19 Molecular Genetic Analysis and Biotechnology

Genetic engineering is being used to modify rice and other crops to grow in environments that are currently unable to support agriculture. *[Friedrich Stark/Peter Arnold.]*

FEEDING THE FUTURE POPULATION OF THE WORLD

In the year 2000, the world's population reached 6 billion. Because the human population has exhibited exponential growth, the pace of increase in the number of people is ever quickening: more than 100,000 years were required for humans to reach 1 billion in number (in 1830); only another 100 years were required for the population to double to 2 billion (in 1930); and only 45 years were required for it to double again (in 1975) to 4 billion. The United Nations projects that the world population will reach somewhere between 7.3 billion and 10.7 billion by the year 2050 and, because the tendency is for people to have smaller families, will eventually level off or even drop in the last part of the twenty-first century.

How will we feed the additional billions of people that will populate the planet Earth 50 years from now? Up to this point, we have been able to sustain the tremendous increase in human numbers because advances in agriculture have greatly increased worldwide food production. Much of this increase was between 1950 and 1980 through the Green Revolution, which utilized traditional techniques of plant breeding and genetics to develop new varieties of corn, wheat, and rice. For example, worldwide grain production increased 260% between 1950 and 1990; worldwide cereal production increased from 275 kg/person in the 1950s to 370 kg/person in the 1980s, during a time in which human population almost doubled. Thus, even though human numbers have increased tremendously in the past 50 years, the world's farmers today produce more food per person than they did in 1950.

What about the next 50 years, when there will be between 1 billion and 5 billion more people to feed? Most of the world's cultivatable land is already in use, and increases in crop yield achievable through traditional breeding and genetics have leveled off. Many experts propose that feeding the future population of the world can be achieved only through the application of genetic engineering to bring about a "second" Green Revolution. Already, genetic engineering has been used to produce crops that are resistant to pests, disease, and herbicides. Genetically engineered (often called genetically modified) crops are today cultivated on more than 60 million hectares (1 hectare = 2.471 acres) of land worldwide; in 2002, 40% of corn, 81% of soybeans, and 73% of cotton grown in North America was genetically engineered.

The potential of genetic engineering to help feed the future world population must be weighed against concerns about the widespread use of genetically modified crops. Although recent scientific reviews contain little evidence of risk to human health from eating genetically modified foods, many consumers remain wary of eating them. Findings from recent studies in the United Kingdom demonstrated that genetically modified beets and oilseed rape reduce the biodiversity of native plants and insects in agricultural fields, and there are concerns that genetically modified plants may hybridize with native plants and cause ecological disruption.

This chapter introduces some of the techniques being used to create genetically engineered crops and other organisms. We begin by considering molecular genetic technology and some of its effects. We then examine a number of methods used to isolate, study, alter, and recombine DNA sequences and place them back into cells. Finally, we explore some of the applications of molecular genetic analysis.

19.1 Techniques of Molecular Genetics Have Revolutionized Biology

In 1973, a group of scientists produced the first organisms with recombinant DNA molecules. Stanley Cohen at Stanford University and Herbert Boyer at the University of California School of Medicine at San Francisco and their colleagues inserted a piece of DNA from one plasmid into another, creating an entirely new, recombinant DNA molecule. They then introduced the recombinant plasmid into *E. coli* cells. These experiments ushered in one of the most momentous revolutions in the history of science.

Recombinant DNA technology is a set of molecular techniques for locating, isolating, altering, and studying DNA segments. The term *recombinant* is used because, frequently, the goal is to combine DNA from two distinct sources. Genes from two different bacteria might be joined, for example, or a human gene might be inserted into a viral chromosome. Commonly called **genetic engineering,** recombinant DNA technology now encompasses many molecular techniques that can be used to analyze, alter, and recombine virtually any DNA sequences from any number of sources.

The Molecular Genetics Revolution

The techniques of recombinant DNA technology are just a part of a vast array of molecular methods that are now available for the study of genetics. These molecular techniques have drastically altered the way that genes are studied. Previously, information about the structure and organization of genes was gained by examining their phenotypic effects, but molecular genetic analysis allows the nucleotide sequences themselves to be read. Methods in molecular genetics have provided new information about the structure and function of genes and has altered many fundamental concepts of genetics. Our detailed understanding of genetic processes such as replication, transcription, translation, RNA processing, and gene regulation has been learned through the use of molecular genetic techniques. These techniques are used in many other fields as well, including biochemistry, microbiology, developmental biology, neurobiology, evolution, and ecology.

Recombinant DNA technology and other molecular techniques are also being used to create a number of commercial products, including drugs, hormones, enzymes, and crops (**Figure 19.1**). A complete industry—**biotechnology**—has grown up around the use of these techniques to develop new products. In medicine, molecular genetics is being used to probe the nature of cancer, diagnose genetic and infectious diseases, produce drugs, and treat hereditary disorders.

CONCEPTS

Molecular genetics and recombinant DNA technology are used to locate, analyze, alter, study, and recombine DNA sequences. These techniques are used to probe the structure and function of genes, address questions in many areas of biology, create commercial products, and diagnose and treat diseases.

Working at the Molecular Level

The manipulation of genes at the molecular level presents a serious challenge, often requiring strategies that may not, at first, seem obvious. The basic problem is that genes are minute and every cell contains thousands of them. Individual nucleotides cannot be seen, and no physical features mark the beginning or the end of a gene.

Let's consider a typical situation faced by a molecular geneticist. Suppose we wanted to use bacteria to produce large quantities of a human protein. The first and most formidable problem is to find the gene that encodes the desired protein. A haploid human genome consists of 3.2 billion base pairs of DNA. Let's assume that the gene that we want to isolate is 3000 bp long. Our target gene occupies only one-millionth of the genome; so searching for our gene in the huge expanse of genomic DNA is more difficult than looking

19.1 Recombinant DNA technology has been used to create genetically modified crops. Genetically engineered corn, which produces a toxin that kills insect pests, now constitutes more than 40% of all corn grown in the United States. *[Chris Knapton/Photo Researchers.]*

for a needle in the proverbial haystack. But, even if we are able to locate the gene, how are we to separate it from the rest of the DNA?

If we did succeed in locating and isolating the desired gene, we would next need to insert it into a bacterial cell. Linear fragments of DNA are quickly degraded by bacteria; so the gene must be inserted in a stable form. It must also be able to successfully replicate or it will not be passed on when the cell divides. If we succeed in transferring our gene to bacteria in a stable form, we must still ensure that the gene is properly transcribed and translated.

Finally, the methods used to isolate and transfer genes are inefficient and, of a million cells that are subjected to these procedures, only *one* cell might successfully take up and express the human gene. So we must search through many bacterial cells to find the one containing the recombinant DNA. We are back to the problem of the needle in the haystack.

Although these problems might seem insurmountable, molecular techniques have been developed to overcome all of them, and human genes are routinely transferred to bacterial cells, where the genes are expressed.

CONCEPTS

Molecular genetic analyses require special methods because individual genes make up a tiny fraction of the cellular DNA and they cannot be seen.

✔ CONCEPT CHECK 1

Briefly outline the steps required to genetically engineer bacteria that will produce a protein encoded by a human gene.

19.2 Molecular Techniques Are Used to Isolate, Recombine, and Amplify Genes

A first step in the molecular analysis of a DNA segment or gene is to isolate it from the remainder of the DNA and to make many copies of it so that further analyses can be carried out. The isolation and amplification of DNA frequently requires that it be recombined with other DNA molecules. In the sections that follow, we will examine some of the molecular techniques that are used to isolate, recombine, and amplify DNA segments.

Cutting and Joining DNA Fragments

A key discovery in the development of molecular genetic methods was the discovery in the late 1960s of **restriction enzymes** (also called **restriction endonucleases**) that recognize and make double-stranded cuts in DNA at specific nucleotide sequences. These enzymes are produced naturally by bacteria, where they are used in defense against viruses. A bacterium protects its own DNA from a restriction enzyme by modifying the recognition sequence, usually by adding methyl groups to its DNA.

Three types of restriction enzymes have been isolated from bacteria. Type I and type III restriction enzymes cut DNA at sites outside of the recognition sequences. Type II restriction enzymes recognize specific sequences and cut the DNA within the recognition sequence. Virtually all molecular genetics work is done with type II restriction enzymes; discussions of restriction enzymes throughout this book refer to type II enzymes.

More than 800 different restriction enzymes that recognize and cut DNA at more than 100 different sequences have been isolated from bacteria. Many of these enzymes are commercially available; examples of some commonly used restriction enzymes are given in **Table 19.1.** The name of each restriction enzyme begins with an abbreviation that signifies its bacterial origin.

Table 19.1 Characteristics of some common type II restriction enzymes used in recombinant DNA technology

Enzyme	Microorganism from Which Enzyme Is Produced	Recognition Sequence	Type of Fragment End Produced
*Bam*HI	*Bacillus amyloliquefaciens*	5′–GGATCC–3′ 3′–CCATAGG–3′	Cohesive
*Cof*I	*Clostridium formicoaceticum*	5′–GCGC–3′ 3′–CGCG–5′	Cohesive
*Eco*RI	*Escherichia coli*	5′–GAATTC–3′ 3′–CTTAAG–5′	Cohesive
*Eco*RII	*Escherichia coli*	5′–CCAGG–3′ 3′–GGTCC–5′	Cohesive
*Hae*III	*Haemophilus aegyptius*	5′–GGCC–3′ 3′–CCGG–5′	Blunt
*Hin*dIII	*Haemophilus influenzae*	5′–AAGCTT–3′ 3′–TTCGAA–5′	Cohesive
*Pvu*II	*Proteus vulgaris*	5′–CAGCTG–3′ 3′–GTCGAC–5′	Blunt

Note: The first three letters of the abbreviation for each restriction enzyme refer to the bacterial species from which the enzyme was isolated (e.g., *Eco* refers to *E. coli*). A fourth letter may refer to the strain of bacteria from which the enzyme was isolated (the "R" in *Eco*RI indicates that this enzyme was isolated from the RY13 strain of *E. coli*). Roman numerals that follow the letters allow different enzymes from the same species to be identified.

The sequences recognized by restriction enzymes are usually from 4 to 8 bp long; most enzymes recognize a sequence of 4 or 6 bp. Most recognition sequences are palindromic—sequences that read the same forward and backward. All type II restriction enzymes recognize palindromic sequences.

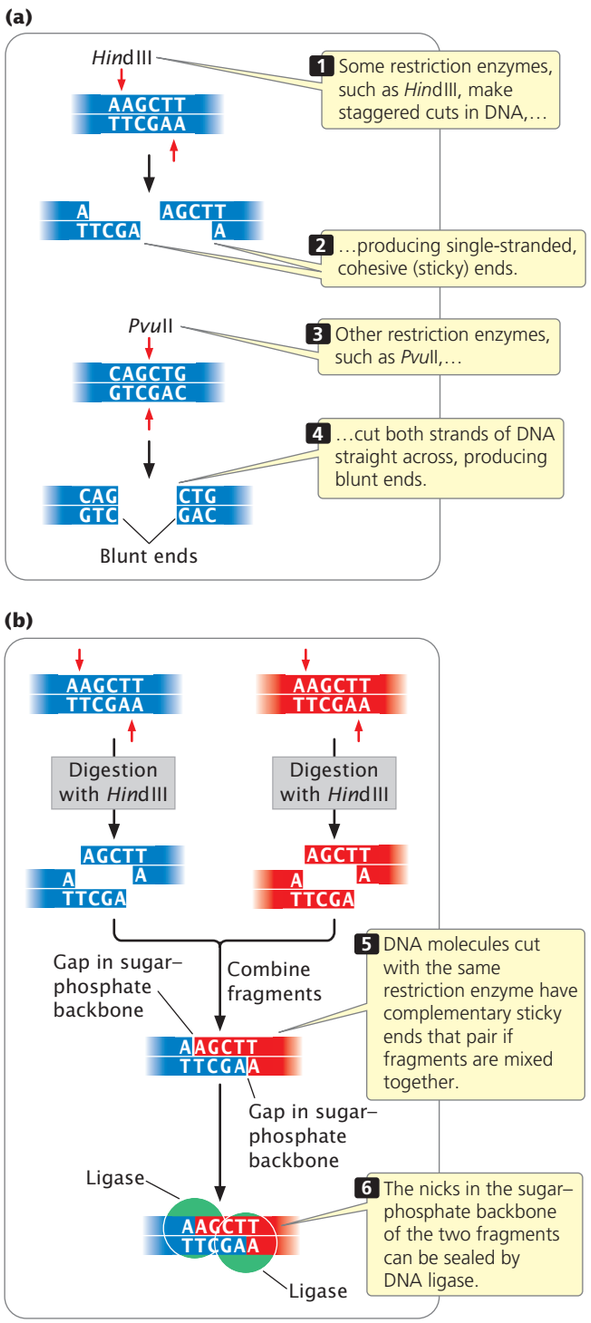
Some of the enzymes make staggered cuts in the DNA. For example, *Hin*dIII recognizes the following sequence:

5′–AAGCTT–3′
3′–TTCGAA–5′

*Hin*dIII cuts the sugar–phosphate backbone of each strand at the point indicated by the arrow, generating fragments with short, single-stranded overhanging ends:

5′–A AGCTT–3′
3′–TTCGA A–5′

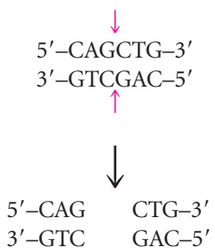
Such ends are called **cohesive ends** or sticky ends, because they are complementary to each other and can spontaneously pair to connect the fragments. Thus, DNA fragments can be "glued" together: any two fragments cleaved by



19.2 Restriction enzymes make double-stranded cuts in DNA, producing cohesive, or sticky, ends.

the same enzyme will have complementary ends and will pair (**Figure 19.2**). When their cohesive ends have paired, two DNA fragments can be joined together permanently by DNA ligase, which seals nicks between the sugar–phosphate groups of the fragments.

Not all restriction enzymes produce staggered cuts and sticky ends. *Pvu*II cuts in the middle of its recognition site, producing blunt-ended fragments:



Fragments with blunt ends must be joined together in other ways.

The sequences recognized by a restriction enzyme are located randomly within the genome. Consequently, there is a relation between the length of the recognition sequence and the number of times that it is present in a genome: there will be fewer longer recognition sequences than shorter recognition sequences, because the probability of the occurrence of a particular sequence consisting of, say, six specific bases is less than the probability of the occurrence of a particular sequence of four specific bases. Consequently, restriction enzymes that recognize longer sequences will cut a given piece of DNA into fewer and longer fragments than will restriction enzymes that recognize shorter sequences.

Restriction enzymes are used whenever DNA fragments must be cut or joined. In a typical restriction reaction, a concentrated solution of purified DNA is placed in a small tube with a buffer solution and a small amount of restriction enzyme. The reaction mixture is then heated at the optimal temperature for the enzyme, usually 37°C. Within a few hours, the enzyme cuts all the appropriate restriction sites in the DNA, producing a set of DNA fragments.

CONCEPTS

Type II restriction enzymes cut DNA at specific base sequences that are palindromic. Some restriction enzymes make staggered cuts, producing DNA fragments with cohesive ends; others cut both strands straight across, producing blunt-ended fragments. There are fewer long recognition sequences in DNA than short sequences.

✔ CONCEPT CHECK 2

Where do restriction enzymes come from?

Viewing DNA Fragments

After the completion of a restriction reaction, a number of questions arise. Did the restriction enzyme cut the DNA?

Into how many fragments was the DNA cut? What are the sizes of the resulting fragments? Gel electrophoresis provides us with a means of answering these questions.

Electrophoresis is a standard biochemical technique for separating molecules on the basis of their size and electrical charge. There are a number of different types of electrophoresis; to separate DNA molecules, **gel electrophoresis** is used. A porous gel is often made from agarose (a polysaccharide isolated from seaweed), which is melted in a buffer solution and poured into a plastic mold. As it cools, the agarose solidifies, making a gel that looks something like stiff gelatin.

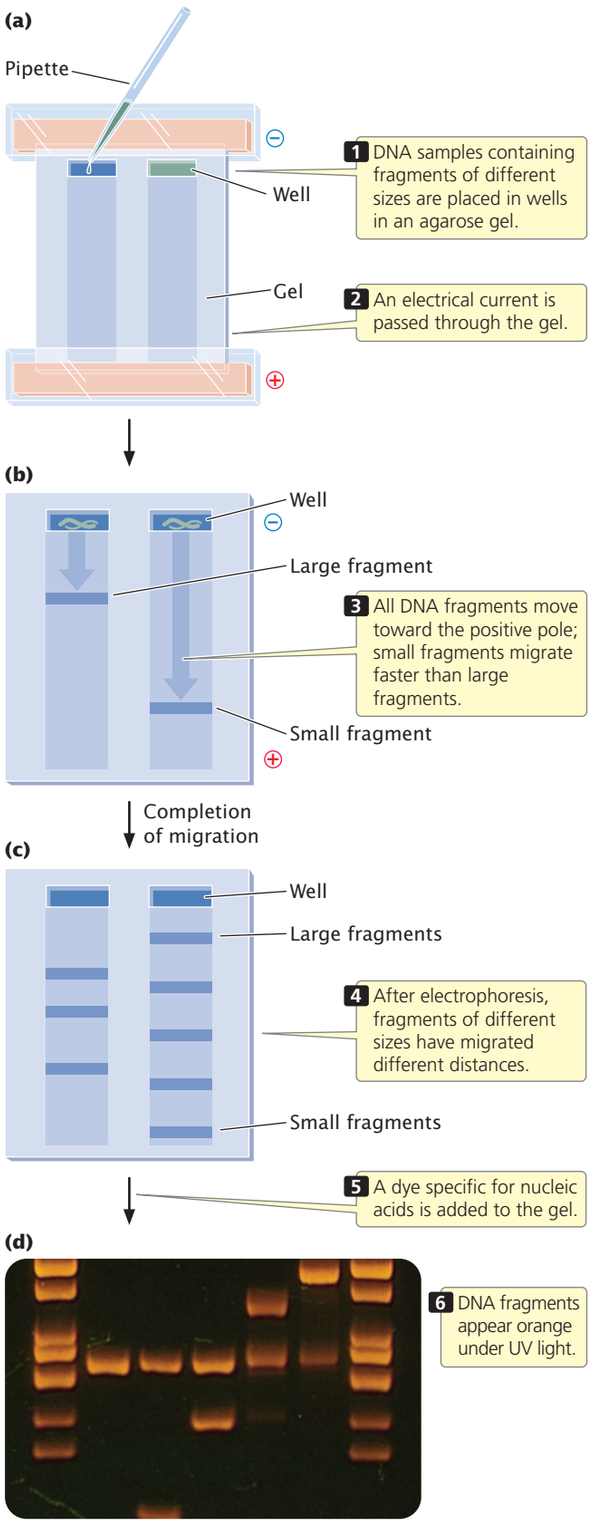
Small wells are made at one end of the gel to hold solutions of DNA fragments (**Figure 19.3a**), and an electrical current is passed through the gel. Because the phosphate group of each DNA nucleotide carries a negative charge, the DNA fragments migrate toward the positive end of the gel (**Figure 19.3b**). In this migration, the porous gel acts as a sieve, separating the DNA fragments by size. Small DNA fragments migrate more rapidly than do large ones and, with the passage of time, the fragments separate on the basis of their size. Typically, DNA fragments of known length (a marker sample) are placed in another well. By comparing the migration distance of the unknown fragments with the distance traveled by the marker fragments, one can determine the approximate size of the unknown fragments (**Figure 19.3c**).

The DNA fragments are still too small to see; so the problem of visualizing the DNA needs to be addressed. Visualization can be accomplished in several ways. The simplest procedure is to stain the gel with a dye specific for nucleic acids, such as ethidium bromide, which wedges itself tightly (intercalates) between the bases of DNA and fluoresces orange when exposed to UV light, producing brilliant orange bands on the gel (**Figure 19.3d**).

Alternatively, DNA fragments can be visualized by adding a radioactive or chemical label to the DNA before it is placed in the gel. Nucleotides with radioactively labeled phosphate (^{32}P) can be used as the substrate for DNA synthesis and will be incorporated into the newly synthesized DNA strand. Radioactively labeled DNA can be detected with a technique called **autoradiography** (see p. xxx), in which a piece of X-ray film is placed on top of the gel. Radiation from the labeled DNA exposes the film, just as light exposes photographic film in a camera. The developed autoradiograph gives a picture of the fragments in the gel, with each DNA fragment appearing as a dark band on the film. Chemical labels can be detected by adding antibodies or other substances that carry a dye and will attach to the relevant DNA, which can be visualized directly.

CONCEPTS

DNA fragments can be separated, and their sizes can be determined with the use of gel electrophoresis. The fragments can be viewed by using a dye that is specific for nucleic acids or by labeling the fragments with a radioactive or chemical tag.



19.3 Gel electrophoresis can be used to separate DNA molecules on the basis of their size and electrical charge. *[Photograph courtesy of Carol Eng.]*

✔ CONCEPT CHECK 3

DNA fragments that are 500 bp, 1000 bp, and 2000 bp in length are separated by gel electrophoresis. Which fragment will migrate farthest in the gel?

a. 2000-bp fragment
b. 1000-bp fragment
c. 500-bp fragment
d. All will migrate equal distances.

Locating DNA Fragments with Southern Blotting and Probes

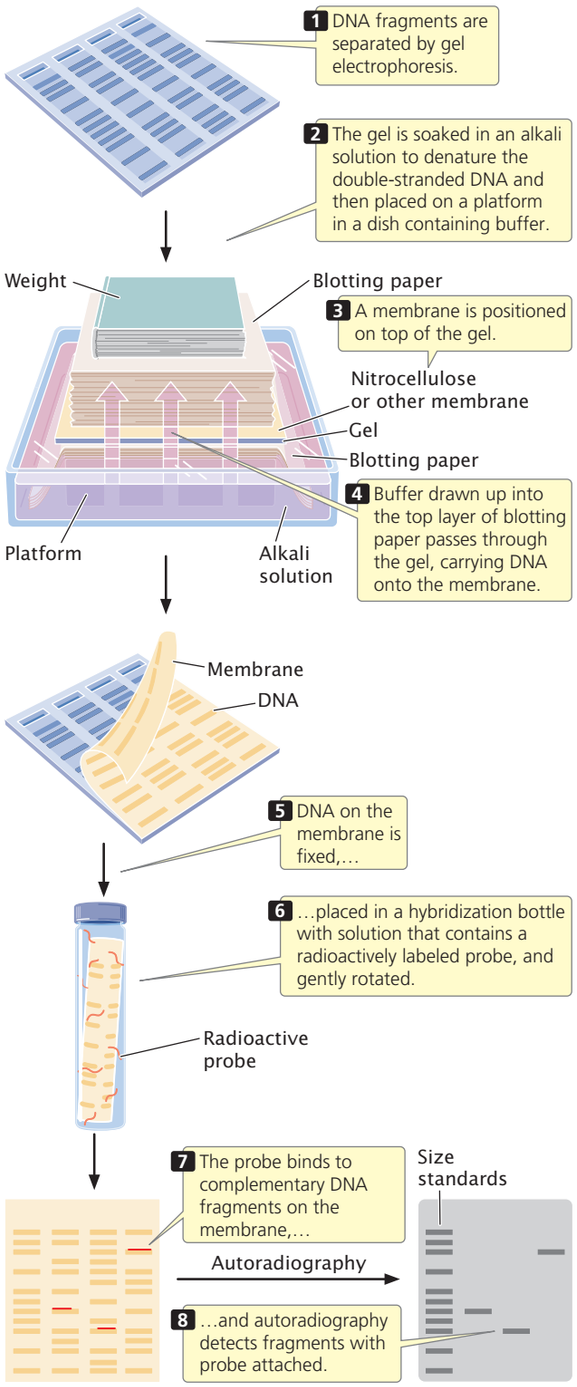
If a small piece of DNA, such as a plasmid, is cut by a restriction enzyme, the few fragments produced can be seen as distinct bands on an electrophoretic gel. In contrast, if genomic DNA from a cell is cut by a restriction enzyme, a large number of fragments of different sizes are produced. A restriction enzyme that recognizes a four-base sequence would theoretically cut about once every 256 bp. The human genome, with 3.3 billion base pairs, would generate more than 12 million fragments when cut by this restriction enzyme. Separated by electrophoresis and stained, this large set of fragments would appear as a continuous smear on the gel because of the presence of so many fragments of differing size. Usually, one is interested in only a few of these fragments, perhaps those carrying a specific gene. How does one locate the desired fragments in such a large pool of DNA?

One approach is to use a **probe,** which is a DNA or RNA molecule with a base sequence complementary to a sequence in the gene of interest. The bases on a probe will pair only with the bases on a complementary sequence and, if suitably labeled, the probe can be used to locate a specific gene or other DNA sequence.

To use a probe, one first cuts the DNA into fragments by using one or more restriction enzymes and then separates the fragments with gel electrophoresis (**Figure 19.4**). Next, the separated fragments must be denatured and transferred to a permanent solid medium (such as nitrocellulose or nylon membrane). **Southern blotting** (named after Edwin M. Southern) is one technique used to transfer the denatured, single-stranded fragments from a gel to a permanent solid medium.

After the single-stranded DNA fragments have been transferred, the membrane is placed in a hybridization solution of a radioactively or chemically labeled probe (see Figure 19.4). The probe will bind to any DNA fragments on the membrane that bear complementary sequences. The membrane is then washed to remove any unbound probe; bound probe is detected by autoradiography or another method for chemically labeled probes.

RNA can be transferred from a gel to a solid support by a related procedure called **Northern blotting** (not named after anyone but capitalized to match Southern). The hybridization of a probe can reveal the size of a particular mRNA molecule, its relative abundance, or the tissues in which the mRNA is transcribed. **Western blotting** is the transfer of protein from a gel to a membrane. Here, the probe is usually an



19.4 Southern blotting and hybridization with probes can locate a few specific fragments in a large pool of DNA.

antibody, used to determine the size of a particular protein and the pattern of the protein's expression.

CONCEPTS

Labeled probes, which are sequences of RNA or DNA that are complementary to the sequence of interest, can be used to locate individual genes or DNA sequences. Southern blotting can be used to transfer DNA fragments from a gel to a membrane such as nitrocellulose.

✔ CONCEPT CHECK 4

How do Northern and Western blotting differ from Southern blotting?

Cloning Genes

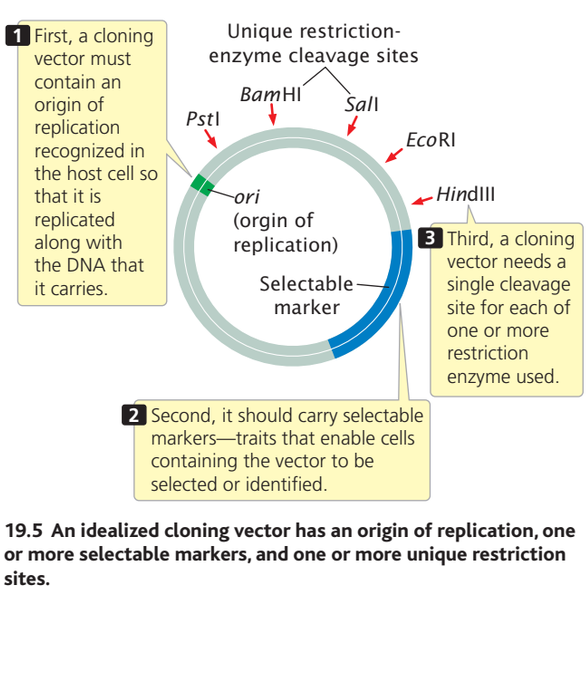
Many recombinant DNA methods require numerous copies of a specific DNA fragment. One way to amplify a specific piece of DNA is to place the fragment in a bacterial cell and allow the cell to replicate the DNA. This procedure is termed **gene cloning,** because identical copies (clones) of the original piece of DNA are produced.

A **cloning vector** is a stable, replicating DNA molecule to which a foreign DNA fragment can be attached for introduction into a cell. An effective cloning vector has three important characteristics (**Figure 19.5**): (1) an origin of replication, which ensures that the vector is replicated within the cell; (2) selectable markers, which enable any cells containing the vector to be selected or identified; and (3) one or more unique restriction sites into which a DNA fragment can be inserted. The restriction sites used for cloning must be unique; if a vector is cut at multiple recognition sites, generating several pieces of DNA, there will be no way to get the pieces back together in the correct order.

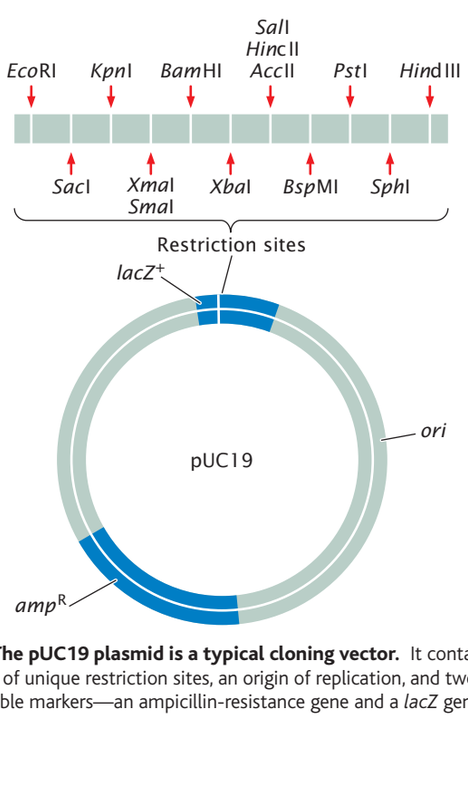
Plasmid vectors Plasmids, circular DNA molecules that exist naturally in bacteria (see Chapter 8), are commonly used vectors for cloning DNA fragments in bacteria. They contain origins of replication and are therefore able to replicate independently of the bacterial chromosome. The plasmids typically used in cloning have been constructed from the larger, naturally occurring bacterial plasmids and have multiple restriction sites, an origin of replication site, and selectable markers (**Figure 19.6**).

The easiest method for inserting a gene into a plasmid vector is to cut the foreign DNA (containing the gene) and the plasmid with the same restriction enzyme (**Figure 19.7**). If the restriction enzyme makes staggered cuts in the DNA, complementary sticky ends are produced on the foreign DNA and the plasmid. The DNA and plasmid are then mixed together; some of the foreign DNA fragments will pair with the cut ends of the plasmid. DNA ligase is used to seal the nicks in the sugar–phosphate backbone, creating a recombinant plasmid that contains the foreign DNA fragment.

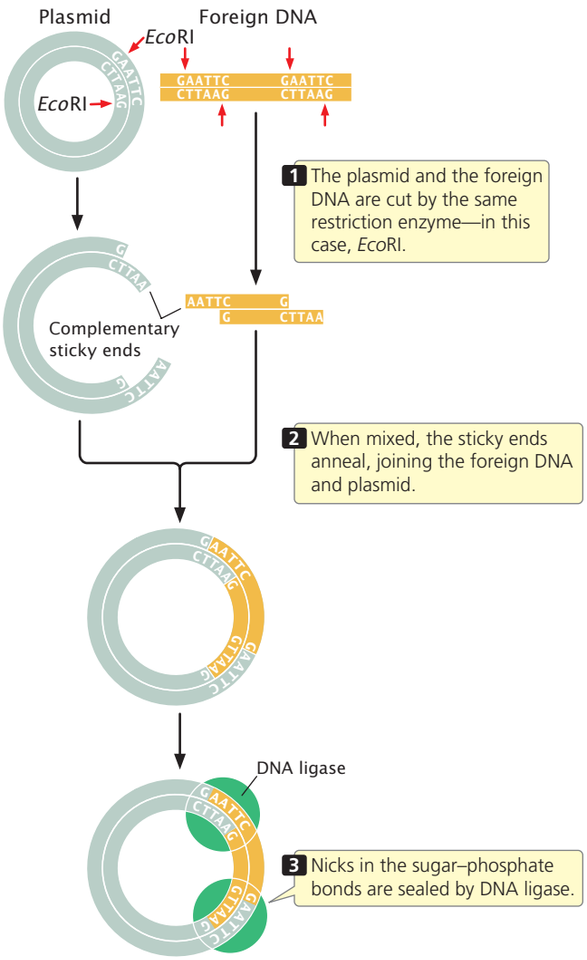
Sometimes restriction sites are not available at a site where the DNA needs to be cut. In that case, a restriction site can be created with the use of **linkers,** which are small, synthetic DNA fragments that contain one or more restriction sites. Linkers can be attached to the ends of any piece of DNA and are then cut by a restriction enzyme, generating sticky ends that are complementary to sticky ends on the plasmid.



19.5 An idealized cloning vector has an origin of replication, one or more selectable markers, and one or more unique restriction sites.



19.6 The pUC19 plasmid is a typical cloning vector. It contains a cluster of unique restriction sites, an origin of replication, and two selectable markers—an ampicillin-resistance gene and a *lacZ* gene.

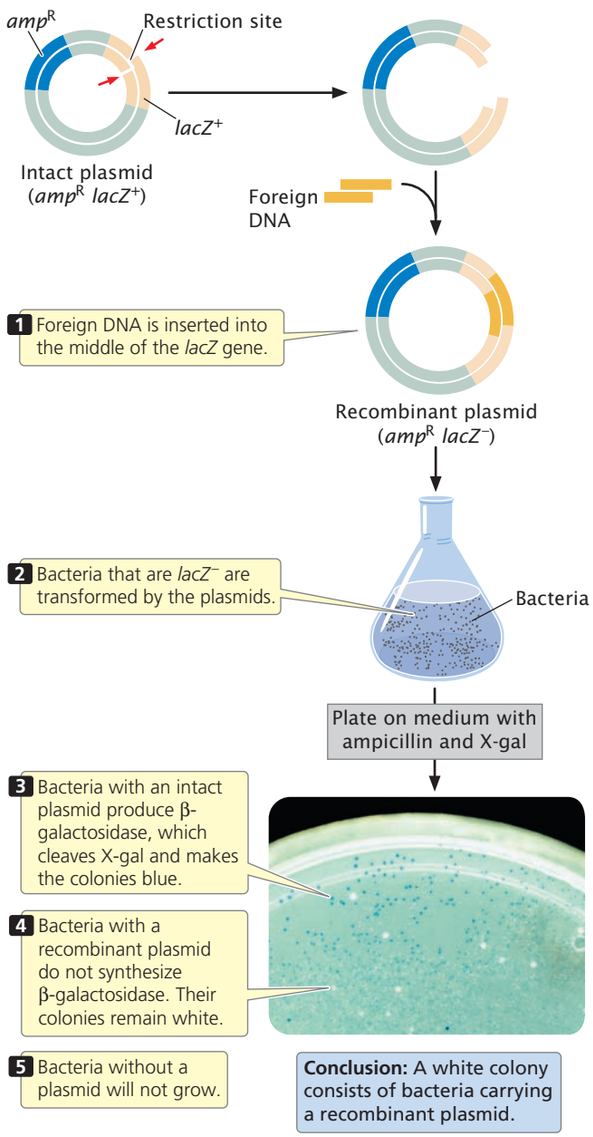


19.7 A foreign DNA fragment can be inserted into a plasmid with the use of restriction enzymes.

Transformation When a gene has been placed inside a plasmid, the plasmid must be introduced into bacterial cells. This task is usually accomplished by *transformation,* which is the capacity of bacterial cells to take up DNA from the external environment (see Chapter 8). Some types of cells undergo transformation naturally; others must be treated chemically or physically before they will undergo transformation. Inside the cell, the plasmids replicate and multiply.

The use of selectable markers Cells bearing recombinant plasmids can be detected by using the selectable markers on the plasmid. One type of selectable marker commonly used with plasmids is a copy of the *lacZ* gene (**Figure 19.8**). The *lacZ* gene contains a series of unique restriction sites into which can be inserted a fragment of DNA to be cloned. In the absence of an inserted fragment, the *lacZ* gene is active and produces β-galactosidase. When foreign DNA is inserted into the restriction site, it disrupts the *lacZ* gene, and β-galactosidase is not produced. The plasmid also usually contains a second selectable marker, which may be a gene that confers resistance to an antibiotic such as ampicillin.

Bacteria that are *lacZ*^{2} are transformed by the plasmids and plated on medium that contains ampicillin. Only cells that have been successfully transformed and contain a plasmid with the ampicillin-resistance gene will survive and grow. Some of these cells will contain an intact plasmid,



19.8 The *lacZ* gene can be used to screen bacteria containing recombinant plasmids. A special plasmid carries a copy of the *lacZ* gene and an ampicillin-resistance gene. *[Photograph: Cytographics/Visuals Unlimited.]*

Table 19.2 Comparison of plasmids, phage lambda vectors, cosmids, and bacterial artificial chromosomes

Cloning Vector	Size of DNA That Can Be Cloned	Method of Propagation	Introduction to Bacteria
Plasmid	As large as 15 kb	Plasmid replication	Transformation
Phage lambda	As large as 23 kb	Phage reproduction	Phage infection
Cosmid	As large as 44 kb	Plasmid reproduction	Phage infection
Bacterial artificial chromosome	As large as 500 kb	Plasmid reproduction	Conjugation

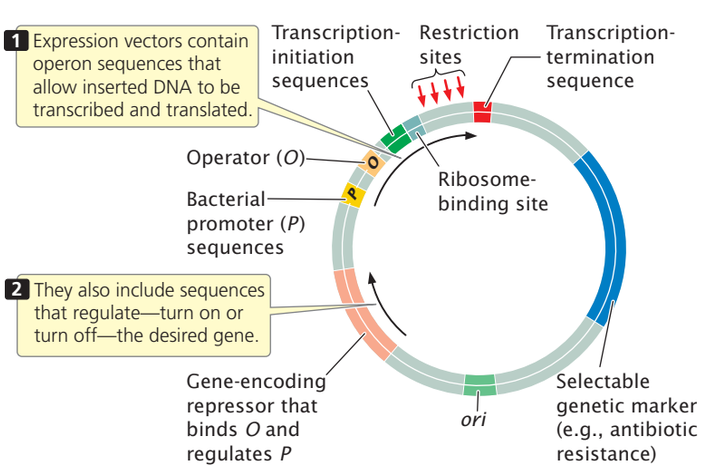
Note: 1 kb = 1000 bp.

whereas others possess a recombinant plasmid. The medium also contains the chemical X-gal, which produces a blue substance when cleaved. Bacterial cells with an intact original plasmid—without an inserted fragment—have a functional *lacZ* gene and can synthesize β-galactosidase, which cleaves X-gal and turns the bacteria blue. Bacterial cells with a recombinant plasmid, however, have a β-galactosidase gene that is disrupted by the inserted DNA; they do not synthesize β-galactosidase and remain white. (In these experiments, the bacterium's own β-galactosidase gene has been inactivated, and so only bacteria with the plasmid turn blue.) Thus, the color of the colony allows quick determination of whether a recombinant or intact plasmid is present in the cell.

Plasmids make ideal cloning vectors but can hold only DNA less than about 15 kb in size. When large DNA fragments are inserted into a plasmid vector, the plasmid becomes unstable.

Bacteriophage vectors A number of other vectors have been developed for cloning larger pieces of DNA in bacteria. For example, bacteriophage λ, which infects *E. coli,* can be used to clone as much as about 23 bp of foreign DNA; it transfers DNA into bacteria cells with high efficiency. **Cosmids** are plasmids that are packaged into empty viral protein coats and transferred to bacteria by viral infection. They can carry more than twice as much foreign DNA as can a phage vector. **Bacterial artificial chromosomes** (BACs) are vectors originally constructed from the F plasmid (a special plasmid that controls mating and the transfer of genetic material in some bacteria; see Chapter 8) and can hold very large fragments of DNA that can be as long as 500,000 bp. **Table 19.2** compares the properties of plasmids, phage λ vectors, cosmids, and BACs.

Sometimes the goal in gene cloning is not just to replicate the gene, but also to produce the protein that it encodes. To ensure transcription and translation, a foreign gene is usually inserted into an **expression vector,** which, in addition to the usual origin of replication, restriction sites, and selectable markers, contains sequences required for transcription and translation in bacterial cells (**Figure 19.9**).



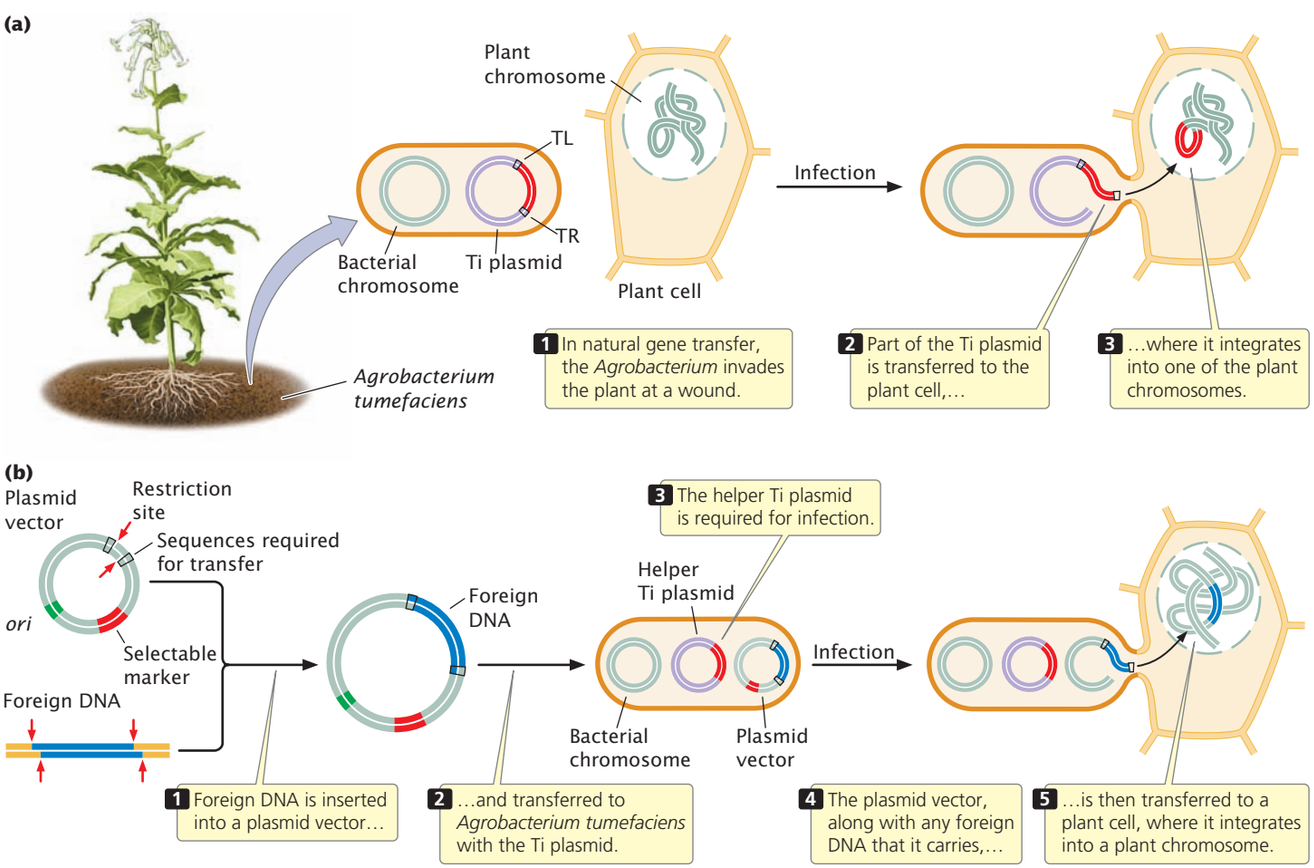
19.9 To ensure transcription and translation, a foreign gene may be inserted into an expression vector—in this example, an *E. coli* expression vector.

Although manipulating genes in bacteria is simple and efficient, the goal may be to transfer a gene into eukaryotic cells. For example, it might be desirable to transfer a gene conferring herbicide resistance into a crop plant or to transfer a gene for clotting factor into a person suffering from hemophilia. Many eukaryotic proteins are modified after translation (e.g., sugar groups may be added). Such modifications are essential for proper function, but bacteria do not have the capacity to carry out the modification; thus a functional protein can be produced only in a eukaryotic cell.

A number of cloning vectors have been developed that allow the insertion of genes into eukaryotic cells. Special plasmids have been developed for cloning in yeast, and retroviral vectors have been developed for cloning in mammals. A **yeast artificial chromosome** (YAC) is a DNA molecule that has a yeast origin of replication, a pair of telomeres, and a centromere. These features ensure that YACs are stable, replicate, and segregate in the same way as yeast chromosomes. YACs are particularly useful because they can carry DNA fragments as large as 600 kb, and some special YACs can carry inserts of more than 1000 kb. YACs have been modified so that they can be used in eukaryotic organisms other than yeast.

The soil bacterium *Agrobacterium tumefaciens,* which invades plants through wounds and induces crown galls (tumors), has been used to transfer genes to plants. This bacterium contains a large plasmid called the **Ti plasmid,** part of which is transferred to a plant cell when *A. tumefaciens* infects a plant. In the plant, part of the Ti plasmid DNA integrates into one of the plant chromosomes, where it is transcribed and translated to produce several enzymes that help support the bacterium (**Figure 19.10a**).

Geneticists have engineered a vector that contains the flanking sequences required to transfer DNA, a selectable marker, and restriction sites into which foreign DNA can be inserted (**Figure 19.10b**). When placed in *A. tumefaciens* with the Ti plasmid, this vector will transfer the foreign DNA that it carries into a plant cell, where it will integrate into a plant chromosome. This vector has been used to transfer genes that confer economically significant attributes such as resistances to herbicides, plant viruses, and insect pests.



19.10 The Ti plasmid can be used to transfer genes into plants. Flanking sequences TL and TR are required for the transfer of the DNA segment from bacteria to the plant cell.

CONCEPTS

DNA fragments can be inserted into cloning vectors, stable pieces of DNA that will replicate within a cell. A cloning vector must have an origin of replication, one or more unique restriction sites, and selectable markers. An expression vector contains sequences that allow a cloned gene to be transcribed and translated. Special cloning vectors have been developed for introducing genes into eukaryotic cells.

✔ CONCEPT CHECK 5

How is a gene inserted into a plasmid cloning vector?

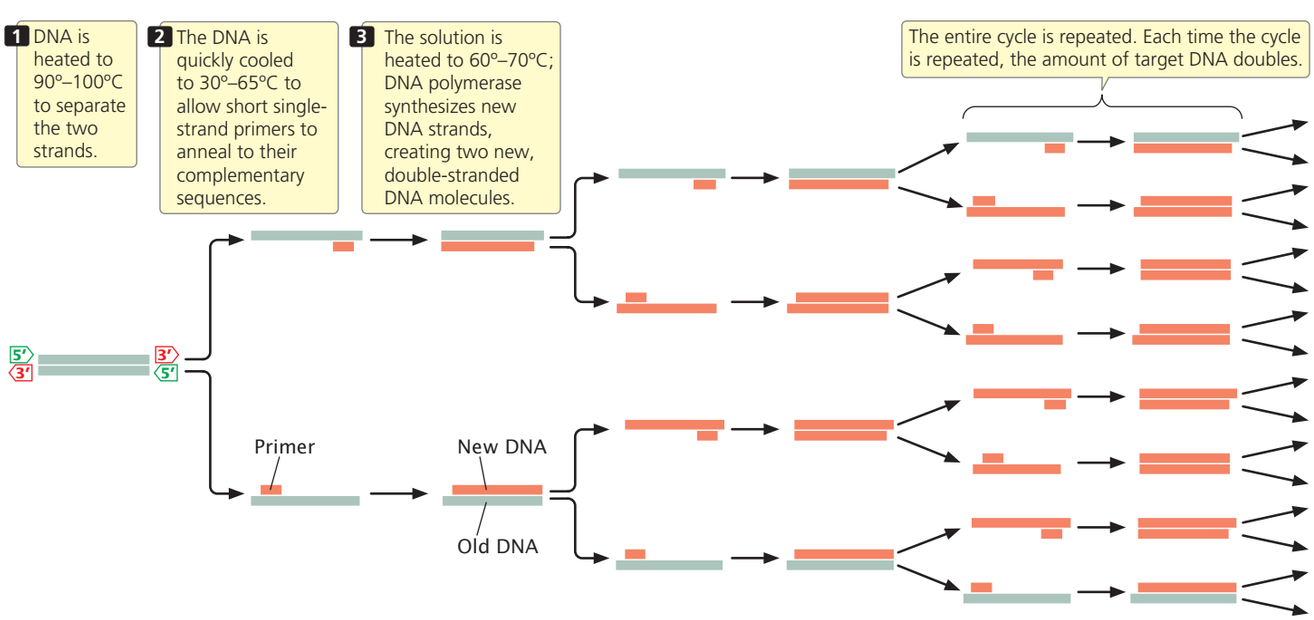
Amplifying DNA Fragments with the Polymerase Chain Reaction

The manipulation and analysis of genes require multiple copies of the DNA sequences used. A major problem in working at the molecular level is that each gene is a tiny fraction of the total cellular DNA. Because each gene is rare, it must be isolated and amplified before it can be studied. As already stated, one way to amplify a DNA fragment is to clone it in bacterial cells. Indeed, for many years, gene cloning was the only way to quickly amplify a DNA fragment, and cloning was a prerequisite for many other molecular methods. Cloning is labor intensive and requires at least several days to grow the bacteria. Today, the polymerase chain reaction makes the amplification of short DNA fragments possible without cloning, but cloning is still widely used for amplifying large DNA fragments and for other manipulations of DNA sequences.

The **polymerase chain reaction** (PCR), first developed in 1983, allows DNA fragments to be amplified a billionfold within just a few hours. It can be used with extremely small amounts of original DNA, even a single molecule. The polymerase chain reaction has revolutionized molecular biology and is now one of the most widely used of all molecular techniques. The basis of PCR is replication catalyzed by a DNA polymerase. Replication in this case has two essential requirements: (1) a single-stranded DNA template from which a new DNA strand can be copied and (2) a primer with a 3′-OH group to which new nucleotides can be added.

Because a DNA molecule consists of two nucleotide strands, each of which can serve as a template to produce a new molecule of DNA, the amount of DNA doubles with each replication event. The primers used in PCR to replicate the templates are short fragments of DNA, typically from 17 to 25 nucleotides long, that are complementary to known sequences on the template.

To carry out PCR, one begins with a solution that includes the target DNA, DNA polymerase, all four deoxyribonucleoside triphosphates (dNTPs—the substrates for DNA polymerase), primers, and magnesium ions and other salts that are necessary for the reaction to proceed. A typical polymerase chain reaction includes three steps (**Figure 19.11**). In step 1, a starting solution of DNA is heated to between 90° and 100°C to break the hydrogen bonds between the strands and thus produce the necessary single-stranded templates. The reaction mixture is held at this temperature for only a minute or two. In step 2, the DNA solution is cooled quickly to between 30° and 65°C and held at this temperature for a minute or less. During this short interval, the DNA strands



19.11 The polymerase chain reaction is used to amplify even very small samples of DNA.

will not have a chance to reanneal, but the primers will be able to attach to the template strands. In step 3, the solution is heated for a minute or less to between 60° and 70°C, the temperature at which DNA polymerase can synthesize new DNA strands. At the end of the cycle, two new double-stranded DNA molecules are produced for each original molecule of target DNA.

The whole cycle is then repeated. With each cycle, the amount of target DNA doubles; so the target DNA increases geometrically. One molecule of DNA increases to more than 1000 molecules in 10 PCR cycles, to more than 1 million molecules in 20 cycles, and to more than 1 billion molecules in 30 cycles. Each cycle is completed within a few minutes; so a large amplification of DNA can be achieved within a few hours.

Two key innovations facilitated the use of PCR in the laboratory. The first was the discovery of a DNA polymerase that is stable at the high temperatures used in step 1 of PCR. The DNA polymerase from *E. coli* that was originally used in PCR denatures at 90°C. For this reason, fresh enzyme had to be added to the reaction mixture in *each* cycle, slowing the process considerably. This obstacle was overcome when DNA polymerase was isolated from the bacterium *Thermus aquaticus,* which lives in the boiling springs of Yellowstone National Park. This enzyme, dubbed ***Taq* polymerase,** is remarkably stable at high temperatures and is not denatured during the strand-separation step of PCR; so it can be added to the reaction mixture at the beginning of the PCR process and will continue to function through many cycles.

The second key innovation was the development of automated thermal cyclers—machines that bring about the rapid temperature changes necessary for the different steps of PCR. Originally, tubes containing reaction mixtures were moved by hand among water baths set at the different temperatures required for the three steps of each cycle. In automated thermal cyclers, the reaction tubes are placed in a metal block that changes temperature rapidly according to a computer program.

In addition to amplifying DNA, PCR can be used to amplify sequences corresponding to RNA. This amplification is accomplished by first converting RNA into complementary DNA (cDNA) with the use of the enzyme reverse transcriptase. The cDNA can then be amplified by the usual PCR reaction. This technique is referred to as **reverse-transcription PCR.**

Limitations of PCR The polymerase chain reaction is now often used in place of gene cloning, but it has several limitations. First, the use of PCR requires prior knowledge of at least part of the sequence of the target DNA to allow the construction of the primers. Second, the capacity of PCR to amplify extremely small amounts of DNA makes contamination a significant problem. Minute amounts of DNA from the skin of laboratory workers and even in small particles in the air can enter a reaction tube and be amplified along with the target DNA. Careful laboratory technique and the use of controls are necessary to circumvent this problem.

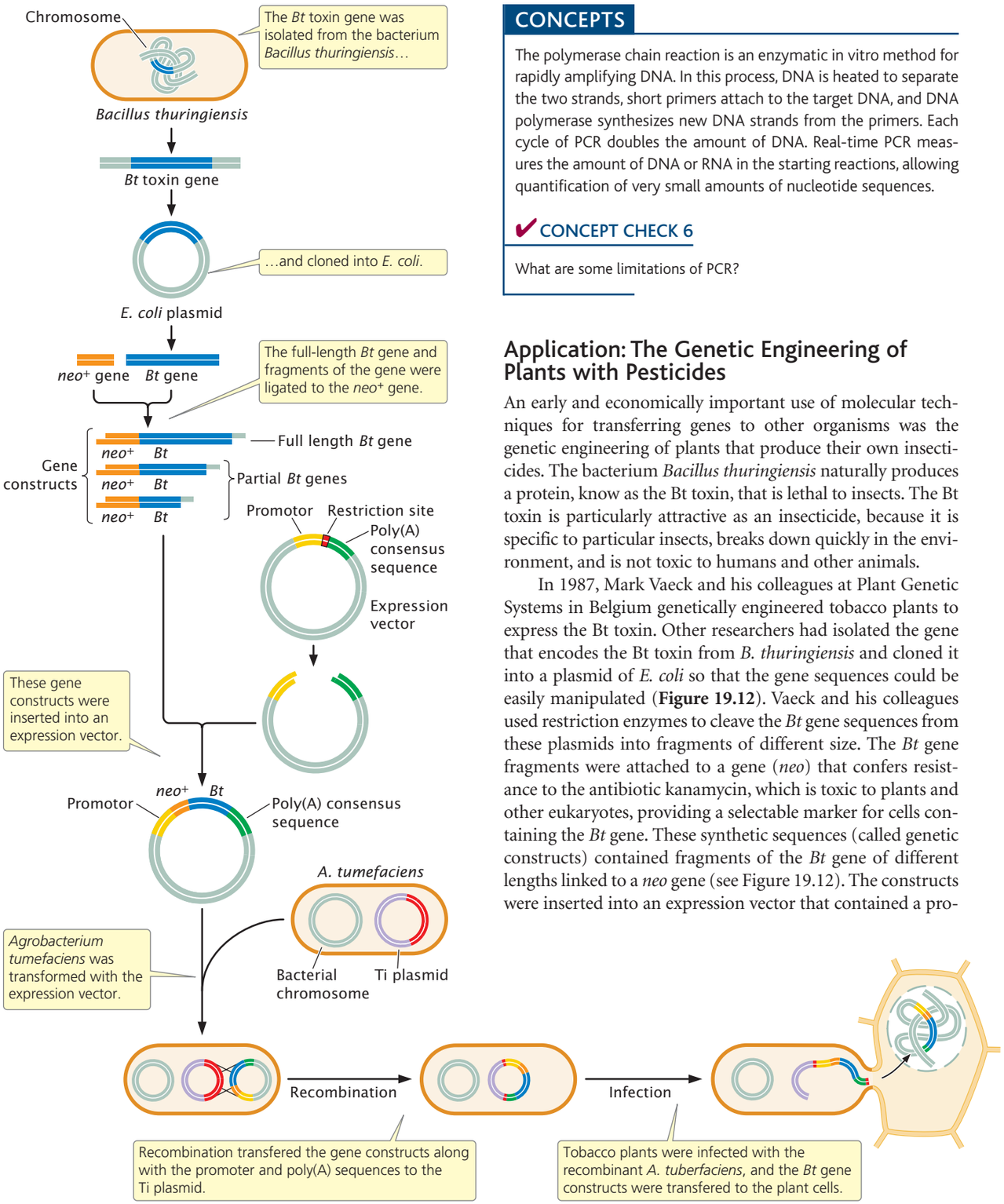
A third limitation of PCR is accuracy. Unlike other DNA polymerases, *Taq* polymerase does not have the capacity to proofread (see p. xxx in Chapter 12) and, under standard PCR conditions, it incorporates an incorrect nucleotide about once every 20,000 bp. New heat-stable DNA polymerases with proofreading capacity have been isolated, giving more accurate PCR results.

A fourth limitation of PCR is that the size of the fragments that can be amplified by standard *Taq* polymerase is usually less than 2000 bp. By using a combination of *Taq* polymerase and a DNA polymerase with proofreading capacity and by modifying the reaction conditions, investigators have been successful in extending PCR amplification to larger fragments, but even these larger fragments are limited in length to 50,000 bp or smaller. In spite of its limitations, PCR is used routinely in a wide array of molecular applications.

PCR as a diagnostic tool Because the primers used in PCR are specific for known DNA sequences, PCR can be used to detect the presence of a particular DNA sequence in a sample. For example, PCR is often used to detect the presence of viruses in blood samples by running the reaction with primers complementary to known viral DNA sequences. If viral DNA is present, the primers will attach to it, and PCR will amplify a DNA fragment of known length. The presence of a DNA fragment of the proper length on a gel indicates the presence of viral DNA in the blood sample. Modern diagnostic tests for infection with HIV, the causative agent of AIDS (see Chapter 8), utilize this type of PCR amplification of HIV sequences.

Real-time PCR A modification of PCR, known as **real-time PCR,** can be used to quantitatively determine the amount of starting nucleic acid. In this procedure, the usual PCR reaction is used to amplify a specific DNA fragment, and a sensitive instrument is used to accurately determine the amount of DNA that is present in solution after each PCR cycle. A probe that fluoresces and is specific to the DNA sequence of interest is frequently used in the reaction, and so only the DNA of interest is measured. The technique is called real-time PCR because the amount of DNA amplified is measured as the reaction proceeds.

Often, real-time PCR is combined with reverse-transcription PCR to measure the amount of mRNA in a sample, allowing biologists to determine the level of gene expression in different cells and under different conditions. For example, researchers interested in how gene expression changes in response to the administration of a drug often use real-time PCR to quantitatively measure the amount of mRNA produced by specific genes in cells exposed to the drug and compare it with the amount of mRNA produced by the genes in control cells with no drug exposure.



19.12 The *Bt* toxin gene, which encodes an insecticide, was isolated from bacteria and transferred to tobacco plants.

CONCEPTS

The polymerase chain reaction is an enzymatic in vitro method for rapidly amplifying DNA. In this process, DNA is heated to separate the two strands, short primers attach to the target DNA, and DNA polymerase synthesizes new DNA strands from the primers. Each cycle of PCR doubles the amount of DNA. Real-time PCR measures the amount of DNA or RNA in the starting reactions, allowing quantification of very small amounts of nucleotide sequences.

✔ CONCEPT CHECK 6

What are some limitations of PCR?

Application: The Genetic Engineering of Plants with Pesticides

An early and economically important use of molecular techniques for transferring genes to other organisms was the genetic engineering of plants that produce their own insecticides. The bacterium *Bacillus thuringiensis* naturally produces a protein, know as the Bt toxin, that is lethal to insects. The Bt toxin is particularly attractive as an insecticide, because it is specific to particular insects, breaks down quickly in the environment, and is not toxic to humans and other animals.

In 1987, Mark Vaeck and his colleagues at Plant Genetic Systems in Belgium genetically engineered tobacco plants to express the Bt toxin. Other researchers had isolated the gene that encodes the Bt toxin from *B. thuringiensis* and cloned it into a plasmid of *E. coli* so that the gene sequences could be easily manipulated (**Figure 19.12**). Vaeck and his colleagues used restriction enzymes to cleave the *Bt* gene sequences from these plasmids into fragments of different size. The *Bt* gene fragments were attached to a gene (*neo*) that confers resistance to the antibiotic kanamycin, which is toxic to plants and other eukaryotes, providing a selectable marker for cells containing the *Bt* gene. These synthetic sequences (called genetic constructs) contained fragments of the *Bt* gene of different lengths linked to a *neo* gene (see Figure 19.12). The constructs were inserted into an expression vector that contained a pro-

19.13 Transgenic crops are broadly cultivated throughout the world. The rapeseed shown here has been genetically modified to resist a specific herbicide, which can then be sprayed on the field to kill weeds. *[Chris Knapton/Photo Researchers.]*

moter, to ensure that the introduced sequences would be transcribed, and poly(A) consensus sequences, to ensure that the mRNAs produced from the *Bt* and *neo* genes would be processed properly and translated in the plant cells.

Agrobacterium tumefaciens bacteria were then transformed with the expression plasmids. After they were inside the *Agrobaceterium,* sequences on the vector recombined with sequences on a Ti plasmid, transferring the gene constructs to the Ti plasmid. Small discs of leaves from a tobacco plant were infected with the genetically engineered *Agrobacterium,* which transferred the Ti plasmids into the plant cells. Whole tobacco plants were regenerated from the leaf disks and selected for resistance to kanamycin.

The resulting transgenic plants were then tested for resistance to insect pests. Leaves of the plants were fed to tobacco hornworms (an important pest of tobacco) and mortality rates of the hornworms were monitored. High mortality of hornworms was observed in about two-thirds of the plants containing fragments of the *Bt* gene. Interestingly, none of the plants containing the full-length *Bt* gene produced any insect-killing toxins; apparently, the plant cells were better able to translate the fragments of the *Bt* gene than to translate the entire gene. The transgenic plants producing Bt toxin grew normally. They were interbred to produce F_1 plants, which also exhibited insect and kanamycin resistance, indicating that the introduced genes were stably integrated into the plant chromosomes.

Other researchers using similar methods subsequently introduced the gene for Bt toxin into cotton, tomatoes, corn and other plants (**Figure 19.13**). Today, transgenic plants expressing the Bt toxin are used broadly throughout the world.

19.3 Molecular Techniques Can Be Used to Find Genes of Interest

To analyze a gene or to transfer it to another organism, the gene must first be located and isolated. For instance, if we wanted to transfer a human gene for growth hormone to bacteria, we must first find the human gene that encodes growth hormone and separate it from the 3.2 billion bp of human DNA. In our consideration of gene cloning, we've glossed over the problem of finding the DNA sequence to be cloned. A discussion of how to solve this problem has been purposely delayed until now because, paradoxically, one must often clone a gene to find it.

This approach—to clone first and search later—is called "shotgun cloning," because it is like hunting with a shotgun: one sprays shots widely in the general direction of the quarry, knowing that there is a good chance that one or more of the pellets will hit the intended target. In shotgun cloning, one first clones a large number of DNA fragments, knowing that one or more contains the DNA of interest, and then searches for the fragment of interest among the clones.

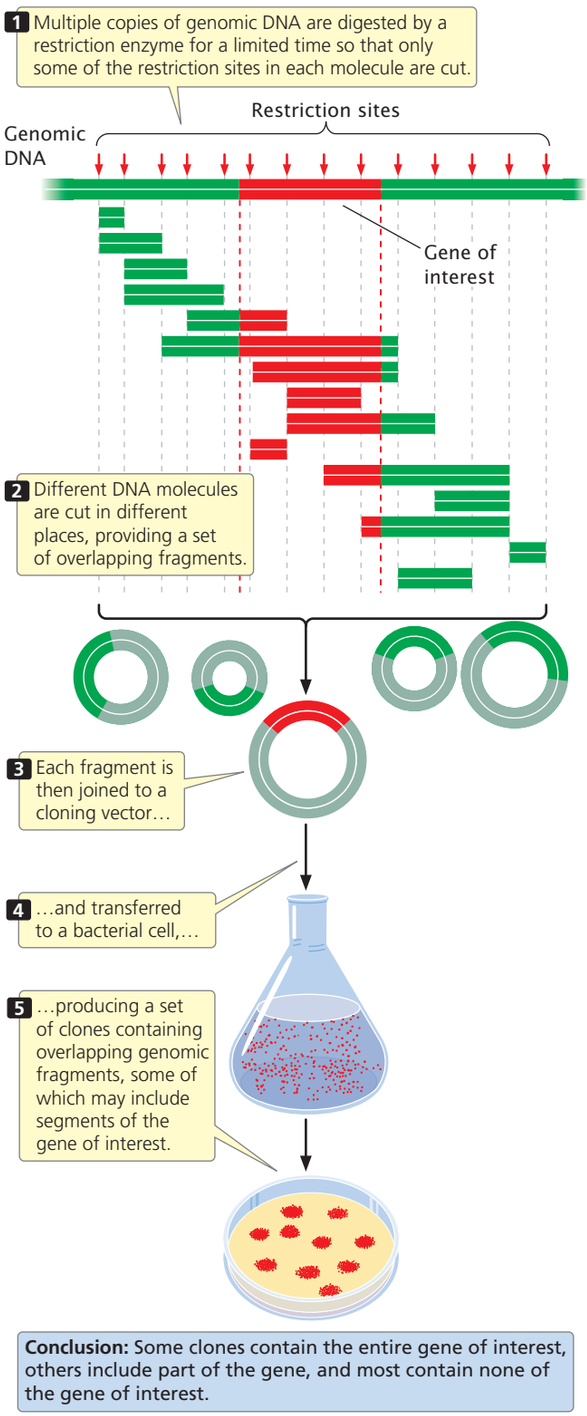
Gene Libraries

A collection of clones containing all the DNA fragments from one source is called a **DNA library.** For example, we might isolate genomic DNA from human cells, break it into fragments, and clone all of them in bacterial cells or phages. The set of bacterial colonies or phages containing these fragments is a human **genomic library,** containing all the DNA sequences found in the human genome.

Creating a genomic library To create a genomic library, cells are collected and disrupted, which causes them to release their DNA and other cellular contents into an aqueous solution, and the DNA is extracted from the solution. After the DNA has been isolated, it is cut into fragments by a restriction enzyme for a limited amount of time only (a partial digestion) so that only *some* of the restriction sites in each DNA molecule are cut. Because which sites are cut is random, different DNA molecules will be cut in different places, and a set of overlapping fragments will be produced (**Figure 19.14**). The fragments are then joined to vectors, which can be transferred to bacteria. A few of the clones contain the entire gene of interest, a few contain parts of the gene, but most contain fragments that have no part of the gene of interest.

A genomic library must contain a large number of clones to ensure that all DNA sequences in the genome are represented in the library. A library of the human genome formed by using cosmids, each carrying a random DNA fragment from 35,000 to 44,000 bp long, would require about 350,000 cosmid clones to provide a 99% chance that every sequence is included in the library.

Creating a cDNA library An alternative to creating a genomic library is to create a library consisting only of those DNA sequences that are transcribed into mRNA (called a **cDNA library** because all the DNA in this library is *complementary* to mRNA). Much of eukaryotic DNA consists of



19.14 A genomic library contains all of the DNA sequences found in an organism's genome.

repetitive (and other DNA) sequences that are not transcribed into mRNA (see p. xxx in Chapter 11), and these sequences are not represented in a cDNA library.

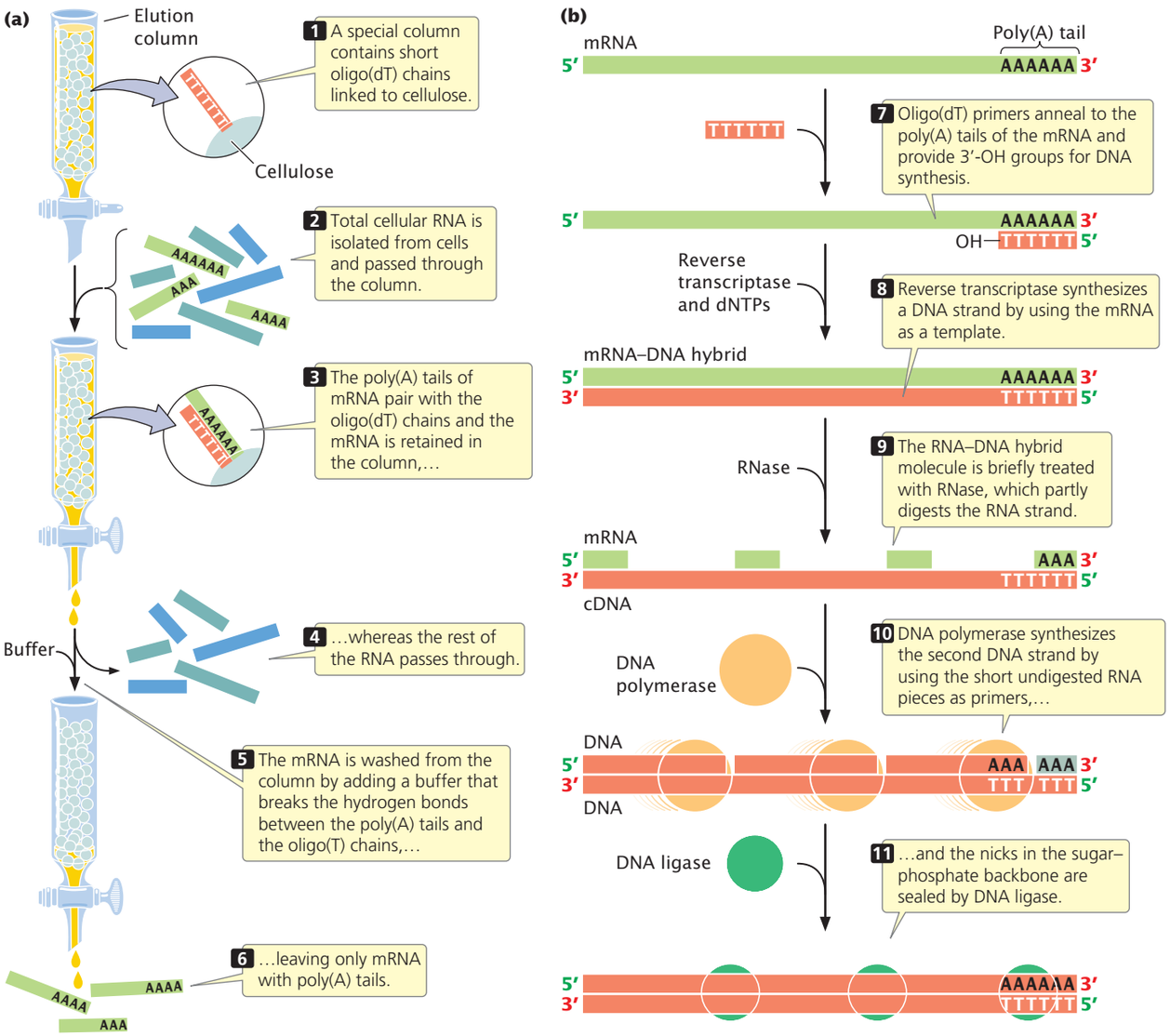
A cDNA library has two additional advantages. First, it is enriched with fragments from actively transcribed genes. Second, introns do not interrupt the cloned sequences; introns would pose a problem when the goal is to produce a eukaryotic protein in bacteria, because most bacteria have no means of removing the introns.

The disadvantage of a cDNA library is that it contains only sequences that are present in mature mRNA. Sometimes, researchers are interested in sequences that are not transcribed, such as those in promoters and enhancers, which are important for transcription but are not themselves transcribed. These sequences are not present in a cDNA library. Furthermore, a cDNA library contains only those gene sequences expressed in the tissue from which the RNA was isolated, and the frequency of a particular DNA sequence in a cDNA library depends on the abundance of the corresponding mRNA in the given tissue. So, if a particular gene is not expressed or is expressed only at low frequency in a particular tissue, it may be absent in a cDNA library prepared from that tissue. In contrast, almost all genes are present at the same frequency in a genomic DNA library.

To create a cDNA library, messenger RNA must first be separated from other types of cellular RNA (tRNA, rRNA, snRNA, etc.). Most eukaryotic mRNAs possess a poly(A) tail, which provides a convenient hook for separating eukaryotic mRNA from the other types. Total cellular RNA is isolated from cells and poured through a column packed with short fragments of DNA consisting entirely of thymine nucleotides—that is, oligo(dT) chains (**Figure 19.15a**). As the RNA moves through the column, the poly(A) tails of mRNA molecules pair with the oligo(dT) chains and are retained in the column, whereas the rest of the RNA passes through it. The mRNA can then be washed from the column by the addition of a buffer that breaks the hydrogen bonds between poly(A) tails and oligo(dT) chains.

The mRNA molecules are then copied into cDNA by reverse transcription. Short oligo(dT) primers are added to the mRNA, which pair with the poly(A) tail at the 3′ end of an mRNA molecule, providing a 3′-OH group for the initiation of DNA synthesis (**Figure 19.15b**). Reverse transcriptase, an enzyme isolated from retroviruses (see p. xxx in Chapter 8), synthesizes single-stranded complementary DNA from the RNA template by adding DNA nucleotides to the 3′-OH group of the primer.

The resulting RNA–DNA hybrid molecule is then converted into a double-stranded cDNA molecule by one of several methods. A common method is to treat the RNA–DNA hybrid with RNase to partly digest the RNA strand. Partial digestion leaves gaps in the RNA–DNA hybrid, allowing DNA polymerase to synthesize a second DNA strand by



19.15 A cDNA library contains only those DNA sequences that are transcribed into mRNA.

using the short undigested RNA pieces as primers and the first DNA strand as a template. DNA polymerase eventually displaces all the RNA fragments, replacing them with DNA nucleotides, and nicks in the sugar–phosphate backbone are sealed by DNA ligase.

CONCEPTS

One method of finding a gene is to create and screen a DNA library. A genomic library is created by cutting genomic DNA into overlapping fragments and cloning each fragment in a separate bacterial cell. A cDNA library is created from mRNA that is converted into cDNA and cloned in bacteria.

✔ CONCEPT CHECK 7

What are some of the advantages and disadvantages of a cDNA library over a genomic library?

Screening DNA libraries Creating a genomic or cDNA library is relatively easy compared with screening the library to find clones that contain the gene of interest. The screening procedure used depends on what is known about the gene.

The first step in screening is to plate the clones of the library. If a plasmid or cosmid vector was used to construct

the library, the cells are diluted and plated so that each bacterium grows into a distinct colony. If a phage vector was used, the phages are allowed to infect a lawn of bacteria on a petri plate. Each plaque or bacterial colony contains a single, cloned DNA fragment that must be screened for the gene of interest.

A common way to screen libraries is with probes. To use a probe, replicas of the plated colonies or plaques in the library must first be made. **Figure 19.16** illustrates this procedure for a cosmid library.

How is a probe obtained when the gene has not yet been isolated? One option is to use a similar gene from another organism as the probe. For example, if we wanted to screen a human genomic library for the growth-hormone gene and the gene had already been isolated from rats, we could use a purified rat-gene sequence as the probe to find the human gene for growth hormone. Successful hybridization does not require perfect complementarity between the probe and the target sequence; so a related sequence can often be used as a probe.

Alternatively, synthetic probes can be created if the protein produced by the gene has been isolated and its amino acid sequence has been determined. With the use of the genetic code and the amino acid sequence of the protein, possible nucleotide sequences of a small region of the gene can be deduced. Although only one sequence in the gene encodes a particular protein, the presence of synonymous codons means that the same protein could be produced by several different nucleotide sequences, and it is impossible to know which is correct. To overcome this problem, a mixture of all the possible nucleotide sequences is used as a probe. To minimize the number of sequences required in the mixture, a region of the protein is selected with relatively little degeneracy in its codons.

When part of the DNA sequence of the gene has been determined, a set of DNA probes can be synthesized chemically by using an automated machine known as an oligonucleotide synthesizer. The resulting probes can be used to screen a library for a gene of interest.

Yet another method of screening a library is to look for the protein product of a gene. This method requires that the DNA library be cloned in an expression vector. The clones can be tested for the presence of the protein by using an antibody that recognizes the protein or by using a chemical test for the protein product. This method depends on the existence of a test for the protein produced by the gene.

Almost any method used to screen a library will identify several clones, some of which will be false positives that do not contain the gene of interest; several screening methods may be needed to determine which clones actually contain the gene.

CONCEPTS

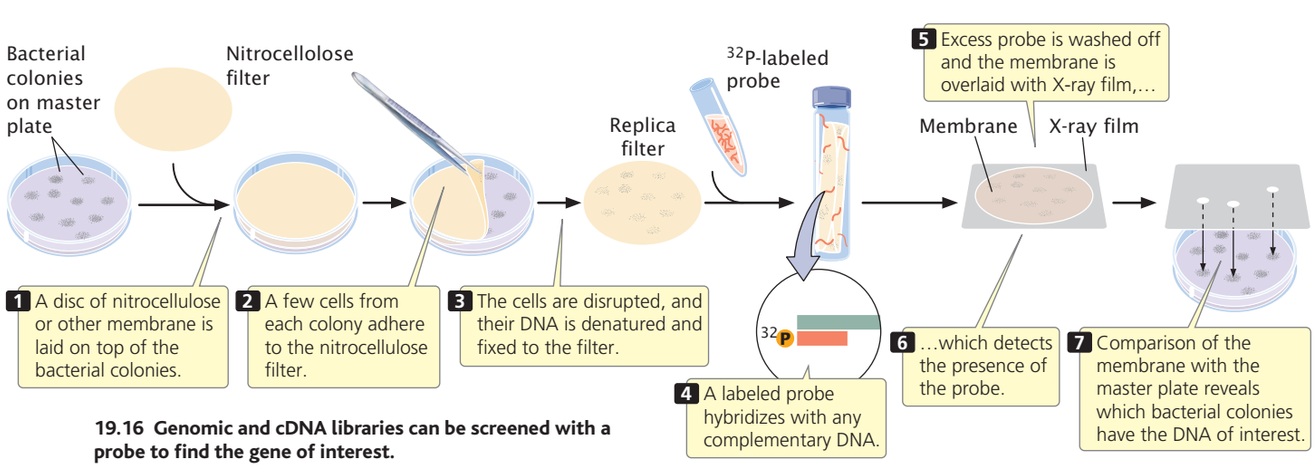
A DNA library can be screened for a specific gene by using complementary probes that hybridize to the gene. Alternatively, the library can be cloned into an expression vector, and the gene can be located by examining the clones for the protein product of the gene.

✔ CONCEPT CHECK 8

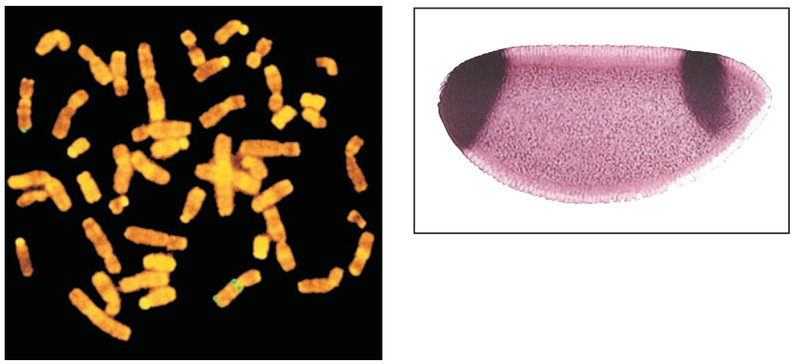
Briefly explain how synthetic probes are created to screen a DNA library when the protein encoded by the gene is known.

In Situ Hybridization

DNA probes can be used to determine the chromosomal location of a gene in a process called **in situ hybridization.** The name is derived from the fact that DNA (or RNA) is visualized while it is in the cell (in situ). This technique requires that the cells be fixed and the chromosomes be spread on a microscope slide and denatured. A labeled probe is then applied to the slide, just as it is can be applied to a gel. Many probes carry attached fluorescent dyes that



19.16 Genomic and cDNA libraries can be screened with a probe to find the gene of interest.



19.17 With in situ hybridization, DNA probes are used to determine the cellular or chromosomal location of a gene or its product. (a) A probe with green fluorescence is specific to the X chromosome, revealing three X chromosomes. A red fluorescent probe reveals one Y chromosome. (b) In situ hybridization is used to detect the presence of mRNA from the tailless gene in a *Drosophila* embryo. *[Part a: Addenbrookes Hospital/Photo Researchers. Part b: Courtesy of L. Tsuda.]*

can be seen directly with the microscope (**Figure 19.17a**). Several probes with different colored dyes can be used simultaneously to investigate different sequences or chromosomes. Fluorescence in situ hybridization (FISH) has been widely used to identify the chromosomal location of human genes.

In situ hybridization can also be used to determine the tissue distribution of specific mRNA molecules, serving as a source of insight into how gene expression differs among cell types (**Figure 19.17b**). A labeled DNA probe complementary to a specific mRNA molecule is added to tissue, and the location of the probe is determined with the use of either autoradiography or fluorescent tags. Determining where a gene is expressed often helps define its function. For example, finding that a gene is highly expressed only in brain tissue might suggest that the gene has a role in neural function.

Positional Cloning

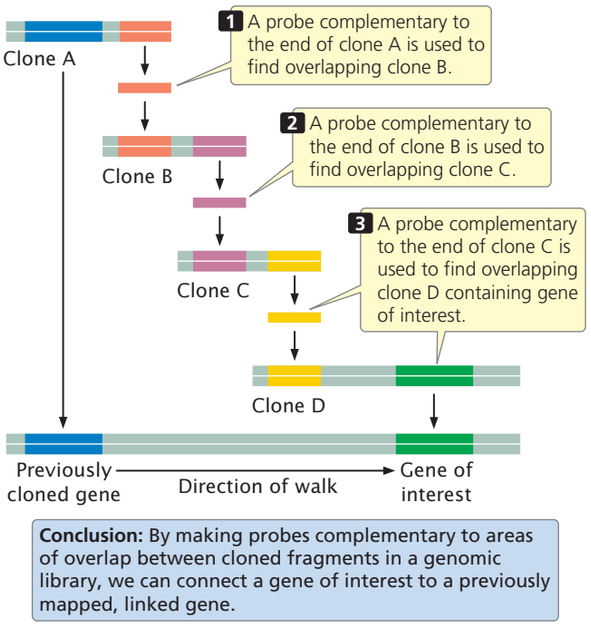
For many genes with important functions, no associated protein product is yet known. The biochemical bases of many human genetic diseases, for example, are still unknown. How can these genes be isolated? One approach is to first determine the general location of the gene on the chromosome by using recombination frequencies derived from crosses or pedigrees (see Chapter 7). After the chromosomal region where the gene is found has been identified, any genes in this region can be cloned and identified. Then other techniques can be used to identify which of the "candidate" genes might be the one that causes the disease. This approach—to isolate genes on the basis of their position on a gene map—is called **positional cloning.**

The development of molecular techniques for identifying gene sequences, such as restriction sites and sequenced DNA fragments, has provided a large number of gene loci that can be used as markers in mapping. Detailed genetic maps of these markers have been created so that their precise location is known. In the first step of positional cloning, geneticists use mapping studies (see Chapter 7) to establish linkage between molecular markers and a phenotype of interest, such as a human disease or a desirable physical trait in a plant or animal. In regard to a human disease, members of one or more families with the disease would be genotyped for a large number of variable genetic markers from each of the human chromosomes. Demonstration of linkage between the disease phenotype and one or more molecular markers would provide information about which chromosome carries the disease locus and its general location on that chromosome.

The next step is to more precisely locate the locus by using additional molecular markers clustered in the chromosomal region where the locus resides. The goal is to identify molecular markers that are within 1 map unit (1% recombination) of the disease locus. In humans, 1% recombination represents about 1,000,000 bp of DNA; so the region in which the gene is located is still large.

After the gene has been placed on a chromosome map, clones that cover the region can be isolated from a genomic library. With the use of a technique called **chromosome walking** (**Figure 19.18**), it is possible to progress from neighboring genes to linked clones, one of which might contain the gene of interest. The basis of chromosome walking is the fact that a genomic library consists of a set of *overlapping* DNA fragments (see Figure 19.14). We start with a cloned gene marker that is close to the new gene of interest so that the "walk" will be as short as possible. One end of the clone of a neighboring marker (clone A in Figure 19.18) is used to make a complementary probe. This probe is used to screen the genomic library to find a second clone (clone B) that overlaps with the first and extends in the direction of the gene of interest. This second clone is isolated and purified and a probe is prepared from its end. The second probe is used to screen the library for a third clone (clone C) that overlaps with the second. In this way, one can systematically "walk" toward the gene of interest, one clone at a time.

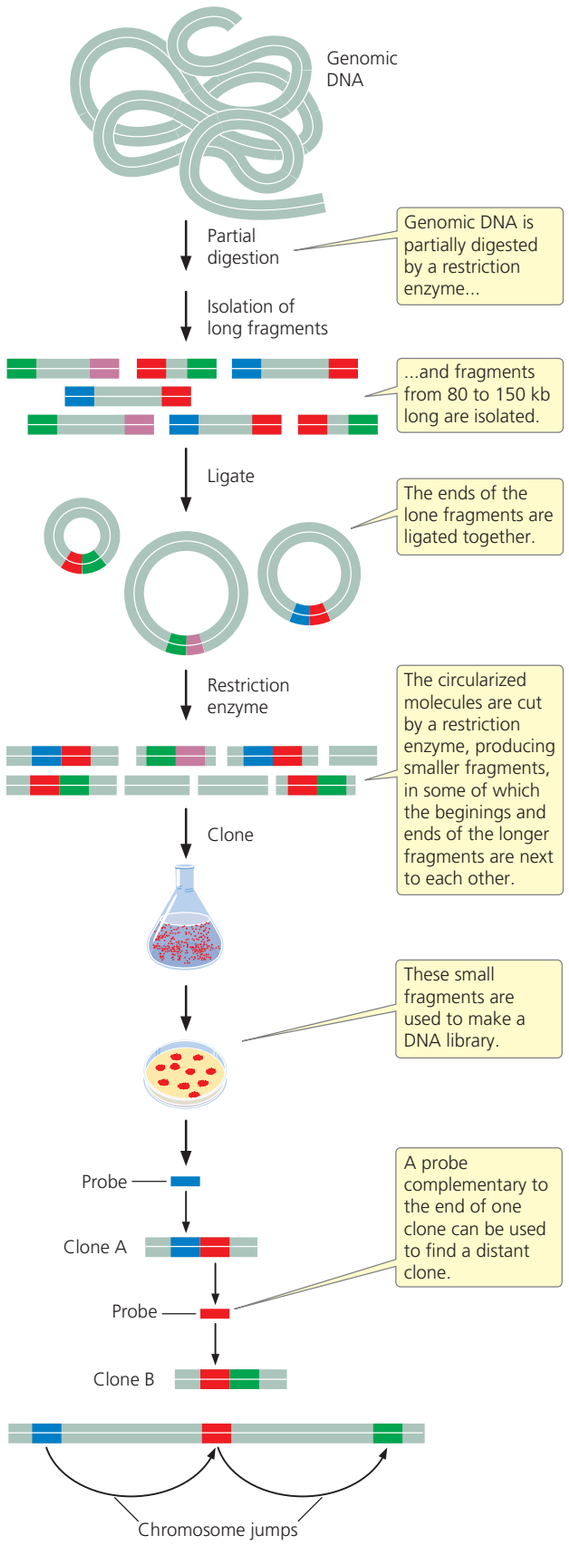
If the region in which the gene of interest lies is very large, chromosome walking can be a laborious task. To cover the region more quickly, geneticists sometimes turn to **chromosome jumping.** In this technique, genomic DNA is cut



19.18 In chromosome walking, neighboring genes are used to locate a gene of interest.

into fragments by partial digestion with a restriction enzyme, and large fragments between 80 kb and 150 kb are isolated (**Figure 19.19**). The ends of the resulting DNA fragments are ligated together so that they create a circular DNA molecule. In each circle, the beginning and end of the long sequence will be brought close together. The circle is then cut with another restriction enzyme that cuts each circle into smaller fragments. Some of these fragments contain the beginnings and the ends of the original large fragments without the intervening DNA. These fragments are used to create a DNA library. Because the cloned fragments begin and end with sequences that are distantly linked, the DNA at the end of one fragment can then be used to make a probe that can find clones that are distantly linked. New probes are then made to the linked clone and used to jump again. In this way, one is able to "jump" to clones that are more distantly linked to a marker gene.

After clones that cover the delineated region have been obtained by chromosome walking or jumping, all genes located within the region are identified. Genes can be distinguished from other sequences by the presence of characteristic features, such as consensus sequences in the promoter and start codon and stop codons within the same reading frame. After potential "candidate" genes have been identified, they can be evaluated to determine which is most likely to be the gene of interest. The expression pattern of the gene—where and when it is transcribed—can often provide clues about its function. For example, genes for neurological disease would likely be expressed in the brain. Geneticists



19.19 Chromosome jumping is used to located distantly linked clones.

often look in the coding region of the gene for mutations among people with the disease. More will be said about determining the function of genes in sections that follow and in Chapter 20.

Through positional cloning, one can identify genes that encode a phenotype without a detailed understanding of the underlying biochemical nature of the phenotype. A number of important human diseases have been identified through positional cloning.

CONCEPTS

Positional cloning allows a gene to be isolated without knowledge of it biochemical basis. Linkage studies are used to map the locus producing a phenotype of interest to a particular chromosome region. Chromosome walking and jumping are used to progress from molecular makers to clones containing sequences that cover the chromosome region. Candidate genes within the region are then evaluated to determine if they encode the phenotype of interest.

✔ CONCEPT CHECK 9

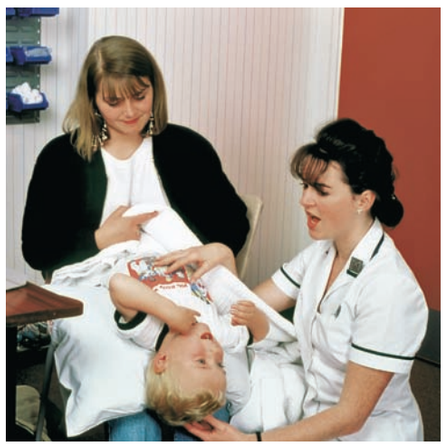
How are candidate genes that are identified by positional cloning evaluated to determine whether they encode the phenotype of interest?

In Silico Gene Discovery

The complete genomes of more and more species are sequenced each year (see Chapter 20), and partial sequences of many other organisms are continually being added to DNA databases. With this growth in sequence information, the task of finding genes is now often carried out by using high-speed computers that analyze and search DNA databases rather than using gene cloning and DNA libraries. Sometimes called "in silico," these methods of locating and characterizing genes rely on the identification of characteristic sequences associated with genes and on comparisons with sequences of known genes in DNA databases. These methods will be discussed in more detail in Chapter 20.

Application: Isolating the Gene for Cystic Fibrosis

The first gene responsible for a human genetic disease that was isolated entirely by positional cloning was the gene for cystic fibrosis. Cystic fibrosis (CF) is an autosomal recessive disorder characterized by chronic lung infections, insufficient production of pancreatic enzymes that are necessary for digestion, and increased salt concentration in sweat (**Figure 19.20**). It is one of the most common genetic diseases in Caucasians, occurring with a frequency of about 1 in 2000 live births. About 5% of all Caucasians are carriers of the CF mutation.

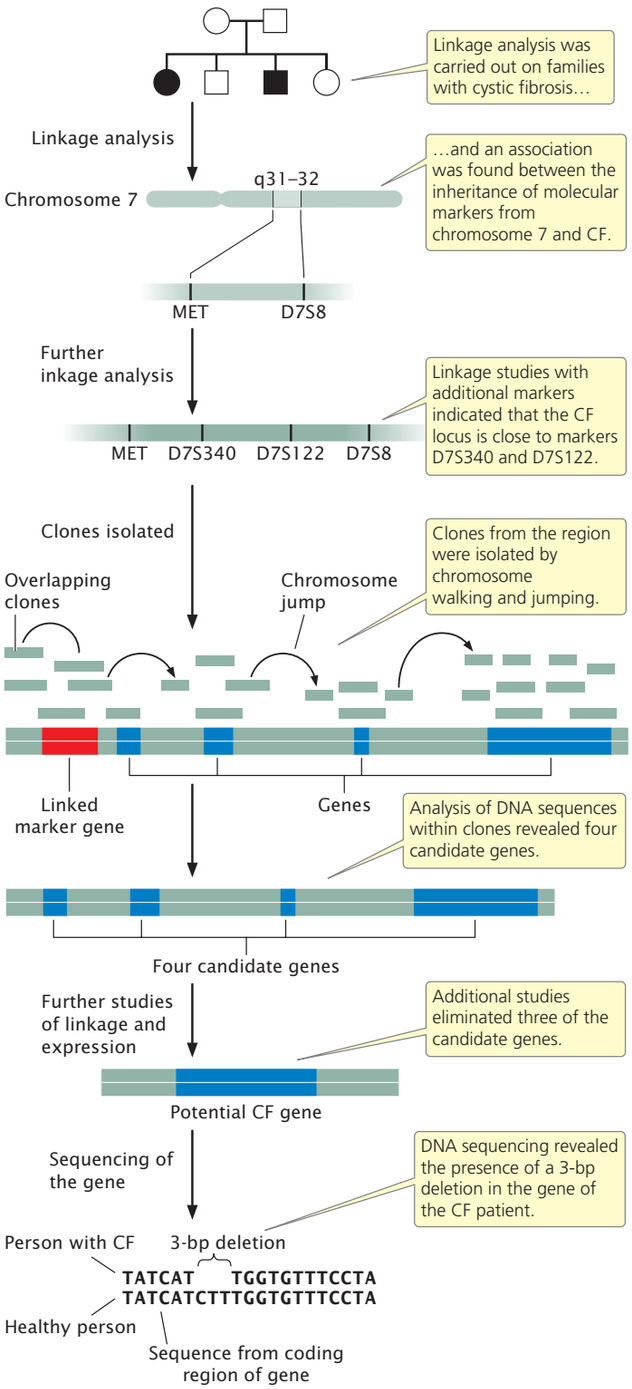


19.20 Cystic fibrosis was the first genetic disease for which the causative gene was isolated entirely by positional cloning.
[Simon Fraser/Science photo Library.]

Geneticists attempting to isolate the gene for CF faced a formidable task. The symptoms of the disease, especially the elevated salt concentration in sweat, suggested that the gene for CF somehow takes part in the movement of ions into and out of the cell, but no information was available about the protein encoded by the gene. Analyses of pedigrees showed that CF is inherited as an autosomal recessive trait, and so it might be located on any one of the 22 pairs of autosomal chromosomes. Thus, geneticists were seeking an unknown gene—probably encompassing a few thousand or tens of thousands of base pairs—among the 3.2 billion base pairs of the human genome.

Researchers began by looking for associations between the inheritance of CF and other traits (**Figure 19.21**). Early studies were limited by the paucity of genetic traits that varied and could be used for gene-mapping studies; but, in the 1980s, advances in molecular biology provided a large number of molecular markers that could be used for linkage analysis (see pp. xxx–xxx in Chapter 7). Geneticists collected pedigrees of families in which several members had CF. They compared the inheritance of CF with that of molecular markers among the members of these families, looking for evidence of linkage. These studies paid off: the gene for CF was found to be closely linked to two markers, MET and D7S8, located on the long arm of chromosome 7. MET and D7S8 are separated by about 1.5 map units (see Chapter 7). In the human genome, each map unit roughly corresponds to 1,000,000 bp; so the gene for CF is located somewhere within a stretch of 1,500,000 bp of DNA, a huge expanse of sequence.

The next step was to carry out further linkage studies with additional markers, to more precisely delineate where in the 1.5-million-base-pair region the CF gene lies. Researchers selected additional molecular markers from the region surrounding MET and D7S8 and performed linkage



19.21 The gene for cystic fibrosis was located by positional cloning.

studies between these new markers and CF (see Figure 19.21). These studies identified two additional markers, D7S122 and D7S340, that are closely linked to CF. Furthermore, they showed that the order of the four markers is MET-D7S340-D7S122-D7S8 and that the CF gene lies very close to D7S122 and D7S340. This finding narrowed the region in which the gene for CF lies to about 500,000 bp.

At this stage, geneticists began isolating clones of sequences from the delineated region. Starting from the molecular markers, they used a combination of chromosome walking and chromosome jumping to identify clones from human genomic libraries that completely covered the region of interest (see Figure 19.21). The examination of sequences within these clones revealed the presence of four genes in the region encompassed by the linked makers (see Figure 19.21). Additional studies were then carried out to better characterize these candidate genes. Three of the candidate genes were eventually eliminated either because linkage analysis suggested that they were not closely linked with the inheritance of CF or because characterization of their DNA sequences suggested they were not the gene for CF.

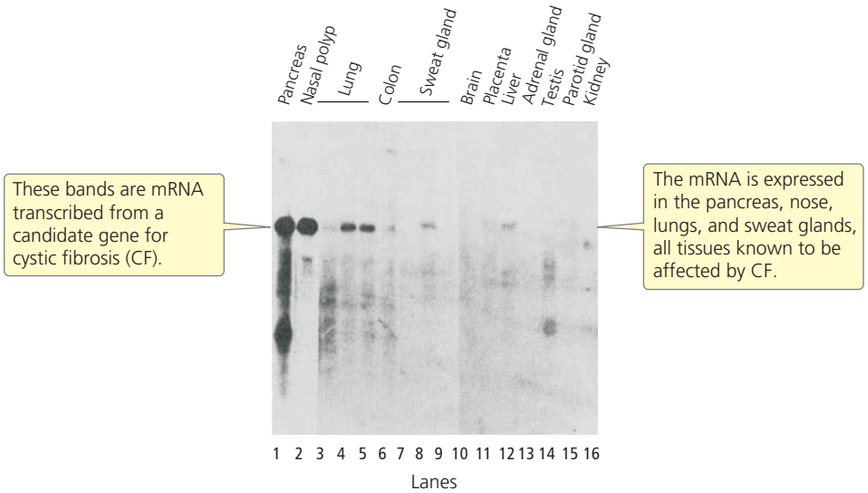
Hybridization studies were carried out with the one remaining gene to determine where it was expressed. Messenger RNA was isolated from different organ tissues, converted into cDNA by reverse transcription, and probed with sequences from the candidate gene. The gene showed high levels of expression in the pancreas, in the lungs, and in sweat glands (**Figure 19.22**), tissues known to be affected by CF.

Copies of the candidate gene from a healthy person and from a person with CF were then sequenced, and the sequence data were examined for differences that might be a mutation causing CF. The findings revealed that the person with CF had a 3-bp deletion in the coding region of the gene; the healthy person did not have this deletion. The deletion resulted in the absence of a phenylalanine amino acid from the protein encoded by the candidate gene. Then, for a large number of patients with CF, geneticists used PCR to amplify the region of the gene where the deletion was found; 68% of the CF patients had this deletion. Subsequent studies demonstrated that the remaining CF patients possessed mutations at other locations within the candidate gene, thus proving that the candidate gene was indeed the locus that caused CF.

Researchers eventually demonstrated that the gene for CF encodes a membrane protein that controls the movement of chloride into and out of cells and is known today as the *cystic fibrosis transmembrane conductance regulator* (*CFTR*) gene. Patients with CF have two mutated forms of *CFTR*, which cause the chloride channels to remain closed. Chloride ions build up in the cell, leading to the formation of thick mucus and the symptoms of the disease.

19.4 DNA Sequences Can Be Determined and Analyzed

In addition to cloning and amplifying DNA, molecular techniques are used to analyze DNA molecules through the study and determination of their sequences.



19.22 A candidate for the cystic fibrosis gene is expressed in pancreatic, respiratory, and sweat-gland tissues—tissues that are affected by the disease. Shown is a Northern blot of mRNA produced by the candidate gene in different tissues. These data provided evidence that the candidate gene is in fact the gene that causes cystic fibrosis. *[After J. R. Riordan et al., 1989,* Science *245:1066–1073.]*

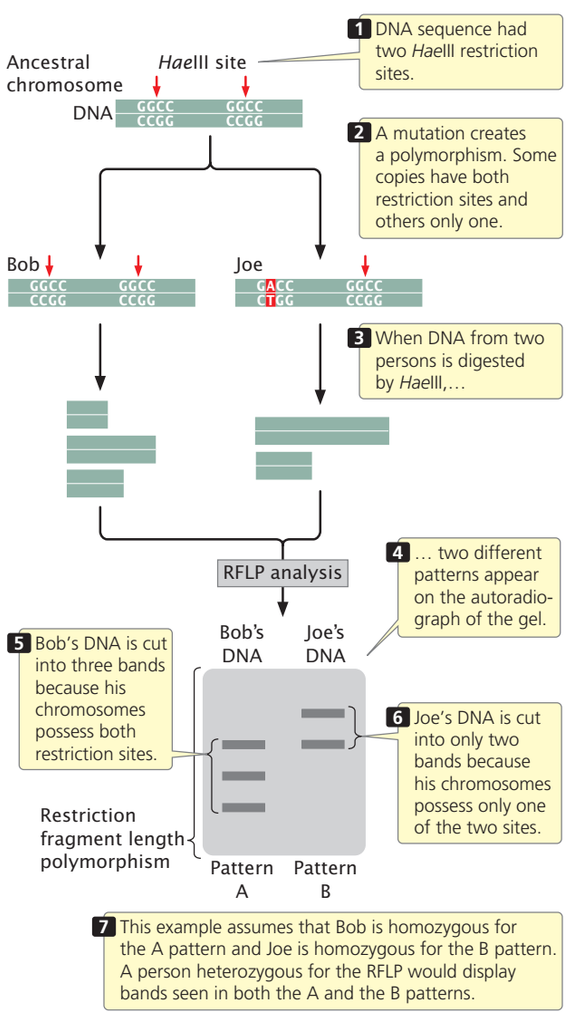
Restriction Fragment Length Polymorphisms

A significant contribution of molecular genetics has been to provide numerous genetic markers that can be used in gene mapping. We saw how these markers are essential to the success of positional cloning. One group of such markers comprises **restriction fragment length polymorphisms** (RFLPs), which are variations (polymorphisms) in the patterns of fragments produced when DNA molecules are cut with the same restriction enzyme (**Figure 19.23**). These differences are inherited and can be used in mapping, similar to the way in which allelic differences are used to map conventional genes.

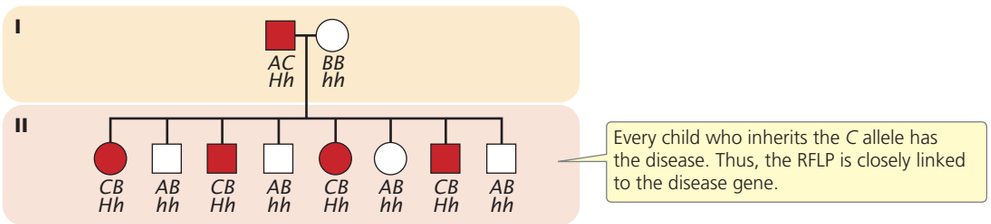
To illustrate mapping with RFLPs, consider Huntington disease, an autosomal dominant disorder. In the family shown in **Figure 19.24**, the father is heterozygous both for Huntington disease (*Hh*) and for a restriction pattern (*AC*). From the father, each child inherits either a Huntington-disease allele (*H*) or a normal allele (*h*); any child inheriting the Huntington-disease allele develops the disease, because it is an autosomal dominant disorder. The child also inherits one of the two RFLP alleles from the father, either *A* or *C,* which produces the corresponding RFLP pattern. In Figure 19.24, every child who inherits the *C* pattern from the father also inherits Huntington disease (and therefore the *H* allele), because the locus for the RFLP is closely linked to the locus for the disease-causing gene. If we had observed no correspondence between the inheritance of the RFLP pattern and the inheritance of the disease, it would indicate that the genes encoding the RFLP and Huntington disease are assorting independently and are not linked.

CONCEPTS

Restriction fragment length polymorphisms are variations in the pattern of fragments produced by restriction enzymes, which reveal variations in DNA sequences. They are used extensively in gene mapping.



19.23 Restriction fragment length polymorphisms are genetic markers that can be used in mapping.



19.24 Restriction fragment length polymorphisms can be used to detect linkage. There is a close correspondence between the inheritance of the RFLP alleles and the presence of Huntington disease, indicating that the genes that encode the RFLP and disease are closely linked.

DNA Sequencing

A powerful molecular method for analyzing DNA is a technique known as **DNA sequencing,** which quickly determines the sequence of bases in DNA. Sequencing allows the genetic information in DNA to be read, providing an enormous amount of information about gene structure and function. In the mid-1970s, Frederick Sanger and his colleagues created the dideoxy-sequencing method based on the elongation of DNA, and it quickly became the standard procedure for sequencing any purified fragment of DNA.

The Sanger, or dideoxy, method of DNA sequencing is based on replication. The fragment to be sequenced is used as a template to make a series of new DNA molecules. In the process, replication is sometimes (but not always) terminated when a specific base is encountered, producing DNA strands of different lengths, each of which ends in the same base.

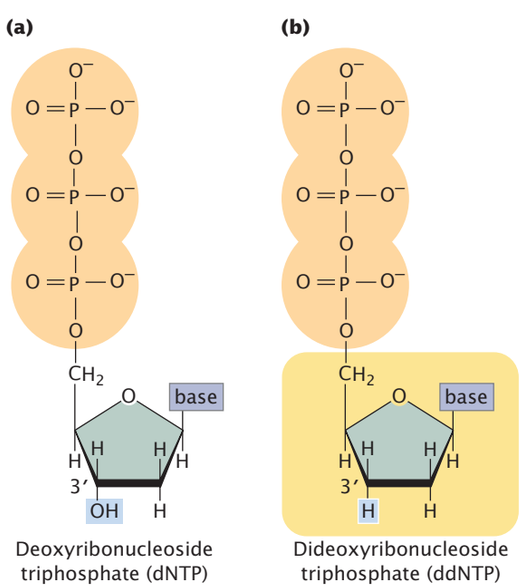
The method relies on the use of a special substrate for DNA synthesis. Normally, DNA is synthesized from deoxyribonucleoside triphosphates (dNTPs), which have an OH group on the 3′-carbon atom (**Figure 19.25a**). In the Sanger method, a special nucleotide, called a **dideoxyribonucleoside triphosphate** (ddNTP; **Figure 19.25b**), is used as one of the substrates. The ddNTPs are identical with dNTPs, except that they lack a 3′-OH group. In the course of DNA synthesis, ddNTPs are incorporated into a growing DNA strand. However, after a ddNTP has been incorporated into the DNA strand, no more nucleotides can be added, because there is no 3′-OH group to form a phosphodiester bond with an incoming nucleotide. Thus, ddNTPs terminate DNA synthesis.

Although the sequencing of a single DNA molecule is technically possible, most sequencing procedures in use today require a considerable amount of DNA; so any DNA fragment to be sequenced must first be amplified by PCR or by cloning in bacteria. Copies of the target DNA are isolated and split into four parts (**Figure 19.26**). Each part is placed in a different tube, to which are added:

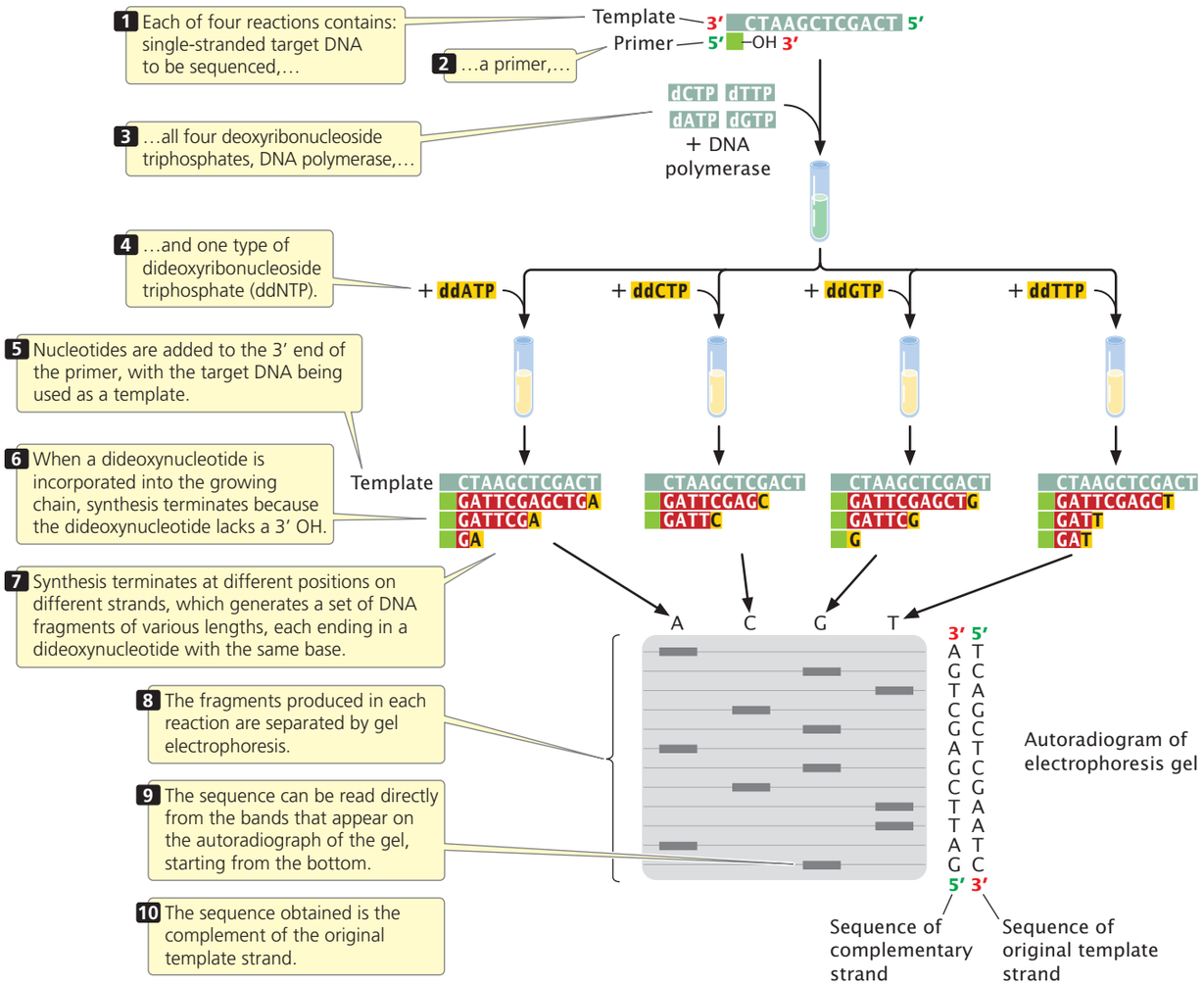
1. many copies of a primer that is complementary to one end of the target DNA strand;
2. all four types of deoxyribonucleoside triphosphates, the normal precursors of DNA synthesis;
3. a small amount of one of the four types of dideoxyribonucleoside triphosphates, which will terminate DNA synthesis as soon as it is incorporated into any growing chain (each of the four tubes received a different ddNTP); and
4. DNA polymerase.

Either the primer or one of the dNTPs is radioactively or chemically labeled so that newly produced DNA can be detected.

Within each of the four tubes, the DNA polymerase enzyme synthesizes DNA. Let's consider the reaction in one of the four tubes; the one that received ddATP. Within this tube, each of the single strands of target DNA serves as a template for DNA synthesis. The primer pairs to its complementary



19.25 The dideoxy sequencing reaction requires a special substrate for DNA synthesis. (a) Structure of deoxyribonucleoside triphosphate, the normal substrate for DNA synthesis. (b) Structure of dideoxyribonucleoside triphosphate, which lacks an OH group on the 3′-carbon atom.



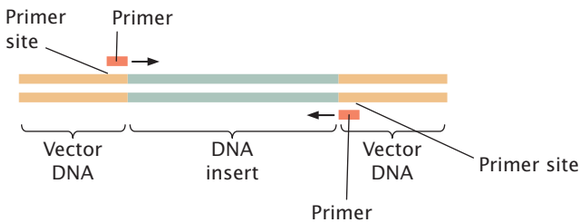
19.26 The dideoxy method of DNA sequencing is based on the termination of DNA synthesis.

sequence at one end of each template strand, providing a 3′-OH group for the initiation of DNA synthesis. DNA polymerase elongates a new strand of DNA from this primer. Wherever DNA polymerase encounters a T on the template strand, it uses at random either a dATP or a ddATP to introduce an A in the newly synthesized strand. Because there is more dATP than ddATP in the reaction mixture, dATP is incorporated most often, allowing DNA synthesis to continue. Occasionally, however, ddATP is incorporated into the strand and synthesis terminates. The incorporation of ddA into the new strand occurs randomly at different positions in different copies, producing a set of DNA chains of different lengths (12, 7, and 2 nucleotides long in the example illustrated in Figure 19.26), each ending in a nucleotide that contains adenine.

Equivalent reactions take place in the other three tubes, except that synthesis is terminated at nucleotides with a different base in each tube. After the completion of the polymerization reactions, all the DNA in the tubes is denatured, and the single-strand products of each reaction are separated by gel electrophoresis.

The contents of the four tubes are separated side by side on an acrylamide gel so that DNA strands differing in length by only a single nucleotide can be distinguished. After electrophoresis, the locations, and therefore the sizes, of the DNA strands in the gel are revealed by autoradiography.

Reading the DNA sequence is the simplest and shortest part of the procedure. In Figure 19.26, you can see that the band closest to the bottom of the gel is from the tube that contained the ddGTP reaction, which means that the first nucleotide synthesized had guanine (G). The next band up is from the tube that contained ddATP; so the next nucleotide in the sequence is adenine (A), and so forth. In this way, the sequence is read from the bottom to the top of the gel, with the nucleotides near the bottom corresponding

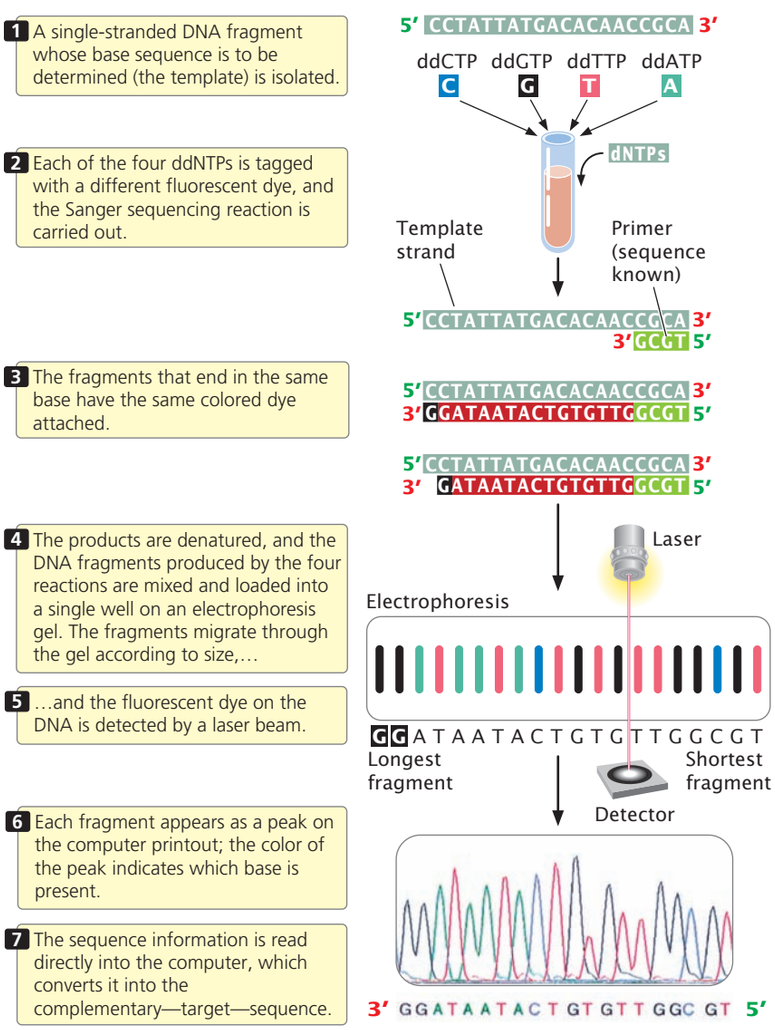


19.27 Sites recognized by sequencing primers are added to the target DNA by cloning the DNA.

to the 5′ end of the newly synthesized DNA strand and those near the top corresponding to the 3′ end. Keep in mind that the sequence obtained is not that of the target DNA but that of its *complement*.

You may have wondered how the primers used in dideoxy sequencing are constructed, because the sequence of the target DNA may not be known ahead of time. The trick is to insert a sequence into the target DNA that will be recognized by the primer. Insertion of the primer sequence is often done by first cloning the target DNA in a vector that contains sequences (called universal sequencing primer sites) recognized by a common primer on either side of the site where the target DNA is inserted. The target DNA is then isolated from the vector and will contain universal sequencing primer sites at each end (**Figure 19.27**). Fragments amplified by PCR also can be sequenced. In this case, the primers used for PCR amplification can also serve as primers for the sequencing reactions.

For many years, DNA sequencing was done largely by hand and was laborious and expensive. Today, sequencing is often carried out by automated machines that use fluorescent dyes and laser scanners to sequence thousands of base pairs in a few hours (**Figure 19.28**). The dideoxy reaction is also used here, but the ddNTPs used in the reaction are labeled with a fluorescent dye, and a different colored dye is used for each type of dideoxynucleotide. In this case, the



19.28 The dideoxy sequencing method can be automated.

four sequencing reactions can take place in the same test tube and can be placed in the same well during electrophoresis. The most recently developed sequencing machines carry out electrophoresis in gel-containing capillary tubes. The different-size fragments produced by the sequencing reaction separate within a tube and migrate past a laser beam and detector. As the fragments pass the laser, their fluorescent dyes are activated and the resulting fluorescence is detected by an optical scanner. Each colored dye emits fluorescence of a characteristic wavelength, which is read by the optical scanner. The information is fed into a computer for interpretation, and the results are printed out as a set of peaks on a graph (see Figure 19.28). Automated sequencing machines may contain 96 or more capillary tubes, allowing from 50,000 to 60,000 bp of sequence to be read in a few hours.

CONCEPTS

DNA can be rapidly sequenced by the dideoxy method, in which ddNTPs are used to terminate DNA synthesis at specific bases. Automated sequencing methods allow tens of thousands of base pairs to be read in just a few hours.

✔ CONCEPT CHECK 10

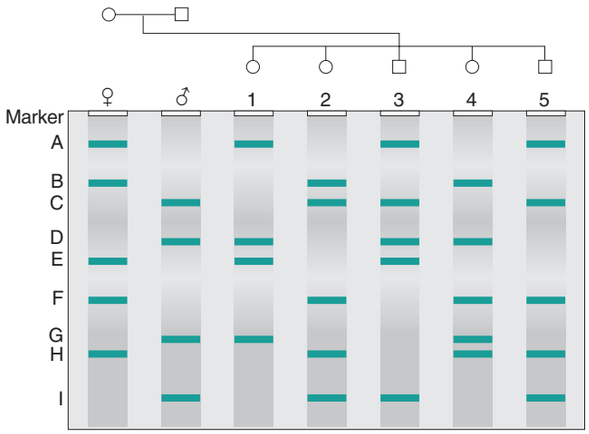
In the dideoxy sequencing reaction, what terminates DNA synthesis at a particular base?

a. The absence of a base on the ddNTP halts the DNA polymerase.
b. The ddNTP causes a break in the sugar–phosphate backbone.
c. DNA polymerase will not incorporate a ddNTP into the growing DNA strand.
d. The absence of a 3′-OH group on the ddNTP prevents the addition of another nucleotide.

DNA Fingerprinting

The use of DNA sequences to identify individual persons is called **DNA fingerprinting** or DNA profiling. Because some parts of the genome are highly variable, each person's DNA sequence is unique and, like a traditional fingerprint, provides a distinctive characteristic that allows identification.

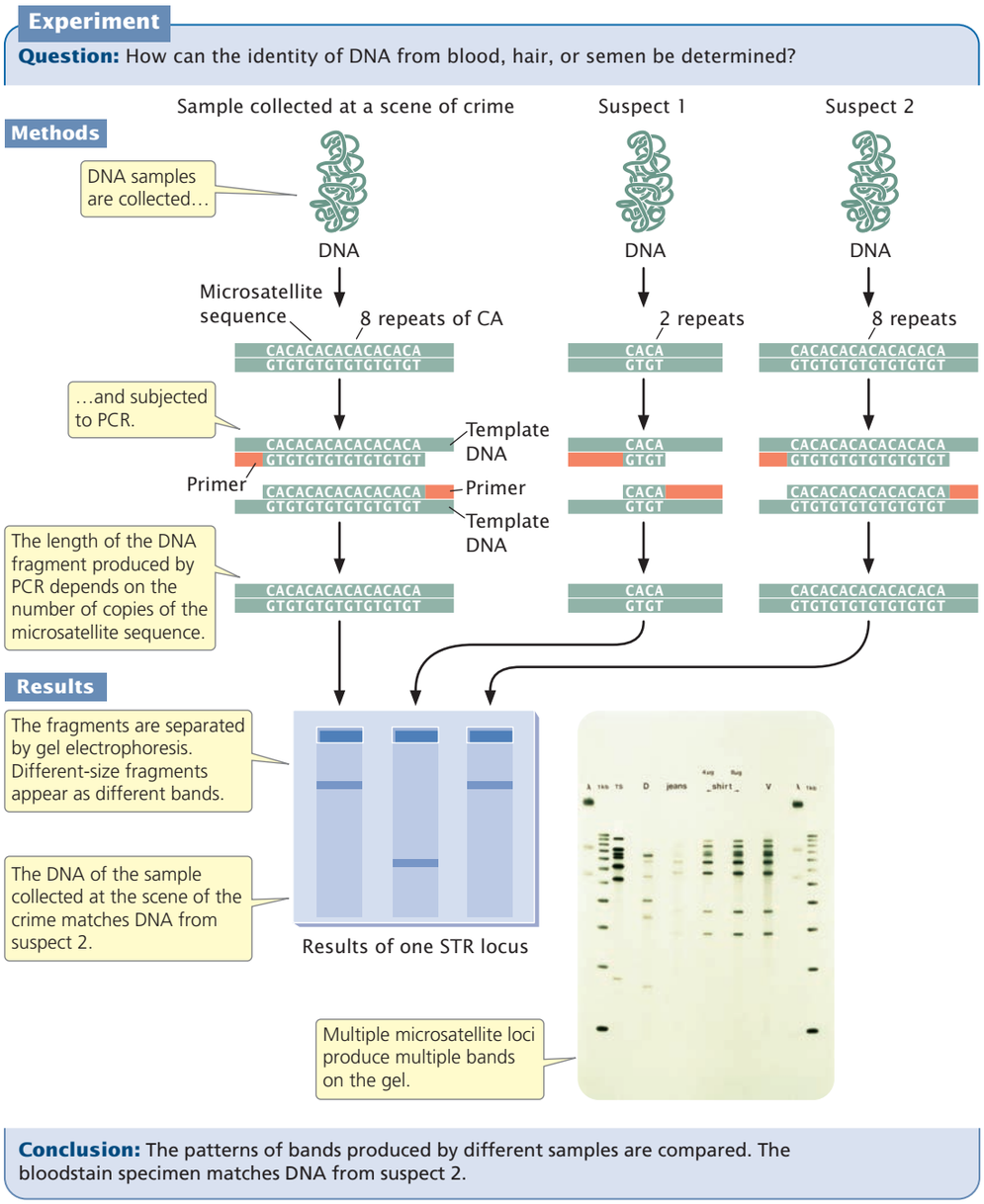
Today, most DNA fingerprinting utilizes **microsatellites,** or **short-tandem repeats** (STRs), which are very short DNA sequences repeated in tandem and are found widely in the human genome (see Chapter 11). People vary in the number of copies of repeat sequences that they possess. Microsatellites are typically detected with PCR, with the use of primers flanking the microsatellite repeats, so that a DNA fragment containing the repeated sequences is amplified. The length of the amplified segment depends on the number of repeats; DNA from a person with more repeats will produce a longer amplified segment than DNA from a person with fewer repeats. After PCR has been completed, the amplified fragments are separated with gel electrophoresis and stained, producing a series of bands on a gel. When several different microsatellite loci are examined,, the probability that two people have the same set of patterns becomes vanishingly small, unless they are identical twins (**Figure 19.29**).



19.29 Variation in banding patterns reveals inherited differences in microsatellite sequences. Microsatellite variation within a family. All bands found in the children are present in the parents. *[From A. Griffiths, S. Wessler, R. Lewontin, W. Gelbart, D. Suzuki, and J. Miller,* Introduction to Genetic Analysis, *8th ed. © 2005 by W. H. Freeman and Company.]*

In a typical application, DNA fingerprinting might be used to confirm that a suspect was present at the scene of a crime (**Figure 19.30**). A sample of DNA from blood, semen, hair, or other body tissue is collected from the crime scene. If the sample is very small, PCR can be used to amplify it so that enough DNA is available for testing. Additional DNA samples are collected from one or more suspects. The pattern of bands produced by DNA fingerprinting from the sample collected at the crime scene is then compared with the patterns produced by DNA fingerprinting of the DNA from the suspects. A match between the sample from the crime scene and one from a suspect can provide evidence that the suspect was present at the scene of the crime.

Since its introduction in the 1980s, DNA fingerprinting has helped convict a number of suspects in murder and rape cases. Suspects in other cases have been proved innocent when their DNA failed to match that from the crime scenes. Initially, calculating the odds of a match (the probability that two people could have the same pattern) was controversial, and there were concerns about quality control (such as the accidental contamination of samples and the reproducibility of results) in laboratories where DNA analysis is done. Today, DNA fingerprinting has become an important tool in forensic investigations. In addition to its application in the analysis of crimes, DNA fingerprinting is used to assess paternity, study genetic relationships among individual organisms in natural populations, identify specific strains of pathogenic bacteria, and identify human remains. For example, DNA fingerprinting was used to determine that several samples of anthrax mailed to different people in 2001 were all from the same source.



19.30 DNA fingerprinting can be used to identify a person. *[Gel courtesy of Orchard Cellmark, Germantown, Maryland.]*

CONCEPTS

DNA fingerprinting detects genetic differences among people by using probes for highly variable regions of chromosomes.

✔ CONCEPT CHECK 11

How are microsatellites detected?

Application: Identifying People Who Died in the Collapse of the World Trade Center

On the morning of September 11, 2001, terrorists hijacked and flew two passenger planes into the World Trade Center towers in New York City. The catastrophic damage and ensuing fire led, within a few hours, to the complete collapse of all 110 floors of both towers, killing almost 3000 building occupants and rescue personnel (**Figure 19.31**). The tremendous

destructive force generated by the collapse of the towers, with their 425,000 cubic yards of concrete and 200,000 tons of steel, pulverized many of the bodies into small pieces that were beyond recognition.

In the days immediately following the World Trade Center collapse, forensic scientists began the task of identifying the remains those who perished in the disaster. The goal was to provide evidence for the ongoing criminal investigation of the attack and to identify the remains at the site for families and friends of the victims. This task was unprecedented in scope and difficulty. There was no complete list of victims (such as a passenger list in an airline crash) with which investigators could match the remains. In all, almost 20,000 individual remains were found, varying from whole bodies to tiny fragments of charred bone. The remains were subjected to fires with temperatures exceeding 1000°C that burned for more than 3 months. The collapse of the buildings intermixed many victims' remains, and many body fragments were not recovered for months, during which time they were exposed to dust, water, bacteria, and decay.

The usual means of victim identification—personal items, fingerprints, dental records—were of little use for most of the World Trade Center remains. Identification of the majority of was made with the use of DNA profiling technology, commonly known as DNA fingerprinting.

DNA was first was extracted from the tissue samples by using sterile techniques to prevent cross-contamination between samples. After the DNA had been extracted, PCR was used to amplify a panel of 13 STR loci (see section on DNA fingerprinting). These loci make up the Combined DNA Index System (CODIS), a system developed by the Federal Bureau of Investigation for use in solving crimes. For the CODIS system, PCR primers that amplify a specific STR locus are used. The length of the amplified fragment depends on the number of repeats. After PCR, the fragments are separated on a gel or by a capillary electrophoresis machine. Each STR locus in the CODIS system has a large number of alleles and is located on a different human chromosome, and so variation at each locus assorts independently. When all 13 CODIS loci are used together, the probability of two randomly selected people having the same DNA profile is less than 1 in 10 billion.

19.31 DNA fingerprinting was used to identify the remains of people who died in the collapse of the World Trade Center. Ground zero after the catastrophic collapse of the 110-story twin towers. *[Michael Reiger/mai/mai/Time Life Pictures/Getty Images.]*

The DNA fingerprint generated from each body sample was compared with that of DNA extracted from reference samples, such as the victims' toothbrushes and blood samples, provided by families and friends. If the DNA from a body part had the same alleles at all 13 loci as the reference sample, then a positive identification was made. When no reference sample was available, investigators collected DNA from family members and tried to match the DNA profiles of remains to those of relatives; in this case, some, but not all STR alleles would match.

Unfortunately, many of the remains were so badly degraded that little DNA remained and one or more of the STR loci could not be amplified. For these remains, DNA fingerprinting was also carried out on mitochondrial DNA (see Chapter 12). Because there are many mitochondria per cell and each mitochondrion contains multiple DNA molecules, there are many more copies of mitochondrial DNA per cell than nuclear DNA; mitochondrial DNA has been successfully extracted and analyzed from ancient remains, such as Neanderthals (see introduction to Chapter 10). Alone, fingerprinting from mitochondrial DNA was insufficient to provide identification with a high degree of confidence (because there are not as many sequences that vary among people as in the CODIS loci) but, when it was used in conjunction with data from at least some STR loci, a positive identification could often be made.

Used in combination, these techniques allowed the remains of many victims to be positively identified. Despite the heroic efforts of hundreds of molecular geneticists, forensic anthropologists, and medical examiners to identify the remains, no positively identified remains were recovered for almost half of the people who are thought to have died in the disaster.

19.5 Molecular Techniques Are Increasingly Used to Analyze Gene Function

In the preceding sections, we have seen how powerful molecular techniques are available for isolating, recombining, and analyzing DNA sequences. Although these methods provide a great deal of information about the organization and nature of gene sequences, the ultimate goal of many molecular studies is to better understand the function of the sequences. In this section, we will explore some advanced molecular techniques that are frequently used to determine

gene function and to better understand the genetic processes that these sequences undergo.

Forward and Reverse Genetics

The traditional approach to the study of gene function begins with the isolation of mutant organisms. For example, suppose a geneticist is interested in genes that affect cardiac function in mammals. A first step would be to find individuals—perhaps mice—that have hereditary defects in heart function. The mutations causing the cardiac problems in the mice could then be mapped, and the implicated genes could be isolated, cloned, and sequenced. The proteins produced by the genes could then be predicted from the gene sequences and isolated. Finally, the biochemistry of the proteins could be studied and their role in heart function discerned. This approach, which begins with a phenotype (a mutant individual) and proceeds to a gene that encodes the phenotype, is called **forward genetics.**

An alternative approach, made possible by advances in molecular genetics, is to begin with a genotype—a DNA sequence—and proceed to the phenotype by altering the sequence or inhibiting its expression. A geneticist might begin with a gene of unknown function, induce mutations in it, and then look to see what effect these mutations have on the phenotype of the organism. This approach is called **reverse genetics.** Today, both forward and reverse genetic approaches are widely used in analysis of gene function. This section introduces some of the molecular techniques that are used in forward and reverse genetics.

Creating Random Mutations

Forward genetics depends on the identification and isolation of random mutations that affect a phenotype of interest. Early in the study of genetics, geneticists were forced to rely on naturally occurring mutations, which are usually rare and can be detected only if large numbers of organisms are examined. The discovery of mutagenic agents—environmental factors that increase the rate of mutation (see Chapter 18)—provided a means increasing the number of mutants in experimental populations of organisms. One of the first examples of experimentally created mutations was H. J. Muller's use of X rays in 1927 to induce X-linked mutations in *Drosophila melanogaster.*

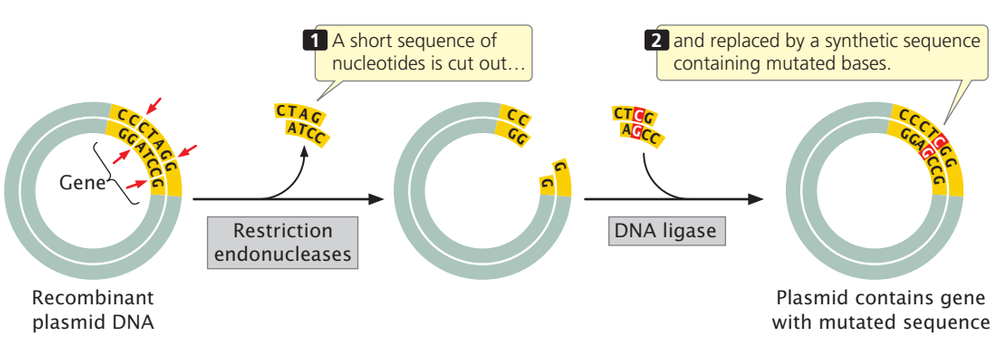
Today, radiation, chemical mutagens, and transposable elements are all used to create mutations for genetic analysis. The choice of mutagen is an important consideration, because different mutagens produce different types of mutations. For example, some mutagens predominately produce base-pair substitutions, whereas other mutagens produce insertions and deletions. The dose of mutagen also is critical. The mutagen needs to increase the number of mutations so that enough mutants are obtained for analysis. However, too high a mutation rate may cause multiple mutations to appear in the same individual organism, making identification of the gene responsible for a mutant phenotype difficult, or may even kill the organism before the mutant phenotype can be observed.

To determine all genes that might affect a phenotype, the creation of mutations in as many genes as possible is desirable—to saturate the genome with mutations. This procedure is done with a mutagenic screen, which will be described in Chapter 20.

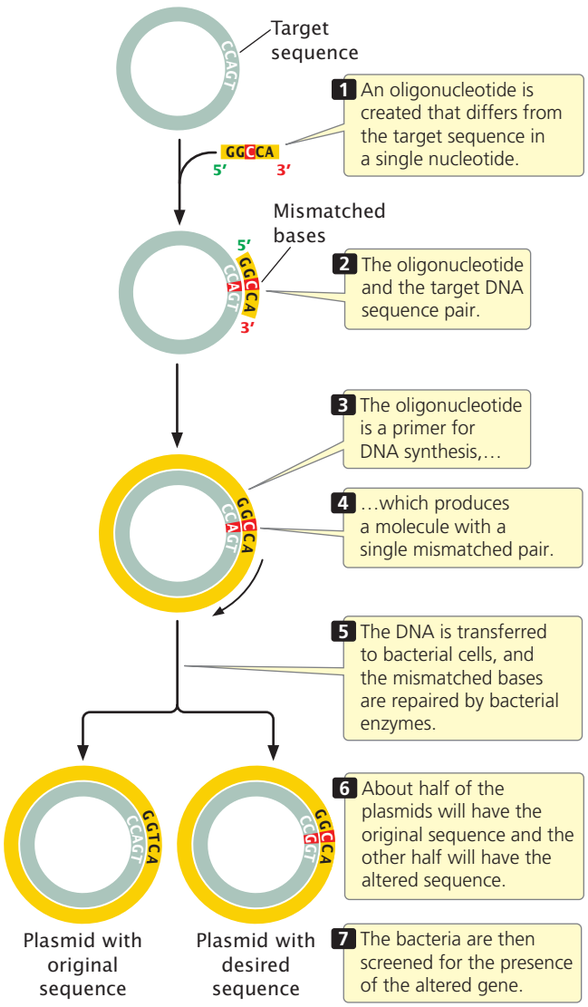
Site-Directed Mutagensis

Reverse genetics depends on the ability to create mutations, not at random, but in particular DNA sequences, and then to study the effects of these mutations on the organism. Mutations are induced at specific locations through a process called **site-directed mutagenesis.**

A number of different strategies have been developed for site-directed mutagenesis. One strategy that is often used in bacteria is to cut out a short sequence of nucleotides with restriction enzymes and replace it with a short, synthetic oligonucleotide that contains the desired mutated sequence (**Figure 19.32**). The success of this method depends on the availability of restriction sites flanking the sequence to be altered.



19.32 In site-directed mutagenesis, restriction enzymes cut out a short sequence of nucleotides that is then replaced by a synthetic mutated DNA sequence.



19.33 Oligonucleotide-directed mutagenesis is used to study gene function when appropriate restriction sites are not available.

If appropriate restriction sites are not available, **oligonucleotide-directed mutagenesis** can be used (**Figure 19.33**). In this method, a single-stranded oligonucleotide is produced that differs from the target sequence by one or a few bases. Because they differ in only a few bases, the target DNA and the oligonucleotide will pair. When successfully paired with the target DNA, the oligonucleotide can act as a primer to initiate DNA synthesis, which produces a double-stranded molecule with a mismatch in the primer region. When this DNA is transferred to bacterial cells, the mismatched bases will be repaired by bacterial enzymes. About half of the time the normal bases will be changed into mutant bases, and about half of the time the mutant bases will be changed into normal bases. The bacteria are then screened for the presence of the mutant gene.

CONCEPTS

Forward genetics begins with a phenotype and detects and analyzes the genotype that causes the phenotype. Reverse genetics begins with a gene sequence and through analysis determines the phenotype that it encodes. Particular mutations can be introduced at specific sites within a gene by means of site-directed and oligonucleotide-directed mutagenesis.

✔ CONCEPT CHECK 12

A geneticist interested in immune function induces random mutations in a number of genes in mice and then determines which of the resulting mutant mice have impaired immune function. This is an example of

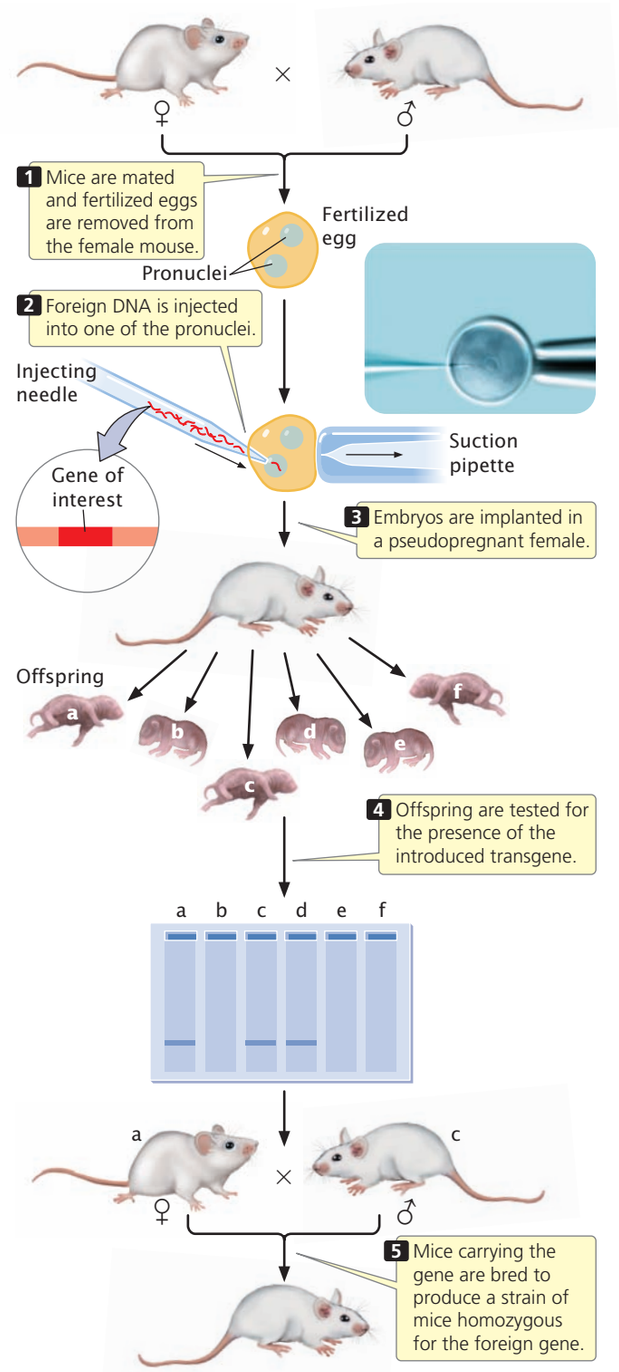
a. forward genetics.
b. reverse genetics.
c. both forward and reverse genetics.
d. neither forward nor reverse genetics.

Transgenic Animals

Another way that gene function can be analyzed is by adding DNA sequences of interest to the genome of an organism that normally lacks such sequences and then seeing what effect the introduced sequence has on the organism's phenotype. This method is a form of reverse genetics. An organism that has been permanently altered by the addition of a DNA sequence to its genome is said to be *transgenic,* and the foreign DNA that it carries is called a **transgene.** Here, we consider techniques for the creation of transgenic mice, which are often used in the study of the function of human genes because they can be genetically manipulated in ways that are impossible with humans and, as mammals, they are more similar to humans than are fruit flies, fish, and other model genetic organisms.

The oocytes of mice and other mammals are large enough that DNA can be injected into them directly. Immediately after penetration by a sperm, a fertilized mouse egg contains two pronuclei, one from the sperm and one from the egg; these pronuclei later fuse to form the nucleus of the embryo. Mechanical devices can manipulate extremely fine, hollow glass needles to inject DNA directly into one of the pronuclei of a fertilized egg (**Figure 19.34**). Typically, a few hundred copies of cloned, linear DNA are injected into a pronucleus, and, in a few of the injected eggs, copies of the cloned DNA integrate randomly into one of the chromosomes through a process called nonhomologous recombination. After injection, the embryos are implanted in a pseudopregnant female—a surrogate mother that has been physiologically prepared for pregnancy by mating with a vasectomized male.

Only about 10% to 30% of the eggs survive and, of those that do survive, only a few have a copy of the cloned



19.34 Transgenic animals have genomes that have been permanently altered through recombinant DNA technology. In the photograph, a mouse embryo is being injected with DNA. *[Photograph: Chad Davis/PhotoDisc.]*

DNA stably integrated into a chromosome. Nevertheless, if several hundred embryos are injected and implanted, there is a good chance that one or more mice whose chromosomes contain the foreign DNA will be born. Moreover, because the DNA was injected at the one-cell stage of the embryo, these mice usually carry the cloned DNA in every cell of their bodies, including their reproductive cells, and will therefore pass the foreign DNA on to their progeny. Through interbreeding, a strain of mice that is homozygous for the foreign gene can be created.

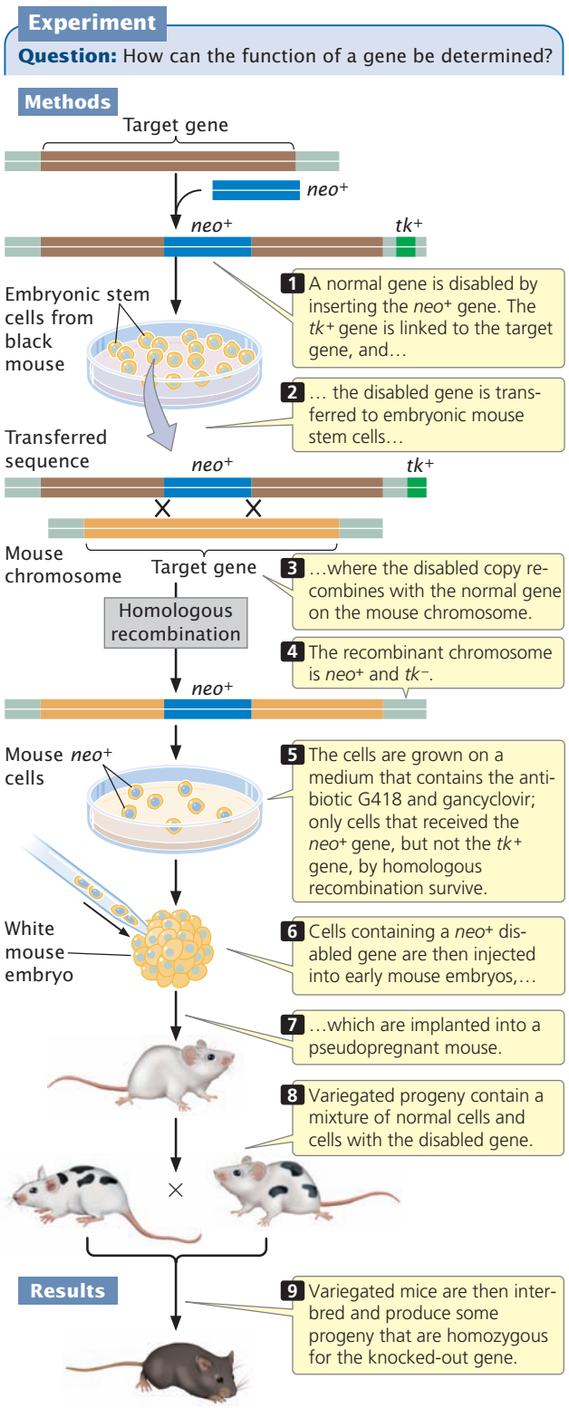
Transgenic mice have proved useful in the study of gene function. For example, proof that the *SRY* gene (see Chapter 4) is the male-determining gene in mice was obtained by injecting a copy of the *SRY* gene into XX embryos and observing that these mice developed as males. In addition, researchers have created a number of transgenic mouse strains that serve as experimental models for human genetic diseases.

Knockout Mice

A useful variant of the transgenic approach is to produce mice in which a normal gene has been not just mutated, but fully disabled. These animals, called **knockout mice,** are particularly helpful in determining the function of a gene: the phenotype of the knockout mouse often gives a good indication of the function of the missing gene.

The creation of knockout mice begins when a normal gene is cloned in bacteria and then "knocked out," or disabled. There are a number of ways to disable a gene, but a common method is to insert a gene called *neo,* which confers resistance to the antibiotic G418, into the middle of the target gene (**Figure 19.35**). The insertion of *neo* both disrupts (knocks out) the target gene and provides a convenient marker for finding copies of the disabled gene. In addition, a second gene, usually the herpes simplex viral thymidine kinase (*tk*) gene, is linked to the disrupted gene. The disabled gene is then transferred to cultured embryonic mouse cells, where it may exchange places with the normal chromosomal copy through homologous recombination.

After the disabled gene has been transferred to the embryonic cells, the cells are screened by adding the antibiotic G418 to the medium. Only cells with the disabled gene containing the *neo* insert will survive. Because the frequency of nonhomologous recombination is higher than that of homologous recombination and because the intact target gene is replaced by the disabled copy only through homologous recombination, a means to select for the rarer homologous recombinants is required. The presence of the viral *tk* gene makes the cells sensitive to gancyclovir. Thus, transfected cells that grow on medium containing G418 and gancyclovir will contain the *neo* gene (disabled target gene) but not the adjacent *tk* gene. These cells contain the desired homologous recombinants. The nonhomologous recombinants (random insertions) will contain both the *neo* and the *tk* genes, and these transfected cells will die on the selection medium owing to the presence of gancyclovir. The surviving cells are injected into an early-stage mouse embryo, which is



19.35 Knockout mice possess a genome in which a gene has been disabled.

then implanted into a pseudopregnant mouse. Cells in the embryo carrying the disabled gene and normal embryonic cells carrying the wild-type gene will develop together, producing a chimera—a mouse that is a genetic mixture of the two cell types. The production of chimeric mice is not itself desirable, but it is not possible to replace all the cells of the embryo with injected cells.

Chimeric mice can be identified easily if the injected embryonic cells came from a black mouse and the embryos into which they are injected came from a white mouse; the resulting chimeras will have variegated black and white fur. The chimeras can then be interbred to produce some progeny that are homozygous for the knockout gene. The effects of disabling a particular gene can be observed in these homozygous mice.

A variant of the knockout procedure is to insert in mice a particular DNA sequence into a known chromosome location. For example, researchers might insert the sequence of a human disease-causing allele into the same locus in mice, creating a precise mouse model of the human disease. Mice that carry inserted sequences at specific locations are called **knock-in mice.**

CONCEPTS

A transgenic mouse is produced by the injection of cloned DNA into the pronucleus of a fertilized egg, followed by implantation of the egg into a female mouse. In knockout mice, the injected DNA contains a mutation that disables a gene. Inside the mouse embryo, the disabled copy of the gene can exchange with the normal copy of the gene through homologous recombination.

✔ CONCEPT CHECK 13

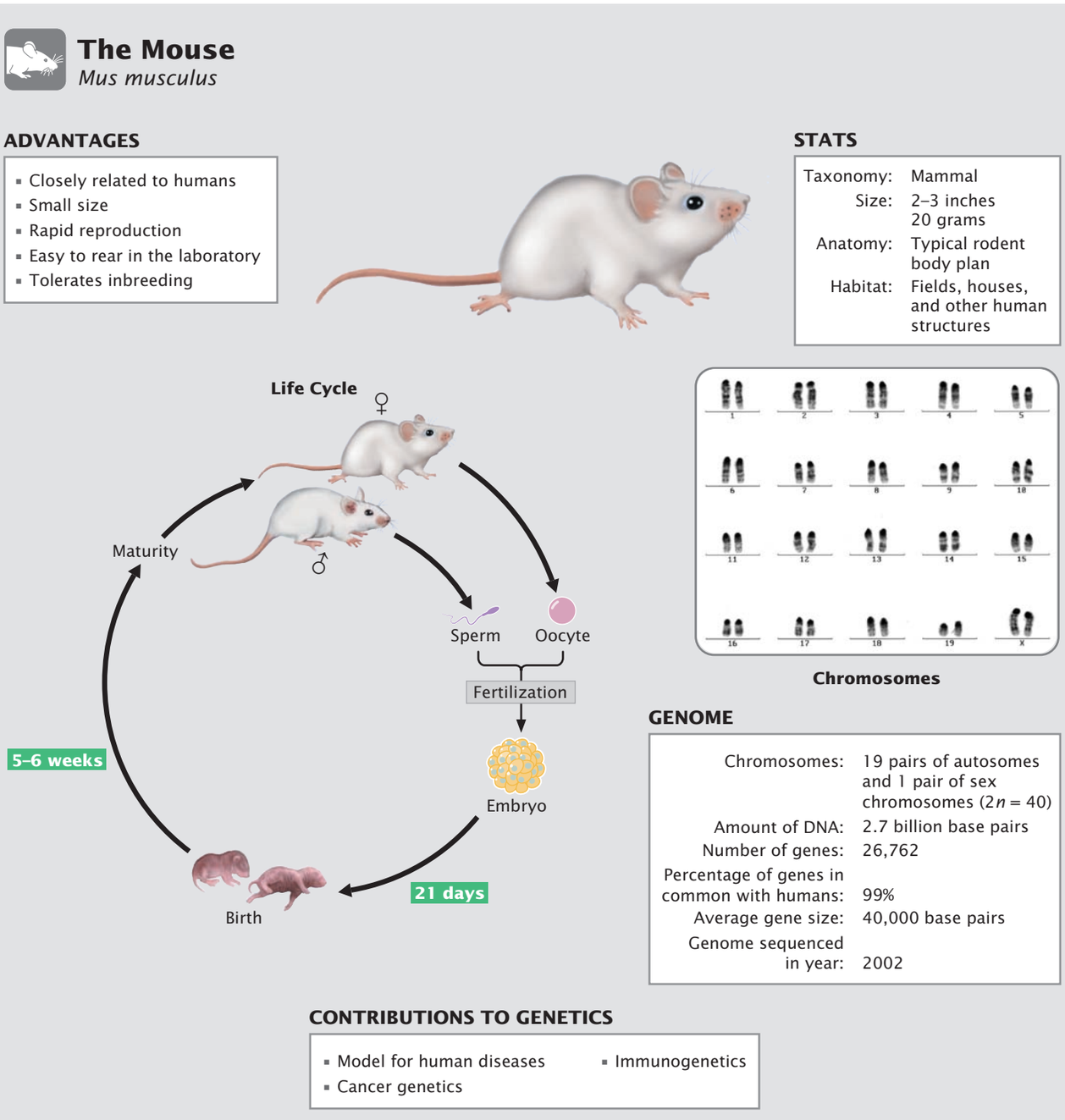
What is the advantage of using the *neo* gene to disrupt the function of a gene in knockout mice?

a. The *neo* gene produces an antibiotic that kills unwanted cells.
b. The *neo* gene is the right size for disabling other genes.
c. The *neo* gene provides a selectable marker for finding cells that contain the disabled gene.
d. The *neo* gene produces a toxin that inhibits transcription of the target gene.

MODEL GENETIC ORGANISM

The Mouse *Mus musculus*

The ability to create transgenic, knockout, and knock-in mice has greatly facilitated the study of human genetics, and these techniques illustrate the power of the mouse as a model genetic organism. The common house mouse, *Mus musculus,* is among the oldest and most valuable subjects for genetic study (**Figure 19.36**). It's an excellent genetic organism—small, prolific, and easy to keep with a short generation time.



19.36 The mouse *Mus musculus* is a model genetic organism. *[Courtesy of Ellen C. Akeson and Muriel T. Davisson, The Jackson Laboratory, Bar Harbor, Maine.]*

Advantages of the mouse as a model genetic organism

Foremost among many advantages that *Mus musculus* has as a model genetic organism is its close evolutionary relationship to humans. Being a mammal, the mouse is genetically, behaviorally, and physiologically more similar to humans than are other organisms used in genetics studies, making the mouse the model of choice for many studies of human and medical genetics. Other advantages include a short generation time compared with that of most other mammals. *Mus musculus* is well adapted to life in the laboratory and

can be easily raised and bred in cages that require little space; thus several thousand mice can be raised within the confines of a small laboratory room. Mice have large litters (8–10 pups), and are docile and easy to handle. Finally, a large number of mutations have been isolated and studied in captive-bred mice, providing an important source of variation for genetic analysis.

Life cycle of the mouse The production of gametes and reproduction in the mouse are very similar to those in humans (see Figure 19.36). Diploid germ cells in the gonads undergo meiosis to produce sperm and oocytes, as outlined in Chapter 2. Male mice begin producing sperm at puberty and continue sperm production throughout the remainder of their lives. Starting at puberty, female mice go through an estrus cycle about every 4 days. If mating takes place during estrus, sperm are deposited into the vagina and swim into the oviduct, where one penetrates the outer layer of the ovum and the nuclei of sperm and ovum fuse. After fertilization, the diploid embryo implants into the uterus. Gestation typically takes about 21 days. Mice reach puberty in about 5 to 6 weeks and will live for about 2 years. A complete generation can be completed in about 8 weeks.

Genetic techniques with the mouse The mouse genome contains about 2.6 billion base pairs of DNA, which is similar in size to the human genome. For most human genes, there are homologous genes in the mouse. An important tool for determining the function of an unknown gene in humans is to search for a homologous gene whose function has already been determined in the mouse. Furthermore, the linkage relations of many mouse genes are similar to those in humans, and the linkage relations of genes in mice often provide important clues to linkage relations among genes in humans. The mouse genome is distributed across 19 pairs of autosomes and one pair of sex chromosomes (see Figure 19.36).

We have already considered three powerful techniques that have been developed for use in the mouse: (1) the creation of transgenic mice by the injection of DNA into a mouse embryo, (2) the ability to disrupt specific genes by the creation of knockout mice, and (3) the ability to insert specific sequences into specific loci. These techniques are made possible by the ability to manipulate the mouse reproductive cycle, including the ability to hormonally induce ovulation, isolate unfertilized oocytes from the ovary, and implant fertilized embryos back into the uterus of a surrogate mother.

A large number of mouse models of specific human diseases have been created—in some cases, by isolating and inbreeding mice with naturally occurring mutations and, in other cases, by using knockout and knock-in techniques to disable and modify specific genes. Mice tolerate inbreeding well, and inbred strains of mice are easily created by brother–sister mating. ■

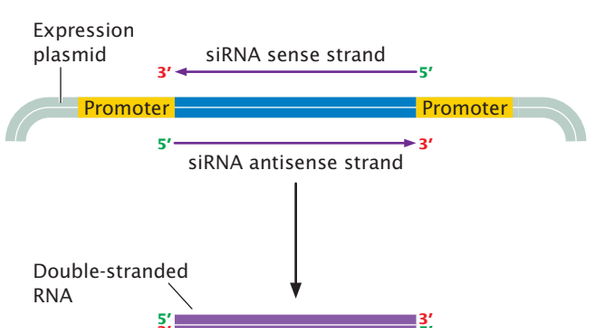
Silencing Genes with RNAi

In the preceding sections, we considered the analysis of gene function by introducing mutations or new DNA sequences into the genome and analyzing the resulting phenotype to provide information about the function of the altered or introduced DNA. We could also analyze gene function by temporarily turning a gene off and seeing what effect the absence of the gene product has on the phenotype. Until recently, there was no method for selectively affecting gene expression. However, the recent discoveries of siRNAs (small interfering RNAs) and miRNAs (microRNAs; see Chapters 14 and 17) have provided powerful tools for controlling the expression of individual genes.

Recall that siRNAs and miRNAs are small RNA molecules that combine with proteins to form the RNA-induced silencing complex (RISC). The RISC pairs with complementary sequences on mRNA and either cleaves the mRNA or prevents the mRNA from being translated. Molecular geneticists have exploited this natural machinery (called RNA interference or RNAi) for turning off the expression of specific genes. Studying the effect of silencing a gene with siRNA can often be a source of insight into the gene's function.

The first step in using RNAi technology is to design the siRNAs such that they will be recognized and cleaved by Dicer (the protein that processes siRNAs; see Chapter 17). The complementary sequence must be unique to the target mRNA and not be found on other mRNAs so that the siRNA will not inhibit nontarget mRNAs. Computer programs are often used to design optimal siRNAs.

After the siRNA sequence has been designed, it must be synthesized. One way to do so is to use an oligonucleotide synthesizer to synthesize a DNA fragment corresponding to the siRNA sequence. The synthesized oligonucleotide can be cloned into a plasmid expression vector between two strong promoters (**Figure 19.37**). *Escherichia coli* are then transformed with the plasmid. Within the bacteria, transcription



19.37 SiRNAs can be produced by cloning DNA sequences corresponding to the siRNAs between two strong promoters. When cloned into an expression vector, both DNA strands will be transcribed and the complementary RNA molecules will anneal to form double-stranded RNA that will be processed into siRNA by Dicer.

from the two promoters will proceed in both directions, producing two complementary RNA molecules that will pair to form a double-stranded RNA molecule recognized by Dicer. Alternatively, double-stranded RNA sequences can be synthesized directly with a gene synthesizer.

The next task is to deliver the double-stranded siRNA to the cells. Delivery can be done in a variety of ways, depending on the cell type. *Caenorhabditis elegans* worms can be fed *E. coli* (their natural food) containing the expression vector. Transcription within the bacteria produces double-stranded RNA, which the worms ingest and incorporate into their cells. Alternatively, double-stranded siRNA can be injected directly into cells or the body cavity. Yet another approach is to synthesize a short sequence of DNA that has internal complementary so that, when transcribed, it folds up into a short hairpin RNA (shRNA) with a double-stranded section. Within a cell, the shRNAs are processed by Dicer to produce siRNAs that bring about gene silencing. DNA sequences containing shRNAs can be introduced into a vector by using standard cloning techniques, and the vector can be used to deliver the DNA into a cell. An advantage of this approach is that, with the addition of a DNA sequence, the RNAi sequence has the potential to become a permanent part of the cell's genome and be passed on to progeny.

The use of RNAi for probing gene function is illustrated by a study of genes that affect male fertility. On the basis of chemical analysis of the chromatin structure in sperm, researchers had identified a group of 132 proteins that might have roles in spermatogenesis and fertility. They created siRNAs that targeted the mRNAs for these 132 proteins and injected the siRNAs into *C. elegans* worms. They then measured the fertility of the treated animals. Of the 132 candidate genes, RNAi treatment in 50 of these genes caused sterility or embryonic death. Many of these same genes are also found in humans, and this study identified a number of genes in humans that potentially take part in human fertility.

Application: Using RNAi for the Treatment of Human Disease

In addition to its value in determining gene function, RNAi holds great potential as a therapeutic agent for the future treatment of human diseases. This potential includes using siRNAs against RNA viruses, such as HIV, as well treating genetic diseases and cancer. This section illustrates the use of RNAi for the potential treatment of disorders of cholesterol metabolism.

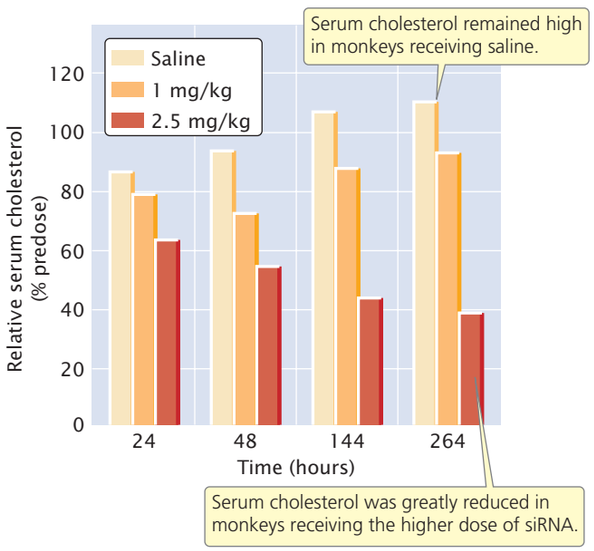
Although cholesterol is essential for life, too much cholesterol is unhealthy: high blood cholesterol is a major contributor to heart disease, the leading cause of death in the United States. Cholesterol is normally transported throughout the body in the form of small particles called lipoproteins, which consist of a core of lipid surrounded by a shell of phospholipids and proteins (see Chapter 6). The ApoB protein is an essential part of lipoproteins. Some people possess genetic mutations that cause elevated levels of ApoB, which predisposes them to coronary artery disease. Findings from studies suggest that lowering the amount of ApoB can reduce the number of lipoproteins and lower blood cholesterol in these people, as well as in people who have elevated cholesterol for other reasons.

In 2006, Tracy Zimmermann and her colleagues at Alnylam Pharmaceuticals and Protiva Biotherapeutics demonstrated that RNAi could be used to reduce the levels of ApoB and blood cholesterol in nonhuman primates. The investigators first created siRNAs that targeted *apoB* gene expression (apoB-siRNAs), based on the known sequence of the gene. The apoB-siRNAs were synthesized in the laboratory and consisted of 21 nucleotides on the sense strand (the strand that was complementary to the *apoB* mRNA) and 23 nucleotides on the complementary antisense strand, with a two-nucleotide overhang.

The next task was to get the apoB-siRNA into the cell. Although siRNAs are readily taken up by the cells of invertebrates such as *C. elegans,* most siRNAs will not readily pass through the membranes of mammalian cells in a form that is still effective in gene silencing. In addition, siRNA are rapidly removed from circulation. To overcome these problems, Zimmermann and her colleagues encapsulated the apoB-siRNAs in lipids, creating stable nucleic-acid–lipid particles (SNALPs). The SNALPs greatly increased the time spent by the siRNAs in circulation and enhanced their uptake by the cell.

The researchers then tested the effects of the apoB-siRNAs on reducing the synthesis of the ApoB protein and on cholesterol levels. They injected cynomolgus monkeys (**Figure 19.38**) with SNALPs containing apoB-siRNA at two doses: 1 mg/kg and 2.5 mg/kg. In addition, they injected a third group of monkeys with saline as a control. They found that the apoB-siRNA clearly silenced the *apoB* gene: 48 hours

19.38 Transfer of siRNAs for the ApoB protein into cynomolgus monkeys demonstrated the potential use of siRNAs for the treatment of hypercholesterolemia. *[Ken Lucas/visuals Unlimited.]*



19.39 Treatment of cynomolgus monkeys with siRNAs for the ApoB protein significantly lowered blood-cholesterol levels. Monkeys were given saline (control), a 1-mg/kg dose, or a 2.5-mg/kg dose of siRNAs. *[T. S. Zimmermann,* Nature *441:112, 2006, Figure 3b.]*

after treatment, *apoB* mRNA in the liver was reduced by 68% for monkeys receiving the 1-mg dose and 90% for monkeys receiving the 2.5-mg dose.

When the researchers examined serum cholesterol levels in the monkeys, they found that monkeys receiving the apoB-siRNA had a significant reduction in blood-cholesterol levels (**Figure 19.39**). Importantly, they observed no negative effects of the siRNA treatment. Although preliminary, this study suggests that siRNAs have great potential for future treatment of human diseases.

19.6 Biotechnology Harnesses the Power of Molecular Genetics

In addition to providing valuable new information about the nature and function of genes, techniques of molecular genes have many practical applications. These applications include the production of pharmaceuticals and other chemicals, specialized bacteria, agriculturally important plants, and genetically engineered farm animals. The technology is also used extensively in medical testing and, in a few cases, is even being used to correct human genetic defects. Hundreds of firms now specialize in developing products through genetic engineering, and many large multinational corporations have invested enormous sums of money in molecular genetics research. As discussed earlier, the analysis of DNA is also used in criminal investigations and for the identification of human remains.

Pharmaceuticals

The first commercial products to be developed with the use of genetic engineering were pharmaceuticals used in the treatment of human diseases and disorders. In 1979, the Eli Lilly corporation began selling human insulin produced with the use of recombinant DNA technology. The gene for human insulin was inserted into plasmids and transferred to bacteria that then produced human insulin. Pharmaceuticals produced through recombinant DNA technology include human growth hormone (for children with growth deficiencies), clotting factors (for hemophiliacs), and tissue plasminogen activator (used to dissolve blood clots in heart-attack patients).

Specialized Bacteria

Bacteria play an important role in many industrial processes, including the production of ethanol from plant material, the leaching of minerals from ore, and the treatment of sewage and other wastes. The bacteria used in these processes are being modified by genetic engineering so that they work more efficiently. New strains of technologically useful bacteria are being developed that will break down toxic chemicals and pollutants, enhance oil recovery, increase nitrogen uptake by plants, and inhibit the growth of pathogenic bacteria and fungi.

Agricultural Products

Recombinant DNA technology has had a major effect on agriculture, where it is now used to create crop plants and domestic animals with valuable traits. For many years, plant pathologists had recognized that plants infected with mild strains of viruses are resistant to infection by virulent strains. Using this knowledge, geneticists have created viral resistance in plants by transferring genes for viral proteins to the plant cells. A genetically engineered squash, called Freedom II, carries genes from the watermelon mosaic virus 2 and the zucchini yellow mosaic virus that protect the squash against viral infections.

Another objective has been to genetically engineer pest resistance into plants to reduce dependence on chemical pesticides. As discussed earlier in the chapter, a gene from the bacterium *Bacillus thuringiensis* that produces an insecticidal toxin has been transferred into corn, tomato, potato, and cotton plants. These BT crops are know grown worldwide. Other genes that confer resistance to viruses and herbicides have been introduced into a number of crop plants. In 2002, more than 37 million hectares of genetically engineered soybeans and 12 million hectares of genetically engineered corn were grown throughout the world.

Recombinant DNA techniques are also applied to domestic animals. For example, the gene for growth hormone was

isolated from cattle and cloned in *E. coli;* these bacteria produce large quantities of bovine growth hormone, which is administered to dairy cattle to increase milk production. Transgenic animals are being developed to carry genes that encode pharmaceutical products; some eukaryotic proteins must be modified after translation, and only other eukaryotes (but not bacteria) are capable of carrying out the modifications. For example, a gene for human clotting factor VIII has been attached to the regulatory region of the sheep gene for β-lactoglobulin, a milk protein. The fused gene was injected in sheep embryos, creating transgenic sheep that produce in their milk the human clotting factor, which is used to treat hemophiliacs.

The genetic engineering of agricultural products is controversial. One area of concern focuses on the potential effects of releasing novel organisms produced by genetic engineering into the environment. There are many examples in which nonnative organisms released into a new environment have caused ecological disruption because they are free of predators and other natural control mechanisms. Genetic engineering normally transfers only small sequences of DNA, relative to the large genetic differences that often exist between species, but even small genetic differences may alter ecologically important traits that might affect the ecosystem.

Another area of concern is the effect of genetically engineered crops on biodiversity. In the largest field test of genetically engineered plants ever conducted, scientists cultivated beets, corn, and oilseed rape that were genetically engineered to resist herbicide along with traditional crops on 200 test plots throughout the United Kingdom and measured the biodiversity of native plants and animals in the agricultural fields. They found that the genetically engineered plants were highly successful in allowing farmers to suppress weeds, but plots with genetically engineered beets and oilseed rape have significantly fewer weeds and insects that feed on weeds. For example, plots with genetically engineered oilseed rape had 24% fewer butterflies than did plots with traditional crops.

There is also concern that transgenic organisms may hybridize with native organisms and transfer their genetically engineered traits. For example, herbicide resistance engineered into crop plants might be transferred to weeds, which would then be resistant to the herbicides that are now used for their control. Other concerns focus on health-safety matters associated with the presence of engineered products in natural foods; some critics have advocated required labeling of all genetically engineered foods that contain transgenic DNA or protein. Such labeling is required in countries of the European Union but not in the United States.

On the other hand, the use of genetically engineered crops and domestic animals has potential benefits. Genetically engineered crops that are pest resistant have the potential to reduce the use of environmentally harmful chemicals, and research findings indicate that lower amounts of pesticides are used in the United States as a result of the adoption of transgenic plants. Transgenic crops also increase yields, providing more food per acre, which reduces the amount of land that must be used for agriculture. As discussed in the introduction to the chapter, genetically engineered plants offer the potential for greater yields that may be necessary to feed the world's future population.

CONCEPTS

Recombinant DNA technology is used to create a wide range of commercial products, including pharmaceuticals, specialized bacteria, genetically engineered crops, and transgenic domestic animals.

✔ CONCEPT CHECK 14

What are some issues of concern about the use of genetically engineered crops?

Oligonucleotide Drugs

A recent application of DNA technology has been the development of oligonucleotide drugs, which are short sequences of synthetic DNA or RNA molecules that can be used to treat diseases. Antisense oligonucleotides are complementary to undesirable RNAs, such as viral RNA. When added to a cell, these antisense DNAs bind to the viral mRNA and inhibit its translation. SiRNAs are currently being tested as potential treatments for some viral and genetic diseases, as well as for cancer. Other oligonucleotides are ribozymes—RNA molecules that function as enzymes (see introduction to Chapter 13). These compounds bind to specific mRNA molecules and cleave them into fragments, destroying their ability to encode proteins. Several oligonucleotide drugs are already being tested for the treatment of AIDS and cancer.

Genetic Testing

The identification and cloning of many important disease-causing human genes have allowed the development of probes for detecting disease-causing mutations. Prenatal testing is already available for several hundred genetic disorders (see Chapter 6). Additionally, presymptomatic genetic tests for adults and children are available for an increasing number of disorders.

The growing availability of genetic tests raises a number of ethical and social questions. For example, is it ethical to test for genetic diseases for which there is no cure or treatment? Other ethical and legal questions concern the confidentiality of test results. Who should have access to the results of genetic testing? Should insurance companies be

allowed to use results from such tests to deny coverage to healthy people who are at risk for genetic diseases? Should relatives who also might be at risk be informed of the results of genetic testing?

Another set of concerns is related to the accuracy of genetic tests. For many genetic diseases, the only predictive tests available are those that identify a *predisposing* mutation in DNA, but many genetic diseases may be caused by dozens or hundreds of different mutations. Probes that detect common mutations can be developed, but they won't detect rare mutations and will give a false negative result. Short of sequencing the entire gene—which is expensive and time consuming—there is no way to identify all predisposed persons. These questions and concerns are currently the focus of intense debate by ethicists, physicians, scientists, and patients.

Gene Therapy

Perhaps the ultimate application of recombinant DNA technology is **gene therapy,** the direct transfer of genes into humans to treat disease. In 1990, gene therapy became reality. W. French Anderson and his colleagues at the U.S. National Institutes of Health (NIH) transferred a functional gene for adenosine deaminase to a young girl with severe combined immunodeficiency disease, an autosomal recessive condition that produces impaired immune function. The researchers removed white blood cells from the girl, and used a retrovirus to transfer a recombinant gene for adenosine deaminase into the cells. The cells were cultured in the laboratory and then implanted back into the patient, where they produced adenosine deaminase and helped alleviate the symptoms of the disease.

Today, thousands of patients have received gene therapy, and many clinical trials are underway. Gene therapy is being used to treat genetic diseases, cancer, heart disease, and even some infectious diseases such as AIDS. A number of different methods for transferring genes into human cells are currently under development. Commonly used vectors include genetically modified retroviruses, adenoviruses, and adeno-associated viruses (**Table 19.3**).

In spite of the growing number of clinical trials for gene therapy, significant problems remain in transferring foreign genes into human cells, getting them expressed, and limiting immune responses to the gene products and the vectors used to transfer the genes to the cells. There are also heightened concerns about safety. In 1999, a patient participating in a gene-therapy trial had a fatal immune reaction after he was injected with a viral vector carrying a gene to treat his

Table 19.3 Vectors used in gene therapy

Vector	Advantages	Disadvantages
Retrovirus	Efficient transfer	Transfers DNA only to dividing cells, inserts randomly; risk of producing wild-type viruses
Adenovirus	Transfers to nondividing cells	Causes immune reaction
Adeno-associated virus	Does not cause immune reaction	Holds small amount of DNA; hard to produce
Herpes virus	Can insert into cells of nervous system; does not cause immune reaction	Hard to produce in large quantities
Lentivirus	Can accommodate large genes	Safety concerns
Liposomes and other lipid-coated vectors	No replication; does not stimulate immune reaction	Low efficiency
Direct injection	No replication; directed toward specific tissues	Low efficiency; does not work well within some tissues
Pressure treatment	Safe, because tissues are treated outside the body and then transplanted into the patient	Most efficient with small DNA molecules
Gene gun (DNA coated on small gold particles and shot into tissue)	No vector required	Low efficiency

Source: After E. Marshall, Gene therapy's growing pains, *Science* 269:1050–1055, 1995.

metabolic disorder. And in 2002, two children who had undergone gene therapy for severe combined immunodeficiency disease developed leukemia that appeared to be directly related to the insertion of the retroviral gene vectors into cancer-causing genes. Despite these setbacks, gene-therapy research has moved ahead. Unequivocal results demonstrating positive benefits from gene therapy for a severe combined immunodeficiency disease and for head and neck cancer were announced in 2000.

Gene therapy conducted to date has targeted only nonreproductive, somatic cells. Correcting a genetic defect in these cells (termed *somatic gene therapy*) may provide positive benefits to patients but will not affect the genes of future generations. Gene therapy that alters reproductive, or germline, cells (termed germ-line gene therapy) is technically possible but raises a number of significant ethical issues, because it has the capacity to alter the gene pool of future generations.

CONCEPTS

Gene therapy is the direct transfer of genes into humans to treat disease. Gene therapy was first successfully implemented in 1990 and is now being used to treat genetic diseases, cancer, and infectious diseases.

✔ CONCEPT CHECK 15

What is the difference between somatic gene therapy and germ-line gene therapy?

CONCEPTS SUMMARY

- Restriction endonucleases are enzymes that make double-stranded cuts in DNA at specific base sequences.
- DNA fragments can be separated with the use of gel electrophoresis and visualized by staining the gel with a dye that is specific for nucleic acids or by labeling the fragments with a radioactive or chemical tag.
- In gene cloning, a gene or a DNA fragment is placed into a bacterial cell, where it will be multiplied as the cell divides.
- Plasmids, small circular pieces of DNA, are often used as vectors to ensure that a cloned gene is stable and replicated within the recipient cells. Expression vectors contain sequences necessary for foreign DNA to be transcribed and translated.
- The polymerase chain reaction is a method for amplifying DNA enzymatically without cloning. A solution containing DNA is heated, so that the two DNA strands separate, and then quickly cooled, allowing primers to attach to the template DNA. The solution is then heated again, and DNA polymerase synthesizes new strands from the primers. Each time the cycle is repeated, the amount of DNA doubles.
- Genes can be isolated by creating a DNA library—a set of bacterial colonies or viral plaques that each contain a different cloned fragment of DNA. A genomic library contains the entire genome of an organism; a cDNA library contains DNA fragments complementary to all the different mRNAs in a cell.
- In situ hybridization can be used to determine the chromosomal location of a gene and the distribution of the mRNA produced by a gene.
- Positional cloning uses linkage relations to determine the location of genes without any knowledge of their products.
- The Sanger (dideoxy) method of DNA sequencing uses special substrates for DNA synthesis (dideoxynucleoside triphosphates, ddNTPs) that terminate synthesis after they are incorporated into the newly made DNA. In a sequencing reaction, DNA fragments of varying length are produced, all of which terminate in nucleotides with the same base. The products of the four reactions are separated by gel electrophoresis, and the sequence of the DNA synthesized is read from the pattern of bands on the gel.
- Short tandem repeats (STRs) and microsatellites are used to identify people by their DNA sequences (DNA fingerprinting).
- Forward genetics begins with a phenotype and conducts analyses to locate the responsible genes. Reverse genetics starts with a DNA sequence and conducts analyses to determine its phenotypic effect.
- Site-directed mutagenesis can be used to produce mutations at specific sites in DNA, allowing genes to be tailored for a particular purpose.
- Transgenic animals, produced by injecting DNA into fertilized eggs, contain foreign DNA that is integrated into a chromosome. Knockout mice are transgenic mice in which a normal gene is disabled.
- The mouse *Mus musculus* is an excellent model genetic organism because of its similarity to humans, small size, and short generation time.
- RNA interference is used to silence the expression of specific genes.
- Techniques of molecular genetics are being used to create products of commercial importance, to develop diagnostic tests, and to treat diseases.
- In gene therapy, diseases are being treated by altering the genes of human cells.

IMPORTANT TERMS

recombinant DNA technology (p. 2)
genetic engineering (p. 2)
biotechnology (p. 2)
restriction enzyme (p. 3)
restriction endonuclease (p. 3)
cohesive end (p. 4)
gel electrophoresis (p. 6)
autoradiography (p. 6)
probe (p. 7)
Southern blotting (p. 7)
Northern blotting (p. 7)
Western blotting (p. 7)
gene cloning (p. 8)
cloning vector (p. 8)
linker (p. 8)
cosmid (p. 10)
bacterial artificial chromosome (BAC) (p. 10)
expression vector (p. 10)
yeast artificial chromosome (YAC) (p. 11)
Ti plasmid (p. 11)
polymerase chain reaction (PCR) (p. 12)
Taq polymerase (p. 13)
reverse-transcription PCR (p. 13)
real-time PCR (p. 13)
DNA library (p. 15)
genomic library (p. 15)
cDNA library (p. 15)
in situ hybridization (p. 18)
positional cloning (p. 19)
chromosome walking (p. 19)
chromosome jumping (p. 19)
restriction fragment length polymorphism (RFLP) (p. 23)
DNA sequencing (p. 24)
dideoxyribonucleoside triphosphate (ddNTP) (p. 24)
DNA fingerprinting (p. 27)
microsatellite (p. 27)
short-tandem repeat (STR) (p. 27)
forward genetics (p. 30)
reverse genetics (p. 30)
site-directed mutagenesis (p. 30)
oligonucleotide-directed mutagenesis (p. 31)
transgene (p. 31)
knockout mice (p. 32)
knock-in mice (p. 33)
gene therapy (p. 39)

ANSWERS TO CONCEPT CHECKS

1. First, the gene must be located and isolated from the rest of the genomic DNA. Then, the gene must be inserted into bacteria in a form that is stable and will be replicated. The gene must be placed in the bacteria in a way that ensures that it will be transcribed and translated. Finally, those bacteria that have taken up an active form of the gene must be separated from other bacteria.
2. Restriction enzymes exist naturally in bacteria, which use them to prevent the entry of viral DNA.
3. c
4. Southern blotting is used to transfer DNA from a gel to a solid medium. Northern blotting is used to transfer RNA from a gel to a solid medium, and Western blotting is used to transfer protein from a gel to a solid medium.
5. An origin of replication, selectable markers, and one or more unique restriction sites.
6. It requires known sequences to which primers attach, is easily contaminated by minute amounts of nontarget DNA, has limited accuracy, and can amplify only small fragments.
7. A cDNA library is smaller than a genomic library, which makes it easier to screen. However, a cDNA library does not contain sequences that are not found in mRNA, such as introns, promoters, and noncoding sequences between genes.
8. With the use of the genetic code and the amino acid sequence of the protein, possible nucleotide sequences that cover a small region of the gene can be deduced. A mixture of all the possible nucleotide sequences that might encode the protein, taking into consideration synonymous codons, are used as probes. To minimize the number of sequences required, a region of the protein that has relatively little degeneracy in its codons is selected.
9. The expression pattern of the gene can be examined, and the coding region of copies of the gene from individuals with mutant phenotype can be compared with the coding region of wild-type individuals.
10. d
11. By using PCR with primers that flank the region containing tandem repeats
12. b
13. c
14. Possible concerns include: (a) ecological damage caused by introducing novel organisms into the environment; (b) negative effects of transgenic organisms on biodiversity; (c) possible spread of transgenes to native organisms by hybridization; and (d) health effects of eating genetically modified foods.
15. Somatic gene therapy modifies genes only in somatic tissue, and these modifications cannot be inherited. Germ-line gene therapy alters genes in germ-line cells and will be inherited.

WORKED PROBLEMS

1. A molecule of double-stranded DNA that is 5 million base pairs long has a base composition that is 62% G + C. How many times, on average, are the following restriction sites likely to be present in this DNA molecule?

a. *Bam*HI (recognition sequence is GGATCC)

b. *Hin*dIII (recognition sequence is AAGCTT)

c. *Hpa*II (recognition sequence is CCGG)

• Solution

The percentages of G and C are equal in double-stranded DNA; so, if G + C = 62%, then %G = %C = 62%/2 = 31%. The percentage of A + T = (100% − G + C) = 38%, and %A = %T = 38%/2 = 19%. To determine the probability of finding a particular base sequence, we use the multiplication rule, multiplying together the probably of finding each base at a particular site.

a. The probability of finding the sequence GGATCC = 0.31 × 0.31 × 0.19 × 0.19 × 0.31 × 0.31 = 0.0003333. To determine the average number of recognition sequences in a 5-million-base-pair piece of DNA, we multiply 5,000,000 bp × 0.00033 = 1666.5 recognition sequences.

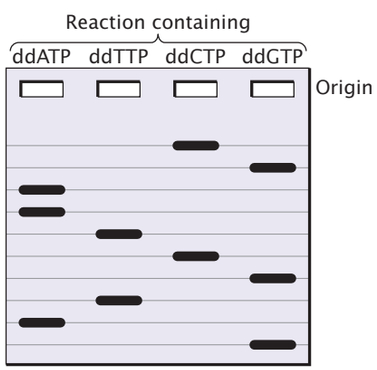
b. The number of AAGCTT recognition sequences is 0.19 × 0.19 × 0.31 × 0.31 × 0.19 × 0.19 × 5,000,000 = 626 recognition sequences.

c. The number of CCGG recognition sequences is 0.31 × 0.31 × 0.31 × 0.31 × 5,000,000 = 46,176 recognition sequences.

2. You are given the following DNA fragment to sequence: 5′–GCTTAGCATC–3′. You first clone the fragment in bacterial cells to produce sufficient DNA for sequencing. You isolate the DNA from the bacterial cells and carry out the dideoxy-sequencing method. You then separate the products of the polymerization reactions by gel electrophoresis. Draw the bands that should appear on the gel from the four sequencing reactions.

• Solution

In the dideoxy-sequencing reaction, the original fragment is used as a template for the synthesis of a new DNA strand; the sequence of the new strand is the sequence that is actually determined. The first task, therefore, is to write out the sequence of the newly synthesized fragment, which will be complementary and antiparallel to the original fragment. The sequence of the newly synthesized strand, written 5′ → 3′ is: 5′–GATGCTAAGC–3′. Bands representing this sequence will appear on the gel, with the bands representing nucleotides near the 5′ end of the molecule at the bottom of the gel.



COMPREHENSION QUESTIONS

Section 19.1

1. List some of the effects and practical applications of molecular genetic analyses.

Section 19.2

2. What common feature is seen in the sequences recognized by type II restriction enzymes?

3. What role do restriction enzymes play in bacteria? How do bacteria protect their own DNA from the action of restriction enzymes?

***4.** Explain how gel electrophoresis is used to separate DNA fragments of different lengths.

***5.** After DNA fragments have been separated by gel electrophoresis, how can they be visualized?

6. What is the purpose of Southern blotting? How is it carried out?

***7.** Give three important characteristics of cloning vectors.

8. Briefly describe four different methods for inserting foreign DNA into plasmids, giving the strengths and weaknesses of each method.

***9.** Briefly explain how an antibiotic-resistance gene and the *lacZ* gene can be used as markers to determine which cells contain a particular plasmid.

***10.** Briefly explain how the polymerase chain reaction is used to amplify a specific DNA sequence. What are some of the limitations of PCR?

11. What is real-time PCR?

Section 19.3

***12.** How does a genomic library differ from a cDNA library? How is each library created?

13. How are probes used to screen DNA libraries? Explain how a synthetic probe can be prepared when the protein product of a gene is known.

***14.** Briefly explain in situ hybridization, giving some applications of this technique.

15. Briefly explain how a gene can be isolated through positional cloning.

16. Explain how chromosome walking can be used to find a gene.

Section 19.4

17. What is the purpose of the dideoxynucleoside triphosphate in the dideoxy sequencing reaction?

***18.** What is DNA fingerprinting? What types of sequences are examined in DNA fingerprinting?

Section 19.5

19. How does a reverse genetics approach differ from a forward genetics approach?

20. Briefly explain how site-directed mutagenesis is carried out.

***21.** What are knockout mice, how are they produced, and for what are they used?

22. What are some advantages that mice possess as model genetic organisms?

23. How is RNA interference used in the analysis of gene function?

Section 19.6

24. What is gene therapy?

APPLICATION QUESTIONS AND PROBLEMS

Section 19.2

***25.** Suppose that a geneticist discovers a new restriction enzyme in the bacterium *Aeromonas ranidae.* This restriction enzyme is the first to be isolated from this bacterial species. Using the standard convention for abbreviating restriction enzymes, give this new restriction enzyme a name (for help, see footnote to Table 19.1).

26. How often, on average, would you expect a type II restriction endonuclease to cut a DNA molecule if the recognition sequence for the enzyme had 5 bp? (Assume that the four types of bases are equally likely to be found in the DNA and that the bases in a recognition sequence are independent.) How often would the endonuclease cut the DNA if the recognition sequence had 8 bp?

***27.** A microbiologist discovers a new type II restriction endonuclease. When DNA is digested by this enzyme, fragments that average 1,048,500 bp in length are produced. What is the most likely number of base pairs in the recognition sequence of this enzyme?

28. Will restriction sites for an enzyme that has 4 bp in its restriction site be closer together, farther apart, or similarly spaced, on average, compared with those of an enzyme that has 6 bp in its restriction site? Explain your reasoning.

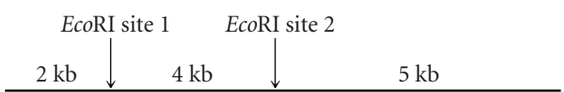
***29.** About 60% of the base pairs in a human DNA molecule are AT. If the human genome has 3.2 billion base pairs of DNA, about how many times will the following restriction sites be present?

a. *Bam*HI (restriction site is 5′–GGATCC–3′)

b. *Eco*RI (restriction site is 5′–GAATTC–3′)

c. *Hae*III (restriction site is 5′–GGCC–3′)

***30.** Restriction mapping of a linear piece of DNA reveals the following *Eco*RI restriction sites.



a. This piece of DNA is cut by *Eco*RI, the resulting fragments are separated by gel electrophoresis, and the gel is stained with ethidium bromide. Draw a picture of the bands that will appear on the gel.

b. If a mutation that alters *Eco*RI site 1 occurs in this piece of DNA, how will the banding pattern on the gel differ from the one that you drew in part *a?*

c. If mutations that alter *Eco*RI sites 1 and 2 occur in this piece of DNA, how will the banding pattern on the gel differ from the one that you drew in part *a?*

d. If 1000 bp of DNA were inserted between the two restriction sites, how would the banding pattern on the gel differ from the one that you drew in part *a?*

e. If 500 bp of DNA between the two restriction sites were deleted, how would the banding pattern on the gel differ from the one that you drew in part *a?*

***31.** Which vectors (plasmid, phage λ, cosmid, bacterial artificial chromosome) can be used to clone a continuous fragment of DNA with the following lengths?

a. 4 kb **c.** 35 kb

b. 20 kb **d.** 100 kb

32. A geneticist uses a plasmid for cloning that has the *lacZ* gene and a gene that confers resistance to penicillin. The

geneticist inserts a piece of foreign DNA into a restriction site that is located within the *lacZ* gene and uses the plasmid to transform bacteria. Explain how the geneticist can identify bacteria that contain a copy of a plasmid with the foreign DNA.

Section 19.3

***33.** Suppose that you have just graduated from college and have started working at a biotechnology firm. Your first job assignment is to clone the pig gene for the hormone prolactin. Assume that the pig gene for prolactin has not yet been isolated, sequenced, or mapped; however, the mouse gene for prolactin has been cloned and the amino acid sequence of mouse prolactin is known. Briefly explain two different strategies that you might use to find and clone the pig gene for prolactin.

34. A genetic engineer wants to isolate a gene from a scorpion that encodes the deadly toxin found in its stinger, with the ultimate purpose of transferring this gene to bacteria and producing the toxin for use as a commercial pesticide. Isolating the gene requires a DNA library. Should the genetic engineer create a genomic library or a cDNA library? Explain your reasoning.

***35.** A protein has the following amino acid sequence:

Met-Tyr-Asn-Val-Arg-Val-Tyr-Lys-Ala-Lys-
Trp-Leu-Ile-His-Thr-Pro

You wish to make a set of probes to screen a cDNA library for the sequence that encodes this protein. Your probes should be at least 18 nucleotides in length.

a. Which amino acids in the protein should be used to construct the probes so that the least degeneracy results? (Consult the genetic code in Figure 15.10.)

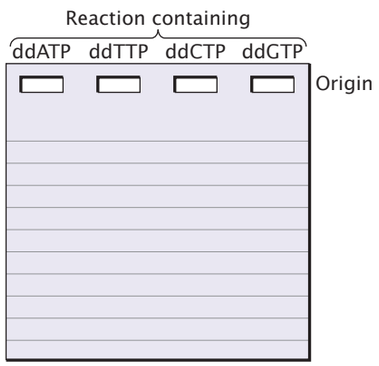
b. How many different probes must be synthesized to be certain that you will find the correct cDNA sequence that specifies the protein?

Section 19.4

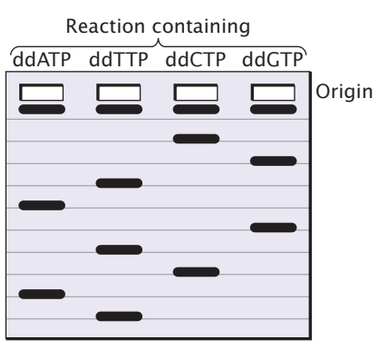
36. Suppose that you want to sequence the following DNA fragment:

5′–TCCCGGGAAA-primer site–3′

You first clone the fragment in bacterial cells to produce sufficient DNA for sequencing. You isolate the DNA from the bacterial cells and apply the dideoxy sequencing method. You then separate the products of the polymerization reactions by gel electrophoresis. Draw the bands that should appear on the gel from the four sequencing reactions.



***37.** Suppose that you are given a short fragment of DNA to sequence. You clone the fragment, isolate the cloned DNA fragment, and set up a series of four dideoxy reactions. You then separate the products of the reactions by gel electrophoresis and obtain the following banding pattern:

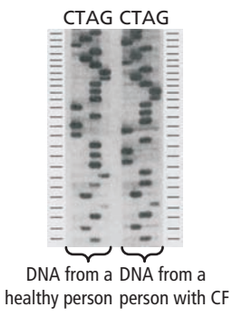


Write out the base sequence of the original fragment that you were given.

Original sequence: 5′– _____ –3′

38. DATA ANALYSIS The autoradiograph on the next page is from the original study that first sequenced the cystic fibrosis gene (J. R. Riordan et al. 1989. *Science* 245:1066–1073).

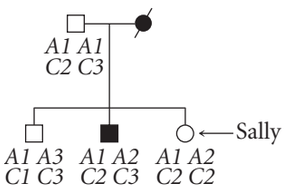
From the autoradiograph, determine the sequence of the normal copy of the gene and the sequence of the mutated copy of the gene. Identify the location of the mutation that causes cystic fibrosis.



***39.** A hypothetical disorder called G syndrome is an autosomal dominant disease characterized by visual, skeletal, and cardiovascular defects. The disorder appears in middle age. Because its symptoms are variable, the disorder is difficult to diagnose. Early diagnosis is important, however, because the cardiovascular defects can be treated if the disorder is recognized early. The gene for G syndrome is known to reside on chromosome 7, and it is closely linked to two RFLPs on the same chromosome, one at the *A* locus and one at the *C* locus. The *G, A,* and *C* loci are very close together, and there is little crossing over between them. The following RFLP alleles are found at the *A* and *C* loci:

A locus: *A1, A2, A3*
C locus: *C1, C2, C3*

Sally, is concerned that she might have G syndrome. Her deceased mother had G syndrome, and she has a brother with the disorder. Her other brother is middle-aged and does not have the disease; so assume that he does not carry genes for it. A geneticist genotypes Sally and her immediate family for the *A* and *C* loci and obtains the genotypes shown on the following pedigree.



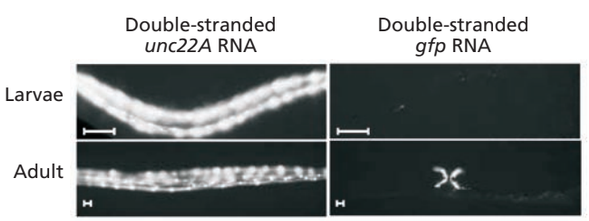
a. Assume that there is no crossing over between the *A, C,* and *G* loci. Does Sally carry the gene that causes G syndrome? Explain why or why not?

b. Draw the arrangement of the *A, C,* and *G* alleles on the chromosomes for all members of the family.

Section 19.5

***40.** You have discovered a gene in mice that is similar to a gene in yeast. How might you determine whether this gene is essential for development in mice?

41. DATA ANALYSIS Andrew Fire, Craig Mello, and their colleagues were among the first to examine the effects of double-stranded RNA on gene expression (A. Fire et al. 1998. *Nature* 391:806–811). In one experiment, they used a transgenic strain of *C. elegans* into which a gene (*gfp*) for a green florescent pigment had been introduced. They injected some worms with double-stranded RNA complementary to coding sequences of the *gfp* gene and injected other worms with double-stranded RNA complementary to the coding region of a different gene (*unc22C*) that encodes a muscle protein. The illustration below includes photographs of larvae and adult progeny of the injected worms. Green florescent pigment appears as bright spots in the photographs.



a. Explain these results.

b. Fire and Mello conducted another experiment in which they injected double-stranded RNA complementary to the introns and promoter sequences of the *gfp* gene. What results would you expect with this experiment? Explain your answer.

CHALLENGE QUESTIONS

Section 19.5

42. Suppose that you are hired by a biotechnology firm to produce a giant strain of fruit flies by using recombinant DNA technology so that genetics students will not be forced to strain their eyes when looking at tiny flies. You go to the library and learn that growth in fruit flies is normally inhibited by a hormone called shorty substance P (SSP). You decide that you can produce giant fruit flies if you can somehow turn off the production of SSP. Shorty substance P is synthesized from a compound called XSP in a single-step reaction catalyzed by the enzyme runtase:

$$\text{XSP} \xrightarrow[\text{Runtase}]{} \text{SSP}$$

A researcher has already isolated cDNA for runtase and has sequenced it, but the location of the runtase gene in the *Drosophila* genome is unknown.

In attempting to devise a strategy for turning off the production of SSP and producing giant flies by using standard recombinant DNA techniques, you discover that deleting, inactivating, or otherwise mutating this DNA sequence in *Drosophila* turns out to be extremely difficult. Therefore you must restrict your genetic engineering to gene augmentation (adding new genes to cells). Describe the methods that you will use to turn off SSP and produce giant flies by using recombinant DNA technology.

Section 19.6

43. Much of the controversy over genetically engineered foods has centered on whether special labeling should be required on all products made from genetically modified crops. Some people have advocated labeling that identifies the product as having been made from genetically modified plants. Others have argued that labeling should be required only to identify the ingredients, not the process by which they were produced. Take one side in this issue and justify your stand.