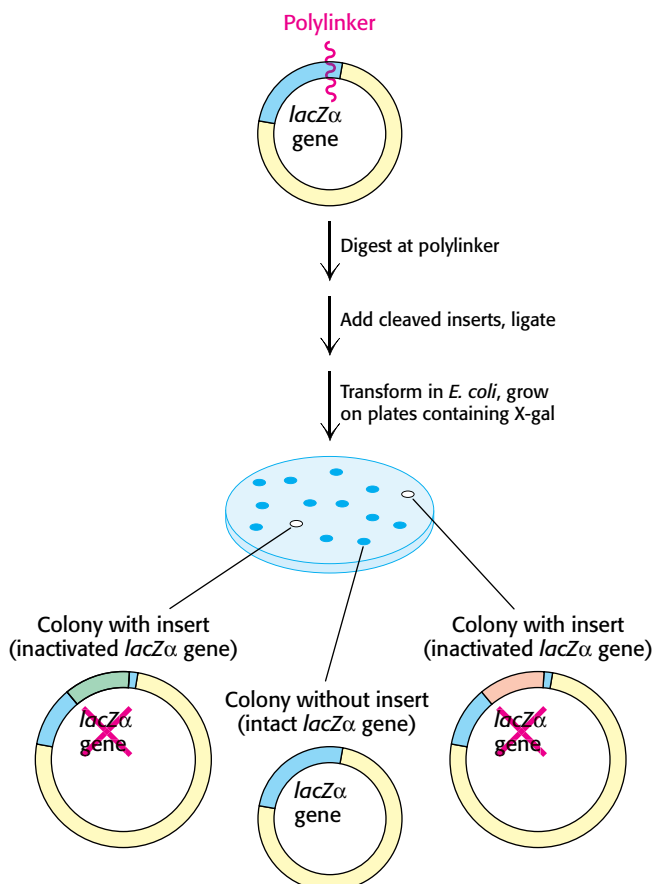


FIGURE 5.12 Insertional inactivation. Successful insertion of DNA fragments into the polylinker region of pUC18 will result in the disruption of the β -galactosidase gene. Bacterial colonies that harbor such plasmids will no longer convert X-gal into a colored product, and will appear white on the plate.

Another class of plasmids has been optimized for use as *expression vectors* for the production of large amounts of protein. In addition to antibiotic-resistance genes, they contain promoter sequences designed to drive the transcription of large amounts of a protein-coding DNA sequence. In addition to polylinkers, these vectors often contain sequences flanking the cloning site that simplify the addition of fusion tags to the protein of interest (Section 3.1), greatly facilitating the purification of the overexpressed protein.

Another widely used vector, λ *phage*, enjoys a choice of life styles: this bacteriophage can destroy its host or it can become part of its host (Figure 5.13). In the *lytic pathway*, viral functions are fully expressed: viral DNA and proteins are quickly produced and packaged into virus particles, leading to the lysis (destruction) of the host cell and the sudden appearance of about 100 progeny virus particles, or *virions*. In the *lysogenic pathway*, the phage DNA becomes inserted into the host-cell genome and can be replicated together with host-cell DNA for many generations, remaining inactive. Certain environmental changes can trigger the expression of this dormant viral DNA, which leads to the formation of progeny viruses and lysis of the host. Large segments of the 48-kb DNA of λ phage are not essential for productive infection and can be replaced by foreign DNA, thus making λ phage an ideal vector.

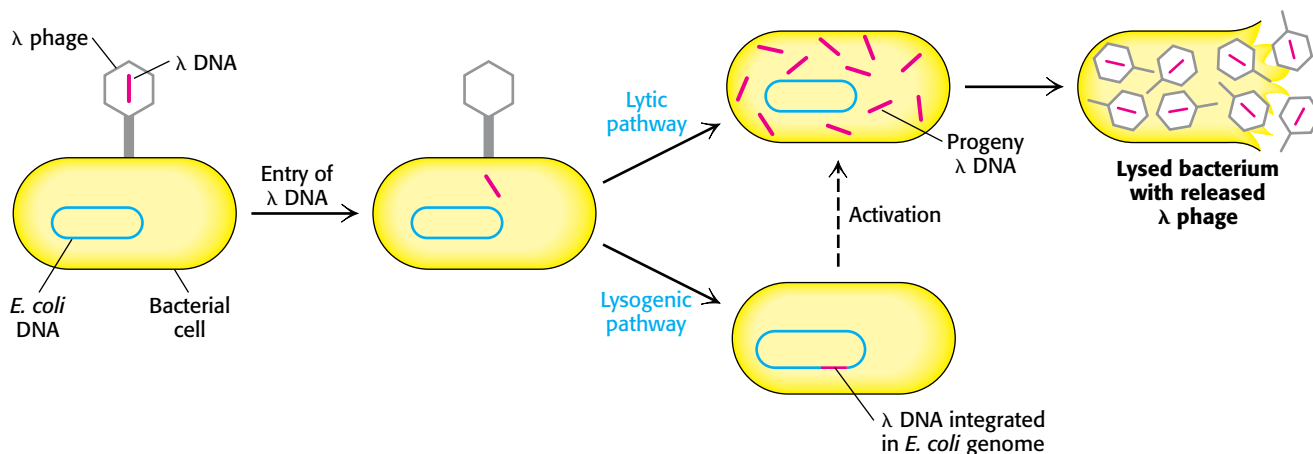


FIGURE 5.13 Alternative infection modes for λ phage. Lambda phage can multiply within a host and lyse it (lytic pathway) or its DNA can become integrated into the host genome (lysogenic pathway), where it is dormant until activated.

Mutant λ phages designed for cloning have been constructed. An especially useful one called λ gt- λ β contains only two EcoRI cleavage sites instead of the five normally present (Figure 5.14). After cleavage, the middle segment of this λ DNA molecule can be removed. The two remaining pieces of DNA (called arms) have a combined length equal to 72% of a normal genome length. This amount of DNA is too little to be packaged into a λ particle, which can take up only DNA measuring from 78% to 105% of a normal genome. However, a suitably long DNA insert (such as 10 kb) between the two ends of λ DNA enables such a recombinant DNA molecule (93% of normal length) to be packaged. Nearly all infectious λ particles formed in this way will contain an inserted piece of foreign DNA. Another advantage of using these modified viruses as vectors is that they enter bacteria much more easily than do plasmids. Among the variety of λ mutants that have been constructed for use as cloning vectors, one of them, called a *cosmid*, is essentially a hybrid of λ phage and a plasmid that can serve as a vector for large DNA inserts (as large as 45 kb).

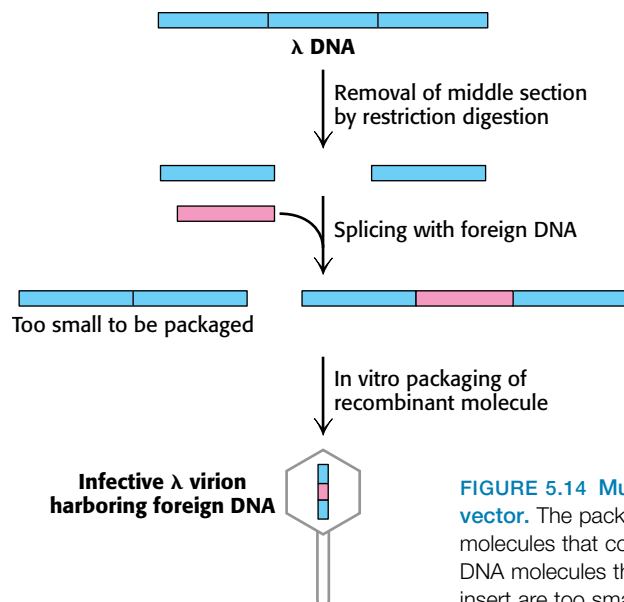


FIGURE 5.14 Mutant λ phage as a cloning vector. The packaging process selects DNA molecules that contain an insert (colored red). DNA molecules that have resealed without an insert are too small to be efficiently packaged.

Bacterial and yeast artificial chromosomes

Much larger pieces of DNA can be propagated in *bacterial artificial chromosomes* (BACs) or *yeast artificial chromosomes* (YACs). BACs are highly engineered versions of the *E. coli* fertility (F) factor that can include inserts as large as 300 kb. YACs contain a centromere, an *autonomously replicating sequence* (ARS, where replication begins), a pair of telomeres (normal ends of eukaryotic chromosomes), selectable marker genes, and a cloning site (Figure 5.15). Inserts as large as 1000 kb can be cloned into YAC vectors.

Specific genes can be cloned from digests of genomic DNA

Ingenious cloning and selection methods have made it possible to isolate small stretches of DNA in a genome containing more than 3×10^6 kb. The approach is to prepare a large collection (*library*) of DNA fragments and then to identify those members of the collection that have the gene of interest. Hence, to clone a gene that is present just once in an entire genome, two critical components must be available: a specific oligonucleotide probe for the gene of interest and a DNA library that can be screened rapidly.

How is a specific probe obtained? In one approach, a probe for a gene can be prepared if a part of the amino acid sequence of the protein encoded by the gene is known. Peptide sequencing of a purified protein (Chapter 3) or knowledge of the sequence of a homologous protein from a related species (Chapter 6) are two potential sources of such information. However, a problem arises because a single peptide sequence can be encoded by a number of different oligonucleotides (Figure 5.16). Thus, for this purpose, peptide sequences containing tryptophan and methionine are preferred, because these amino acids are specified by a single codon, whereas other amino acid residues have between two and six codons (Table 4.5). All the possible DNA sequences (or their complements) that encode the targeted peptide sequence are synthesized by the solid-phase method and made radioactive by phosphorylating their 5' ends with ^{32}P .

To prepare the DNA library, a sample containing many copies of total genomic DNA is first mechanically sheared or partly digested by restriction enzymes

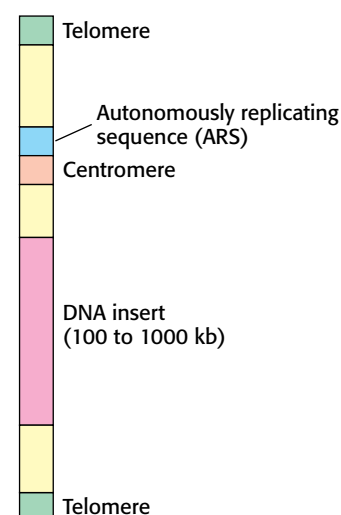


FIGURE 5.15 Diagram of a yeast artificial chromosome (YAC). These vectors include features necessary for replication and stability in yeast cells.

Amino acid sequence	...	Cys	Pro	Asn	Lys	Trp	Thr	His	...
Potential oligonucleotide sequences		TG ^C _T	CC ^C _G	AA ^C _T	AA ^A _G	TGG	AC ^C _G	CA ^C _T	

FIGURE 5.16 Probes generated from a protein sequence. A probe can be generated by synthesizing all possible oligonucleotides encoding a particular sequence of amino acids. Because of the degeneracy of the genetic code, 256 distinct oligonucleotides must be synthesized to ensure that the probe matching the sequence of seven amino acids in this example is present.

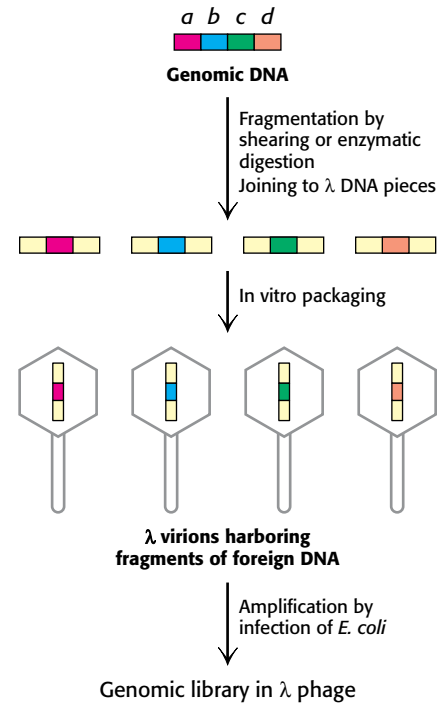


FIGURE 5.17 Creation of a genomic library. A genomic library can be created from a digest of a whole complex genome. After fragmentation of the genomic DNA into overlapping segments, the DNA is inserted into the λ phage vector (shown in yellow). Packaging into virions and amplification by infection in *E. coli* yields a genomic library.

into large fragments (Figure 5.17). This process yields a nearly random population of overlapping DNA fragments. These fragments are then separated by gel electrophoresis to isolate the set of all fragments that are about 15 kb long. Synthetic linkers are attached to the ends of these fragments, cohesive ends are formed, and the fragments are then inserted into a vector, such as λ phage DNA, prepared with the same cohesive ends. *E. coli* bacteria are then infected with these recombinant phages. These phages replicate themselves and then lyse their bacterial hosts. The resulting lysate contains fragments of human DNA housed in a sufficiently large number of virus particles to ensure that nearly the entire genome is represented. These phages constitute a *genomic library*. Phages can be propagated indefinitely such that the library can be used repeatedly over long periods.

This genomic library is then screened to find the very small number of phages harboring the gene of interest. For the human genome, a calculation shows that a 99% probability of success requires screening about 500,000 clones; hence, a very rapid and efficient screening process is essential. Rapid screening can be accomplished by DNA hybridization.

A dilute suspension of the recombinant phages is first plated on a lawn of bacteria (Figure 5.18). Where each phage particle has landed and infected a bacterium, a *plaque* containing identical phages develops on the plate. A replica of this master plate is then made by applying a sheet of nitrocellulose. Infected bacteria and phage DNA released from lysed cells adhere to the sheet in a pattern of spots corresponding to the plaques. Intact bacteria on this sheet are lysed with NaOH, which also serves to denature the DNA so that it becomes accessible for hybridization with a ^{32}P -labeled probe. The presence of a specific DNA sequence in a single spot on the replica can be detected by using a radioactive complementary DNA or RNA molecule as a probe. Autoradiography then reveals the positions of spots harboring recombinant DNA. The corresponding plaques are picked out of the intact master plate and grown. A single investigator can readily screen a million clones in a day. This method makes it possible to isolate virtually any gene, provided that a probe is available.

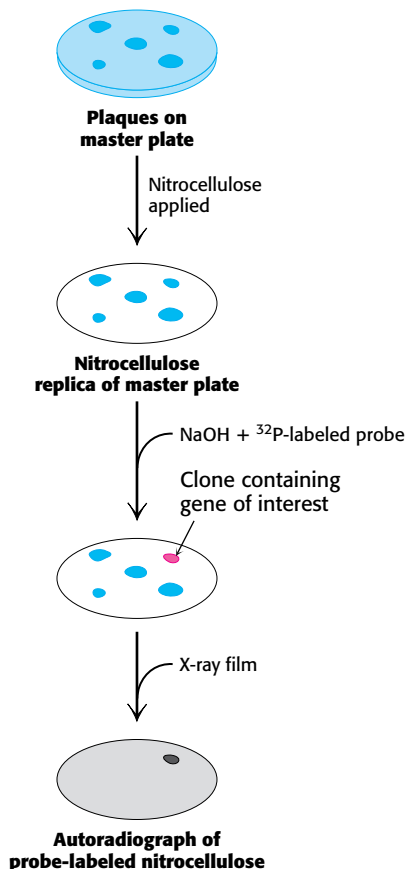


FIGURE 5.18 Screening a genomic library for a specific gene. Here, a plate is tested for plaques containing gene *a* of Figure 5.17.

Complementary DNA prepared from mRNA can be expressed in host cells

The preparation of eukaryotic DNA libraries presents unique challenges, especially if the researcher is interested primarily in the protein-coding region of a particular gene. Recall that most mammalian genes are mosaics of introns and exons. These interrupted genes cannot be expressed by bacteria, which lack the machinery to splice introns out of the primary transcript. However, this difficulty can be circumvented by causing bacteria to take up recombinant DNA that is complementary to mRNA, where the intronic sequences have been removed.

The key to forming *complementary DNA* is the enzyme *reverse transcriptase*. As discussed in Section 4.3, a retrovirus uses this enzyme to form a DNA–RNA hybrid in replicating its genomic RNA. Reverse transcriptase synthesizes a DNA strand complementary to an RNA template if the transcriptase is provided with a DNA primer that is base-paired to the RNA and contains a free 3′-OH group. We can use a simple sequence of linked thymidine [oligo(T)] residues as the primer. This oligo(T) sequence pairs with the poly(A) sequence at the 3′ end of most eukaryotic mRNA molecules (Section 4.4), as shown in Figure 5.19. The reverse transcriptase then synthesizes the rest of the cDNA strand in the presence of the four deoxyribonucleoside triphosphates (step 1). The RNA strand of this RNA–DNA hybrid is subsequently hydrolyzed by raising the pH (step 2). Unlike RNA, DNA is resistant to alkaline hydrolysis. The single-stranded DNA is converted into double-stranded DNA by creating another primer site. The enzyme *terminal transferase* adds nucleotides—for instance, several residues of dG—to the 3′ end of DNA (step 3). Oligo(dC) can bind to dG residues and prime the synthesis of the second DNA strand (step 4). Synthetic linkers can be added to this double-helical DNA for ligation to a suitable vector. Complementary DNA for all mRNA that a cell contains can be made, inserted into vectors, and then inserted into bacteria. Such a collection is called a *cDNA library*.

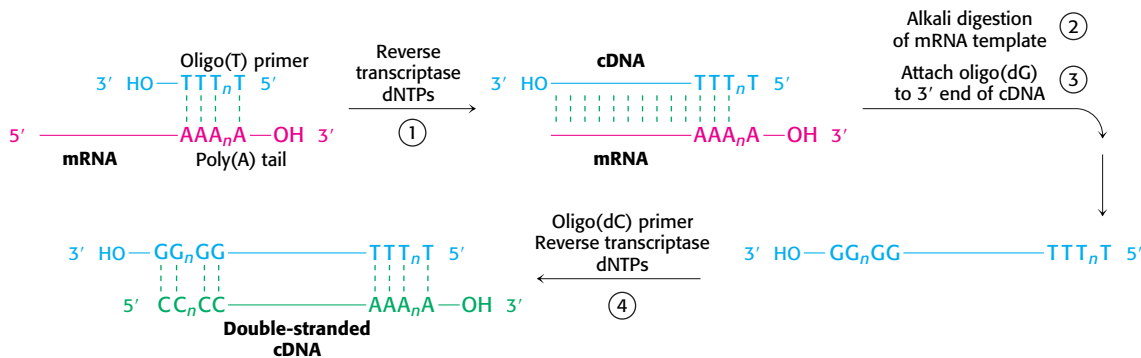


FIGURE 5.19 Formation of a cDNA duplex. A complementary DNA (cDNA) duplex is created from mRNA by (1) use of reverse transcriptase to synthesize a cDNA strand, (2) digestion of the original RNA strand, (3) addition of several G bases to the DNA by terminal transferase, and (4) synthesis of a complementary DNA strand using the newly synthesized cDNA strand as a template.

Complementary DNA molecules can be inserted into expression vectors to enable the production of the corresponding protein of interest. Clones of cDNA can be screened on the basis of their capacity to direct the synthesis of a foreign protein in bacteria, a technique referred to as *expression cloning*. A labeled antibody specific for the protein of interest can be used to identify colonies of bacteria that express the corresponding protein product (Figure 5.20). As described earlier, spots of bacteria on a replica plate are lysed to release proteins, which bind to an applied nitrocellulose

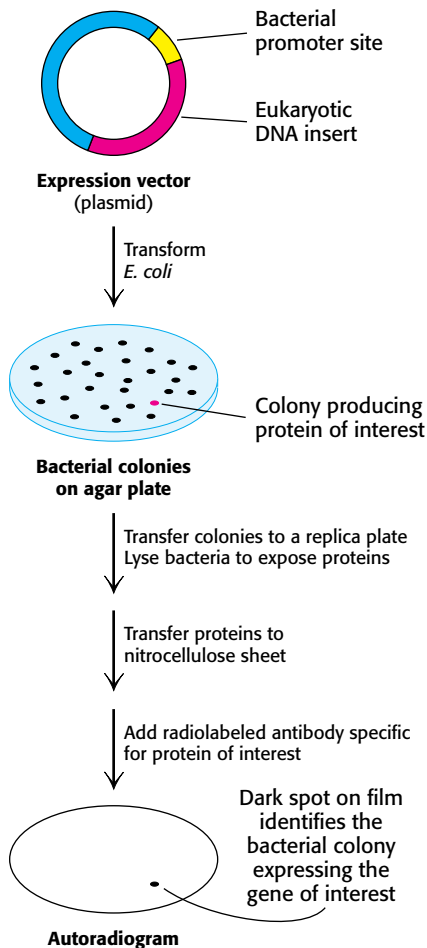
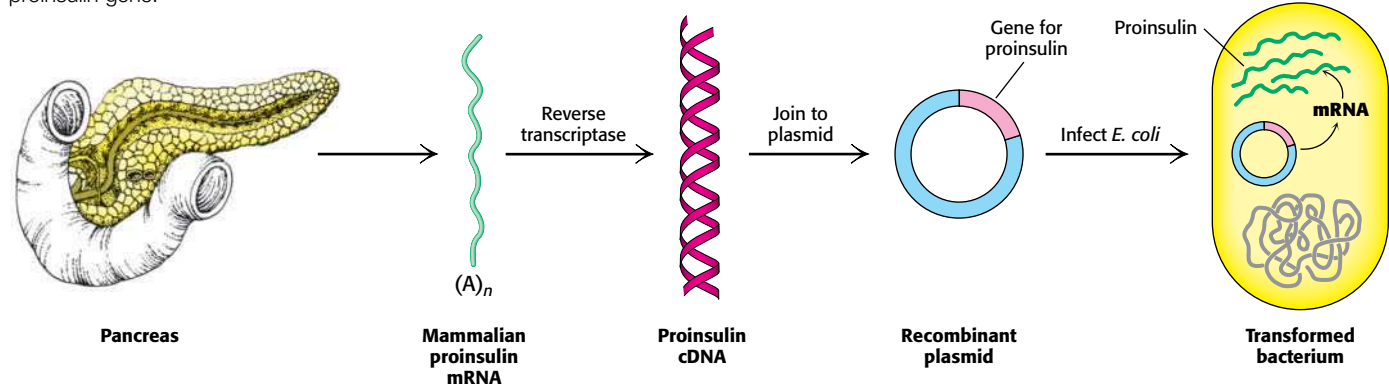


FIGURE 5.20 Screening of cDNA clones. A method of screening for cDNA clones is to identify expressed products by staining with specific antibody.

FIGURE 5.21 Synthesis of proinsulin by bacteria. Proinsulin, a precursor of insulin, can be synthesized by transformed (genetically altered) clones of *E. coli*. The clones contain the mammalian proinsulin gene.



filter. With the addition of labeled antibody specific for the protein of interest, the location of the desired colonies on the master plate can be readily identified. This immunochemical screening approach can be used whenever a protein is expressed and corresponding antibody is available.

Complementary DNA has many applications beyond the generation of genetic libraries. The overproduction and purification of most eukaryotic proteins in prokaryotic cells necessitates the insertion of cDNA into plasmid vectors. For example, proinsulin, a precursor of insulin, is synthesized by bacteria-harboring plasmids that contain DNA complementary to mRNA for proinsulin (Figure 5.21). Indeed, bacteria produce much of the insulin used today by millions of diabetics.

Proteins with new functions can be created through directed changes in DNA

Much has been learned about genes and proteins by analyzing the effects that mutations have on their structure and function. In the classic genetic approach, mutations are generated randomly throughout the genome of a host organism, and those individuals exhibiting a phenotype of interest are selected. Analysis of these mutants then reveals which genes are altered, and DNA sequencing identifies the precise nature of the changes. *Recombinant DNA technology makes the creation of specific mutations feasible in vitro*. We can construct new genes with designed properties by making three kinds of directed changes: *deletions*, *insertions*, and *substitutions*. A variety of methods can be used to introduce these types of mutations, including the following examples.

Site-directed mutagenesis. Mutant proteins with single amino acid substitutions can be readily produced by *site-directed mutagenesis* (Figure 5.22). Suppose that we want to replace a particular serine residue with cysteine. This mutation can be made if (1) we have a plasmid containing the gene or cDNA for the protein and (2) we know the base sequence around the site to be altered. If the serine of interest is encoded by TCT, mutation of the central base from C to G yields the TGT codon, which encodes cysteine. This type of mutation is called a *point mutation* because only one base is altered. To introduce this mutation into our plasmid, we prepare an oligonucleotide primer that is complementary to this region of the gene except that it contains TGT instead of TCT. The two strands of the plasmid are separated, and the primer is then annealed to the complementary strand. The mismatch of 1 of 15 base pairs is tolerable if the annealing is carried out at an appropriate temperature. After annealing to the complementary strand, the primer is elongated by DNA polymerase, and the double-stranded circle is closed by adding DNA ligase. Subsequent replication of this duplex yields two kinds of progeny plasmid, half with the original TCT sequence and half with the mutant TGT sequence. Expression of the plasmid containing the new TGT

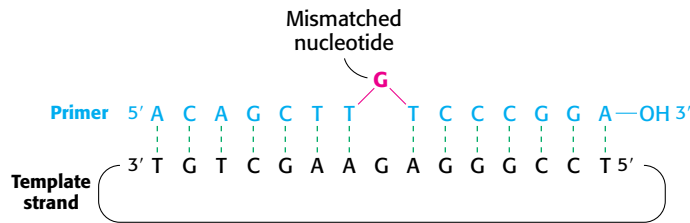


FIGURE 5.22 Oligonucleotide-directed mutagenesis. A primer containing a mismatched nucleotide is used to produce a desired change in the DNA sequence.

sequence will produce a protein with the desired substitution of cysteine for serine at a unique site. We will encounter many examples of the use of site-directed mutagenesis to precisely alter regulatory regions of genes and to produce proteins with tailor-made features.

Cassette mutagenesis. In *cassette mutagenesis*, a variety of mutations, including insertions, deletions, and multiple point mutations, can be introduced into the gene of interest. A plasmid harboring the original gene is cut with a pair of restriction enzymes to remove a short segment (Figure 5.23). A synthetic double-stranded oligonucleotide—the *cassette*—carrying the genetic alterations of interest is prepared with cohesive ends that are complementary to the ends of the cut plasmid. Ligation of the cassette into the plasmid yields the desired mutated gene product.

Mutagenesis by PCR. In Section 5.1, we learned how PCR can be used to amplify a specific region of DNA using primers that flank the region of interest. In fact, the creative design of PCR primers enables the introduction of specific insertions, deletions, and substitutions into the amplified sequence. A number of methods have been developed for this purpose. Here, we shall consider one: *inverse PCR* to introduce deletions into plasmid DNA (Figure 5.24). In this approach, primers are designed to flank the sequence to be deleted. However, these primers are oriented in the opposite

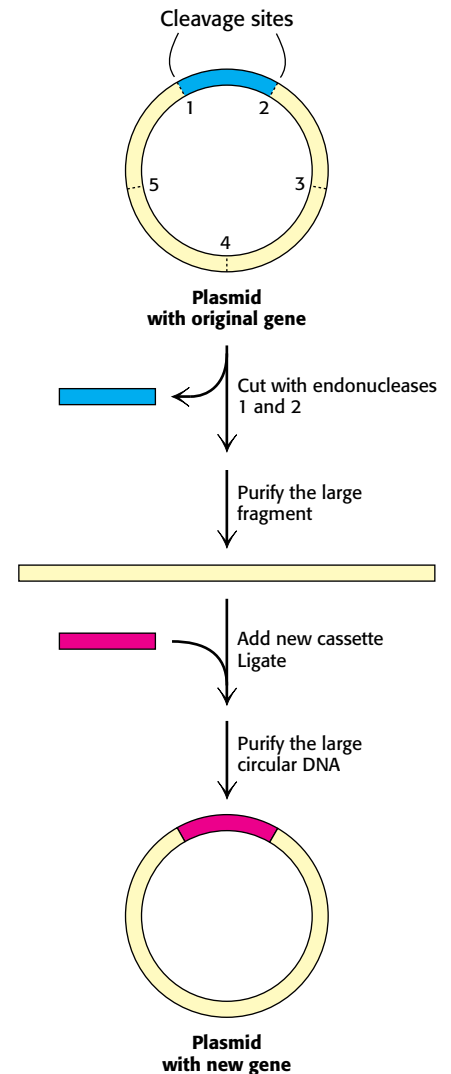


FIGURE 5.23 Cassette mutagenesis. DNA is cleaved at a pair of unique restriction sites by two different restriction endonucleases. A synthetic oligonucleotide with ends that are complementary to these sites (the cassette) is then ligated to the cleaved DNA. The method is highly versatile because the inserted DNA can have any desired sequence.

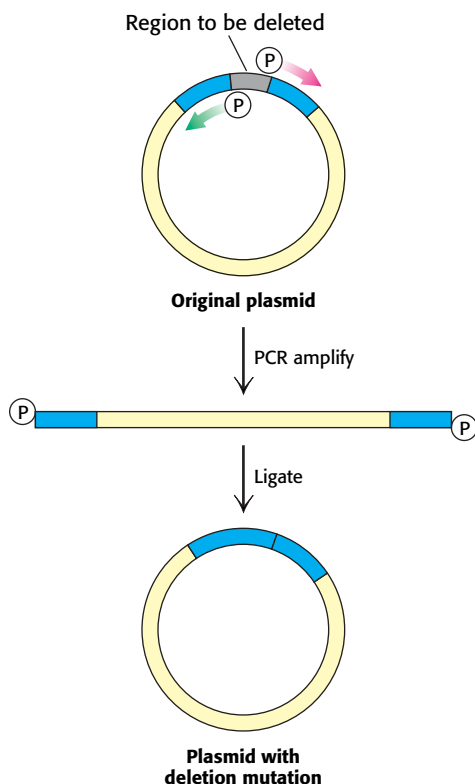


FIGURE 5.24 Deletion mutagenesis by inverse PCR. A deletion can be introduced into a plasmid with primers that flank this region but are oriented away from the segment to be removed. PCR amplification yields a linear product that contains the entire plasmid minus the unwanted sequence. If the primers contained a 5' phosphate group, this product can be recircularized using DNA ligase, generating a plasmid with the desired mutation.

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/726103225243010100>