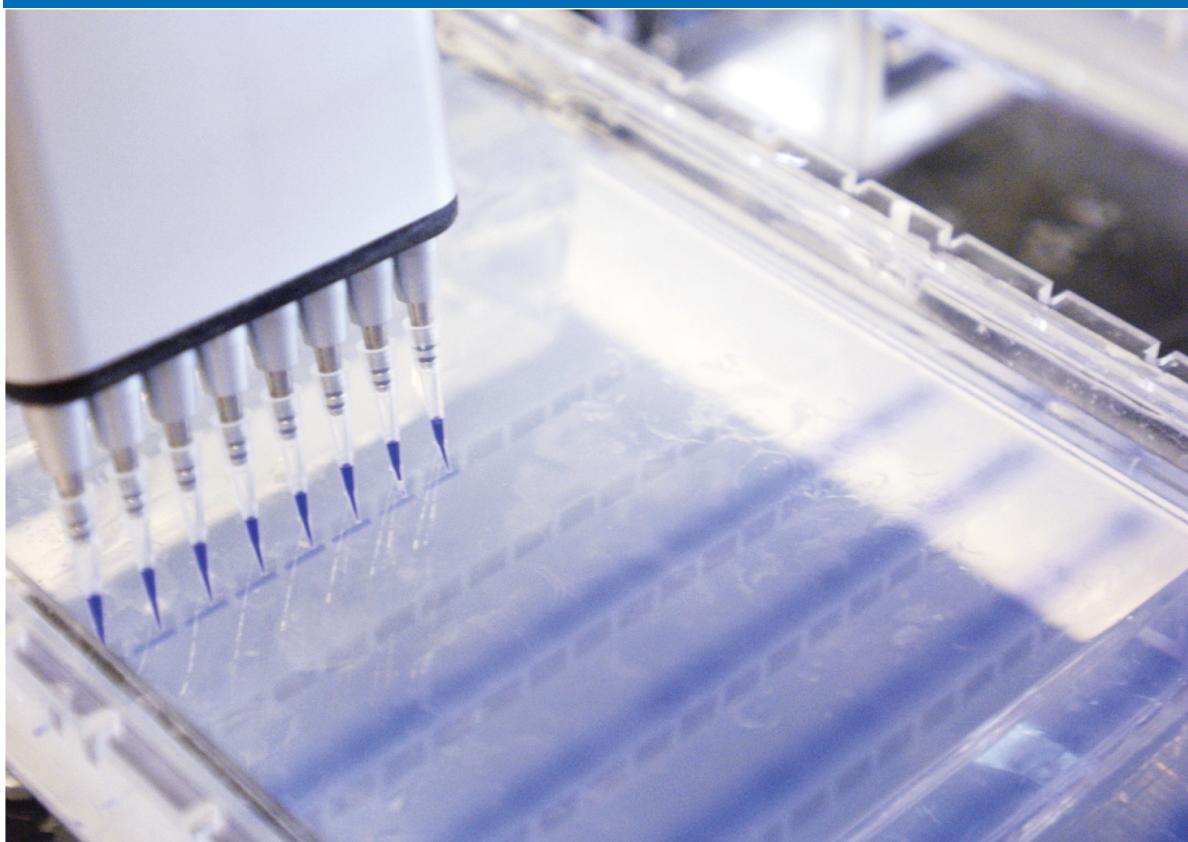


CHAPTER 10

Gene Isolation and Manipulation



SPL/Science Source.

Agarose gel electrophoresis is used to separate DNA fragments based on their size.

CHAPTER OUTLINE AND LEARNING OBJECTIVES

10.1 DETECTING AND QUANTIFYING DNA, RNA, AND PROTEIN

LO 10.1 Describe methods for detecting and quantifying specific nucleic acid and protein molecules *in vitro* and *in vivo*.

10.2 GENERATING RECOMBINANT DNA

LO 10.2 Describe the functional components of vectors that are useful for cloning DNA.

LO 10.3 Describe methods for generating and isolating recombinant DNA molecules.

10.3 SEQUENCING DNA

LO 10.4 Diagram the steps of dideoxy DNA sequencing.

10.4 ENGINEERING GENOMES

LO 10.5 Describe methods for generating transgenic organisms.

LO 10.6 Describe the CRISPR-Cas9 technique for precise engineering of genomes.

CHAPTER OBJECTIVE

In prior chapters, we saw that the genome contains the information needed to build and maintain an organism. Researchers are often interested in studying the function of one or a few genes in a genome. The main objective of this chapter is to present methods that are used to do this, including methods to detect and quantify RNAs, proteins, and specific regions of DNA as well as methods to alter the sequence and amounts of these molecules.

In this chapter, we describe experimental techniques used to isolate and manipulate genes and their products, RNAs and proteins. There are many good reasons for including this information in an introductory genetics textbook. To develop these techniques, researchers relied on knowledge of the chemical and functional properties of DNA, RNAs, and proteins as well as the mechanisms underlying fundamental molecular genetic processes such as DNA replication, transcription, and translation. Thus, an understanding of experimental techniques will reinforce principles that are presented in other chapters. It will also aid in the comprehension and evaluation of the primary research literature, as well as the design of experiments to address yet-to-be-solved genetic problems. Lastly, whether or not a problem can be solved is often determined by the techniques that are available. Consequently, some major advances in genetics have become possible only because of the development of a new technique. Standouts include techniques for isolating and manipulating fragments of DNA (DNA cloning), amplifying DNA (PCR), sequencing DNA (dideoxy sequencing), and introducing DNA into an organism (transgenesis). This chapter ends with recently developed techniques such as CRISPR-Cas9 for the precise engineering of genomes. New genome engineering techniques have made possible reverse genetic studies that aim to understand the function of a gene by analyzing the phenotypic consequences of altering the gene sequence or its expression.

Genes are the central focus of genetics, and so, clearly, it is desirable to isolate a gene of interest (or any DNA region) from the genome to study it. Isolating individual genes and producing enough copies of them to analyze can be a daunting task because a single gene is a tiny fraction of an entire genome. For example, the haploid human genome contains over 3 billion base pairs, whereas the coding region of an average gene contains only a few thousand base pairs. How do scientists find the proverbial needle in the haystack—the gene—and then produce suitable quantities of it for analysis?

Many investigations in genetics begin with the desire to study a trait or a disease. In [Chapter 2](#), we described forward genetic approaches to search for mutants that exhibit an altered phenotype and crosses or pedigree analysis to determine whether that phenotype is determined by a single gene.

In [Chapter 4](#), we discussed how mapping by recombination helps locate the gene at the DNA level. In this chapter, we continue by presenting molecular methods for identifying a gene of interest and studying its molecular function.

The first step in studying gene function is to isolate its DNA and reproduce it in quantities suitable for study. Just like a construction worker, a genetic engineer needs tools. Most toolboxes that we are familiar with are filled with tools like hammers, screwdrivers, and wrenches that are designed by people and manufactured in factories. In contrast, the tools of the genetic engineer are molecules isolated from cells. Most of these tools were the product of scientific discovery—where the objective was to answer a biological question. Only later did some scientists appreciate the potential practical value of these molecules and invent ways to put them to use with the goal of isolating and amplifying DNA fragments. As an example, one way to separate our gene of interest from the rest of the genome is to cut the genome with “molecular scissors” and isolate the small fragment containing the gene. Werner Arber discovered these molecular scissors, and for this discovery he was awarded the Nobel Prize in Physiology or Medicine in 1978. However, Arber was not looking for a tool to cut DNA precisely. Rather, he was trying to understand why some bacteria are resistant to infection by bacterial viruses. By answering this biological question, he discovered that resistant bacteria possess a previously unknown type of enzyme—a restriction endonuclease—that cuts DNA at specific sequences. As we will see in this chapter, restriction enzymes are one of the cornerstones of genetic engineering and a common tool found in the genetic engineer’s toolbox.

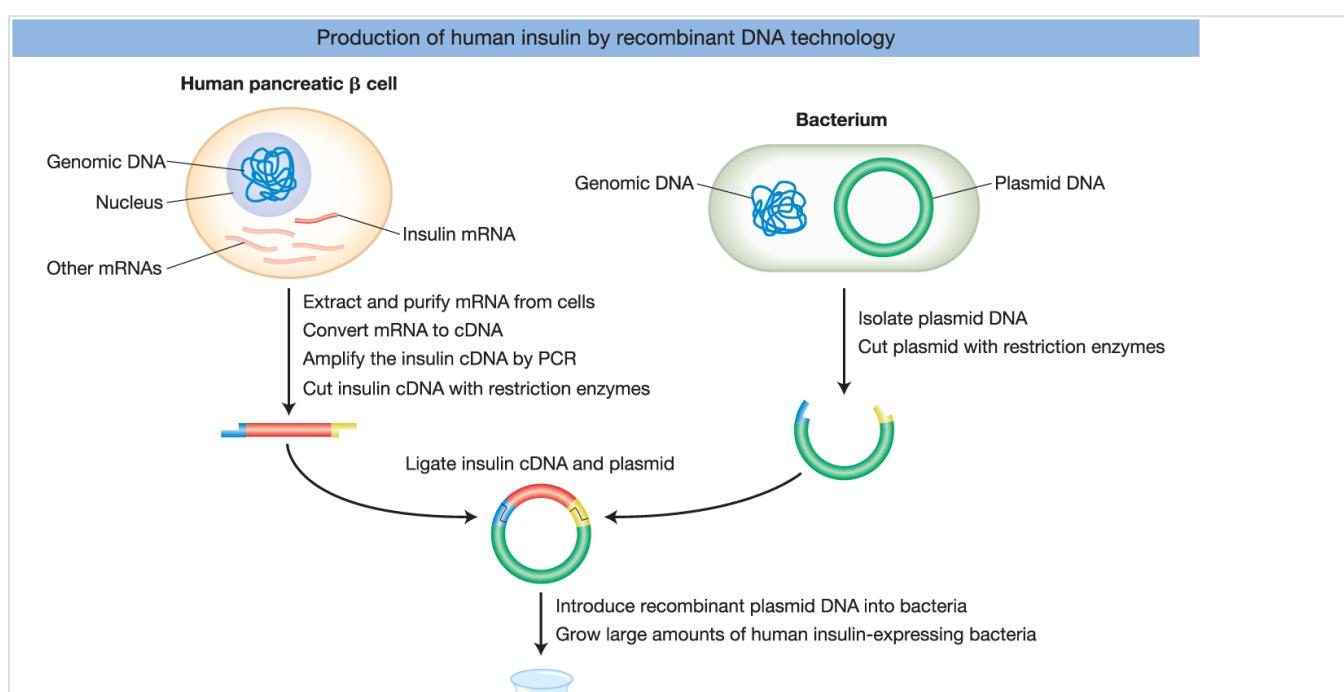
As another example, it is unlikely that anyone would have predicted that DNA polymerase, discovered by Arthur Kornberg, could be fashioned into two powerful tools for DNA isolation and analysis. To this day, many of the techniques used to determine the nucleotide sequence of DNA rely on synthesizing it with DNA polymerase. Similarly, most of the protocols used to isolate and amplify specific regions of DNA from sources as disparate as a crime scene or a fossil embedded in amber rely on the activity of DNA polymerase.

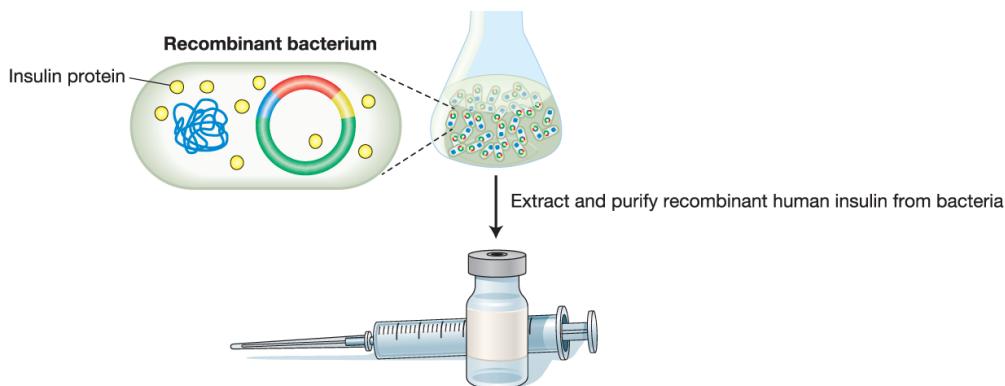
DNA technologies are the collective techniques for obtaining, amplifying, and manipulating specific DNA fragments. Since the mid-1970s, the development of DNA technologies has revolutionized the study of biology, opening many areas of research to molecular investigation.

Genetic engineering, the application of DNA technologies to specific biological, medical, or agricultural problems, is now a well-established branch of technology. **Genomics** is the ultimate extension of the technologies to the global analysis of the nucleic acids present in a nucleus, a cell, an organism, or a group of related species (see [Chapter 14](#)).

In this chapter, we will illustrate gene isolation and manipulation techniques through their application to the insulin gene. The insulin gene encodes a protein that functions to maintain normal blood glucose levels by promoting the uptake of glucose from blood into cells and by regulating the metabolism of carbohydrates, lipids, and proteins. In mammals, including humans and mice, the insulin gene is present in the genome of all cells, but it is expressed (i.e., transcribed and translated) only in beta (β) cells in the pancreas. Diabetes is a human disease in which blood glucose levels are abnormally high, either because β cells do not produce enough insulin (type I diabetes) or because cells are unable to respond to insulin (type II diabetes). Mild forms of type I diabetes can be treated by dietary restrictions, but for many patients, daily insulin injections are necessary.

Until about 35 years ago, cows were the major source of insulin protein. The protein was harvested from pancreases of animals processed in meatpacking plants and purified on a large scale to eliminate the majority of proteins and other contaminants in pancreas extracts. It took about 8,000 pounds of pancreas from 23,500 animals to purify one pound of insulin. Then, in 1982, the first recombinant human insulin came on the market. Because it could be produced on an industrial scale in bacteria by recombinant DNA techniques using the human gene sequence, insulin could be made in a purer form and at a lower cost than the previous method. We will use the generation of recombinant human insulin as an example of the general steps necessary for making any recombinant DNA molecule. These steps are summarized in [Figure 10-1](#). The uses of recombinant DNA technologies are quite broad, ranging from gene isolation for basic biological research to gene therapy to treat human disease to the production of herbicides and pesticides for crop plants.





Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020 W. H. Freeman and Company

FIGURE 10-1 Recombinant human insulin is currently being produced in bacteria. The first step in this process is the construction of an expression plasmid that contains a human insulin cDNA insert. The plasmid is then transformed into bacteria, and the bacteria are grown in large quantities. As the recombinant bacteria grow, they transcribe insulin mRNA and translate the mRNA into insulin protein. Finally, the bacteria are harvested, and the insulin protein is extracted and purified for therapeutic use in humans.

Fr

10.1 DETECTING AND QUANTIFYING DNA, RNA, AND PROTEIN

LO 10.1 Describe methods for detecting and quantifying specific nucleic acid and protein molecules in vitro and in vivo.

The ability to detect and quantify DNA, RNA, and protein in vivo and in vitro is an essential part of investigating the function of these molecules in normal and disease states. For example, we might want to know whether the size of the insulin gene, mRNA, or protein varies among human populations. Alternatively, we might want to determine whether a similar insulin gene, mRNA, or protein is present in other organisms such as mice. Techniques described in this section and summarized in [Figure 10-2](#) were used to determine that the human genome contains a single insulin gene (*Ins*) on chromosome 11. Transcription of *Ins* produces a 1431-nucleotide pre-mRNA that contains three exons and two introns ([Figure 10-3a](#), top). Translation of the spliced *Ins* mRNA produces a protein of 110 amino acids. In addition, other animals have insulin genes that are similar in sequence to human *Ins*. For example, mice have two insulin genes, *Ins1* and *Ins2*, on chromosomes 19 and 7, respectively ([Figure 10-3a](#), middle and bottom). The open reading frame of the human *Ins* mRNA is 81 percent identical in nucleotide sequence to the mouse *Ins1* mRNA and 83 percent to *Ins2*, and the human Ins protein is 78 percent identical in amino acid sequence to the mouse *Ins1* protein and 82 percent to *Ins2*. As we move through the chapter, consider how the methods presented could be used to isolate and manipulate insulin genes in other animals such as the fruit fly *Drosophila melanogaster* that lack a pancreas and β cells.

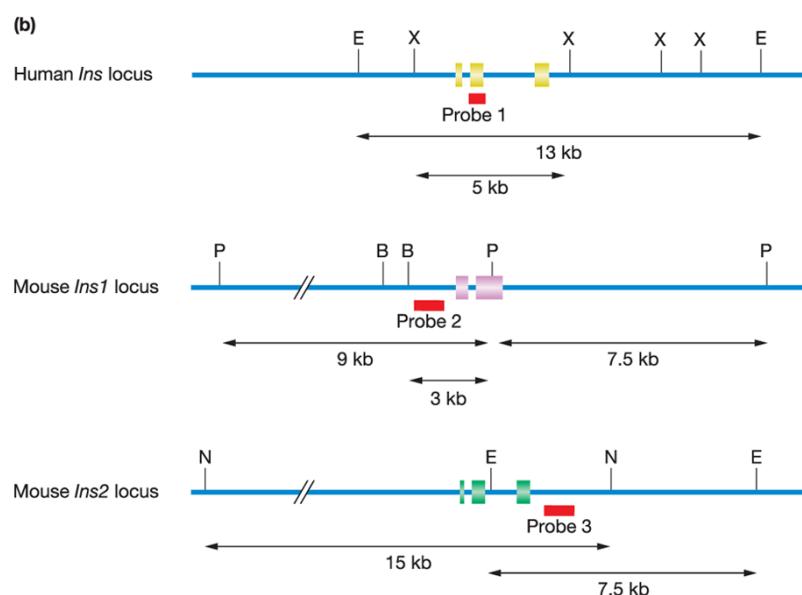
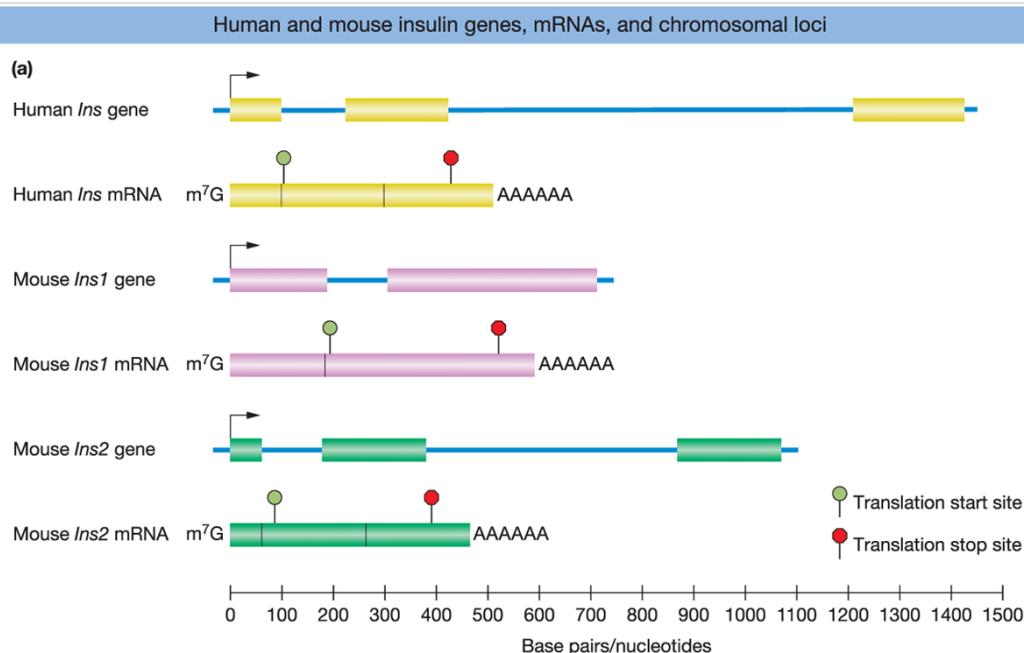
Methods for detecting and quantifying DNA, RNA, and protein

	DNA	RNA	Protein
In vitro	Southern blot Probe: DNA or RNA fragment Polymerase chain reaction (PCR) Probe: DNA primers	Northern blot Probe: DNA or RNA fragment Reverse transcription-PCR (RT-PCR) Probe: DNA primers	Western blot Probe: Antibody
In vivo	Fluorescence in situ hybridization (FISH) Probe: DNA or RNA fragment	In situ hybridization Probe: DNA or RNA fragment	Immunofluorescence Probe: Antibody

Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020 W. H. Freeman and

Company

FIGURE 10-2 A summary of the main methods used to detect and quantify specific DNA regions, RNAs, and proteins in vitro (i.e., after purification from cells) and in vivo (i.e., in cells, tissues, and whole organisms).



Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020 W. H. Freeman and Company

FIGURE 10-3 (a) Structures of insulin genes and mRNAs in humans and mice. Boxes represent exons. The mRNAs are spliced, capped at the 5' end with m^7G , and polyadenylated at the 3' end ([Chapter 8](#)). (b) Structures of insulin gene loci in humans and mice. Red lines indicate regions used as probes for Southern and Northern blot analyses. Results for probe 1 are shown in [Figure 10-6](#), and questions related to probes 2 and 3 are in the Working with the Figures section at the end of the chapter. Restriction enzymes used for Southern blot analysis are indicated by single letters: B = *Bam*HI, E = *Eco*RI, N = *Nsi*I, P = *Pvu*II, and X = *Xba*I (restriction sites for these enzymes are shown in [Table 10-1](#)). Lines with arrows indicate the sizes of some restriction fragments.

Detecting and quantifying molecules by Southern, Northern, and Western blot analysis

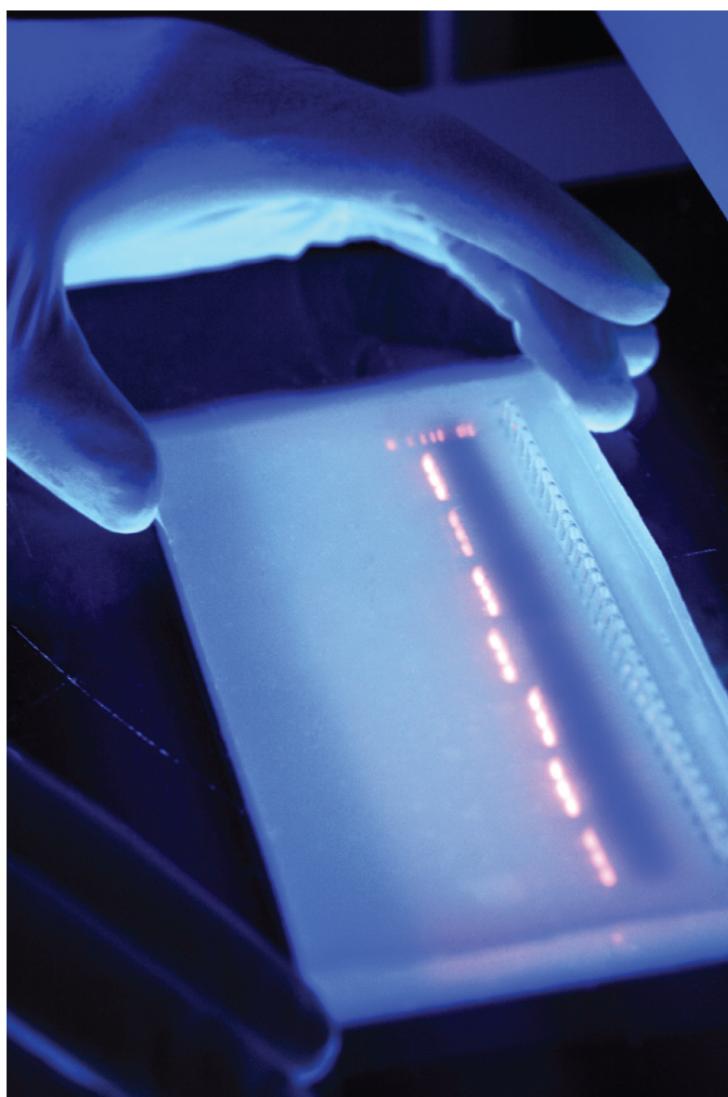
Blotting is a commonly used *in vitro* method to detect and quantify a specific DNA, RNA, or protein molecule within a mixture of many different DNA, RNA, or protein molecules. Blotting for DNA is called **Southern blotting** because the technique was developed by Edwin Southern. Similar blotting techniques for RNA and protein were invented later, and researchers could not resist the temptation to call blotting for RNA **Northern blotting** and blotting for protein **Western blotting**.

Blotting starts with **gel electrophoresis** to separate molecules in a mixture based on their physical properties such as size and charge. The term “gel” refers to the matrix used to separate molecules. Usually, agarose gels are used to separate DNA fragments, whereas polyacrylamide gels are used to separate RNAs as well as proteins. Agarose is a polysaccharide polymer extracted from seaweed. Agarose gels are made by melting agarose powder in a hot buffer and cooling the solution in a rectangular tray to form a slab of agarose that is similar to Jell-O. In contrast, polyacrylamide gels are produced by polymerization of acrylamide and a cross-linker such as bis-acrylamide between two glass plates. Wells that hold experimental samples are formed when agarose hardens or acrylamide polymerizes around square teeth of a comb that is set into the tray or between the glass plates.

The term “electrophoresis” refers to the voltage that is applied to gels that are submerged in a buffer solution. Gels are oriented with electrodes at the top and bottom. The cathode (negative charge) is at the top of the gel, where samples are loaded into wells, and the anode (positive charge) is at the bottom of the gel. Because of their negatively charged phosphate backbone, DNA and RNA migrate out of the wells toward the positive charge (opposite charges attract each other). Smaller nucleic acid molecules move faster through gels than larger ones, so, after electrophoresis, molecules are separated by size; larger molecules are near the top of the gel and smaller ones are near the bottom. Molecules of the same size will all migrate the same distance and form a band in the gel. Bands can be visualized by staining gels with dyes such as coomassie blue for proteins and ethidium bromide for DNA. Ethidium bromide binds DNA by intercalating between base pairs, and enables DNA to fluoresce when exposed to ultraviolet (UV) light (**Figure 10-4**). The size of the molecules within each band in the gel can be determined by comparing a band’s migration distance with a set of standard molecules of known sizes (also known as size markers). If the bands are well separated, an individual band can be cut from the gel, and the DNA sample can be purified from the gel matrix. Therefore, DNA electrophoresis can be either

diagnostic (showing sizes and relative amounts of DNA fragments present) or preparative (useful in isolating specific DNA fragments).

Gel electrophoresis



SPL/Science Source.

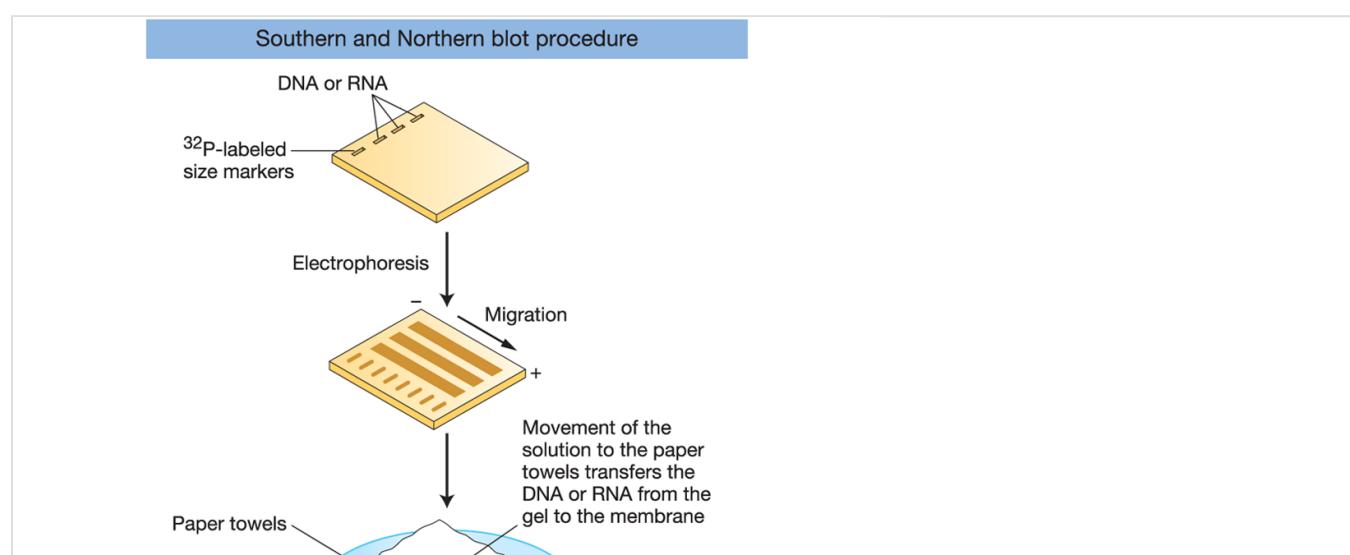
FIGURE 10-4 Agarose gel electrophoresis is used to separate DNA fragments based on size. After the DNA fragments are separated, they are stained with ethidium bromide and visualized with ultraviolet (UV) light. The visualized pink/white lines are bands of DNA of a particular size.

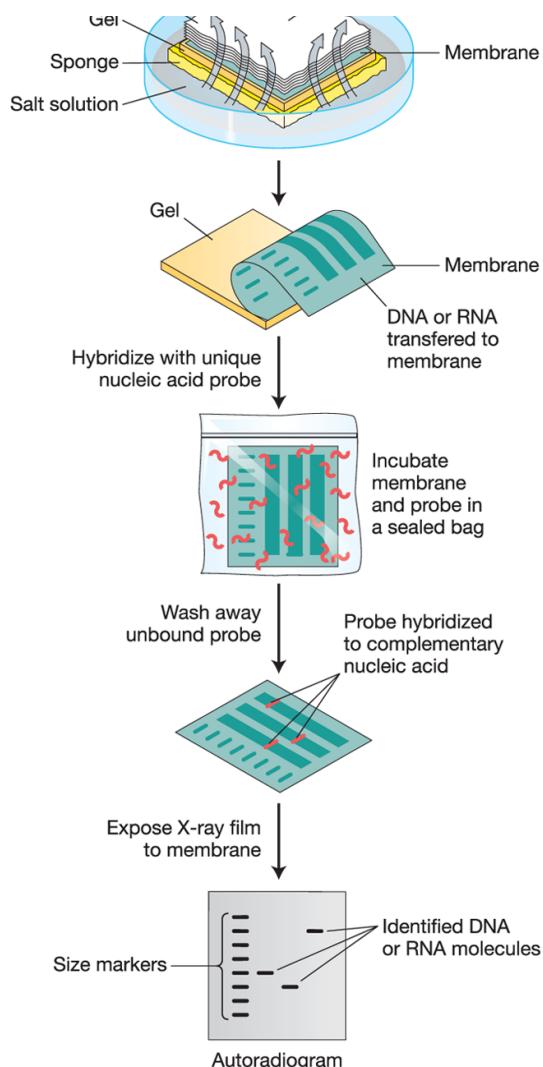
Gel electrophoresis can be carried out under non-denaturing or denaturing conditions. Non-denaturing conditions maintain the higher-order structures of molecules, including base pairing between DNA strands, base pairing in folded single-stranded RNAs, secondary and tertiary structures in folded proteins, and interactions between molecules. Typically, agarose gel electrophoresis of DNA molecules is carried out under non-denaturing conditions to maintain the double-stranded structure of DNA, whereas polyacrylamide gel electrophoresis of RNAs and proteins is carried out under denaturing conditions to eliminate higher-order structures and allow

molecules to migrate true to size. For example, denaturing SDS-PAGE (sodium dodecyl sulfate-polyacrylamide gel electrophoresis) is typically used to separate proteins. SDS is a detergent that not only disrupts protein folding, but also uniformly coats proteins with a negative charge, blocking the charges on amino acid R-groups, so that the rate of migration of a protein toward the positive anode depends on its molecular weight, and not on its amino acid composition.

KEY CONCEPT Gel electrophoresis separates complex mixtures of DNA fragments, RNAs, or proteins, based on size and charge.

After gel electrophoresis is used to separate complex mixtures of DNA fragments, RNAs, or proteins, the next step in blotting is to transfer the molecules from the gel to a special type of paper called a membrane that has high affinity for these molecules. The transfer procedure maintains the molecules' positions relative to each other on the membrane, just as they were in the gel. Transfer is carried out either by capillary action, as illustrated in [Figure 10-5](#), or by electrophoresis. At this point, the transferred molecules on the membrane are invisible to the naked eye. The last step in blotting is to use a [probe](#) to visualize a specific molecule on the membrane. In Southern and Northern blotting, probes are radioactive (^{32}P -labeled) single-stranded nucleic acids that are complementary to the nucleic acid of interest. When the membrane is incubated with a solution containing the probe, the probe anneals with complementary nucleic acid sequences bound to the membrane. This annealing process is commonly called [hybridization](#). Unbound probe is washed away, and places where the probe has hybridized are revealed by [autoradiography](#), that is, exposing the membrane to X-ray film. Because hybridization requires single-stranded molecules, Southern blotting has an extra step in which the gel is soaked in an alkaline solution such as NaOH (sodium hydroxide) to denature the double-stranded DNA into single-stranded DNA prior to the membrane transfer step.





Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020
W. H. Freeman and Company

FIGURE 10-5 Southern and Northern blotting procedures are similar. The key difference is that in Southern blotting, DNA is transferred to the membrane, while in Northern blotting, RNA is transferred to the membrane. Western blotting for proteins is carried out by a comparable procedure, except that the probe is an antibody rather than a radioactive nucleic acid, and the size markers are proteins of different sizes.

Stu ANIMATED ART Sapling Plus

Northern blot analysis

In Western blotting, probes are **antibodies** that bind specific proteins and are detected in a variety of ways, including light emitting chemiluminescence or fluorescence. Antibodies are proteins made by the immune system of some animals that bind foreign substances called antigens with high affinity. One way to produce an antibody is to inject a large amount of an antigen (e.g., a

protein of interest) into an animal (usually rabbits or chickens, but sometimes larger animals such as goats), allow time for the animal to raise an immune response to the antigen, and then, from the blood of the animal, collect serum that contains the antibody.

KEY CONCEPT In Southern blotting, the material that is transferred to the membrane is DNA, while in Northern blotting it is RNA, and in Western blotting it is protein.

Techniques analogous to Southern, Northern, and Western blotting are used to detect DNA, RNAs, and proteins *in vivo* in cultured cells or in whole organisms. Detection of DNA and RNA *in vivo* is carried out by hybridization with radioactive single-stranded nucleic acid probes, followed by autoradiography. This technique is called **in situ hybridization (ISH)**. A modified form of ISH that uses fluorescently labeled probes rather than radioactive probes is called **fluorescence in situ hybridization (FISH)** (for example, see [Figure 13-6](#)). Proteins are detected *in vivo* by **immunofluorescence**, which uses antibody probes (for example, see [Figure 13-15](#)). A fluorescence microscope is used to reveal the location and abundance of the fluorescence signal in FISH and immunofluorescence.

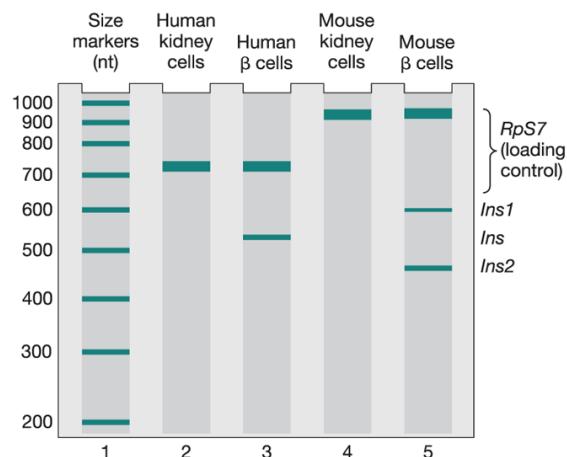
KEY CONCEPT Specific DNA fragments or RNAs are detected *in vitro* and *in vivo* by hybridization with nucleic acid probes, and specific proteins are detected by interaction with antibody probes.

The expected results of Northern and Western blot analyses for insulin are shown in [Figures 10-6a](#) and [10-6b](#), respectively. Northern blot analysis of the thousands of different mRNAs expressed in human cells with a probe complementary to the human *Ins* mRNA should detect one band, that is, a single type of insulin mRNA in β cells but not in kidney cells ([Figure 10-6a](#), lanes 2 and 3). Because the nucleotide sequence of the insulin mRNA is very similar between humans and mice, the human probe should also detect the two insulin mRNAs (*Ins1* and *Ins2*) in mouse β cells ([Figure 10-6a](#), lane 5). Standardly, blots are probed not only for a molecule of interest, but also for other molecules that serve to confirm that the experiment worked as expected or that serve as the basis for comparison within or between samples, which is referred to as a loading, normalization, or specificity control. In this case, analysis of *ribosomal protein S7 (RpS7)* mRNA, a transcript that is present in all cells at similar levels, shows that the failure to detect insulin mRNA in the kidney cell samples ([Figure 10-6a](#), lanes 2 and 4) was not due to a problem with the experiment. Furthermore, the intensity of the bands provides information about the abundance of the mRNAs. The data show that, in human β cells, *Ins* mRNA is less abundant than *RpS7* mRNA, and, in

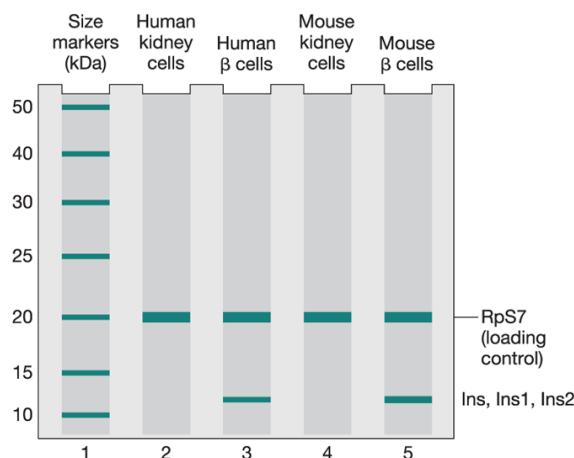
mouse β cells, *Ins2* is more abundant than *Ins1*. Western blot analysis of proteins from the same types of cells with an antibody to human insulin shows that cell type-specific expression of the insulin protein is similar to that of the insulin mRNA ([Figure 10-6b](#)). The expected size of a protein on a Western blot can be estimated based on the average molecular weight of an amino acid, 110 Daltons (Da). One Da is one gram per mole, and one kilodalton (kDa) is 1000 grams per mole. Thus, the 110-amino-acid insulin protein is about 12 kDa (110 amino acids \times 110 Da/amino acid). Note that only one insulin protein band should be detected in mouse β cells because the *Ins1* and *Ins2* proteins are the same size and, thus, should migrate to the same position upon gel electrophoresis.

Northern, Western, and Southern blots for insulin

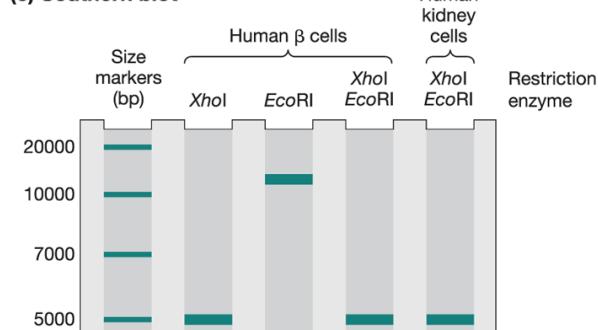
(a) Northern blot

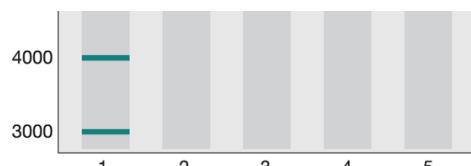


(b) Western blot



(c) Southern blot





Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020 W. H. Freeman and Company

FIGURE 10-6 (a) Northern, (b) Western, (c) and Southern blot analyses for insulin in different human and mouse tissues.

The Southern and Northern blots were analyzed with probe 1 to exon 2 of the human *Ins* gene (see [Figure 10-3b](#)). The Northern blot was also analyzed with a probe to the *RpS7* mRNA. The Western blot was probed with antibodies to human insulin and *RpS7* proteins.



Western blot analysis

Sometimes, the starting material for Southern blot analysis is chromosome-sized DNA molecules of genomic DNA. Such large molecules can be analyzed more precisely when cut into fragments of much smaller size. Cutting is carried out by bacterial **restriction enzymes**. These enzymes are endonucleases that cleave phosphodiester bonds between nucleotides at specific DNA sequences, called **restriction sites**, that are usually 4 to 8 base pairs long. [Table 10-1](#) shows several examples of restriction enzymes and their restriction sites. Restriction sites are **palindromic**, which means that both strands have the same nucleotide sequence but in antiparallel orientation (the same sequence reads 5' to 3' on each strand). For example, the restriction enzyme *EcoRI* recognizes the

sequence $\begin{matrix} 5'-\overset{\downarrow}{G}AATT\overset{\downarrow}{C}-3' \\ 3'-CTTAA\overset{\uparrow}{G}-5' \end{matrix}$ and cleaves the bond between G and A on each strand.

Restriction enzyme names are based on the organism in which they were discovered. For example, the enzyme *EcoRI* was discovered in *E. coli*, which also explains why the first part of a restriction enzyme name is italicized. To date, approximately 3000 restriction enzymes have been identified that recognize over 230 restriction sites. Some restriction enzymes cut in the same position on each strand, leaving *blunt* ends (e.g., *MspI*; see [Table 10-1](#)), while others make cuts that are offset, producing *staggered* ends (e.g., *EcoRI*). Thus, a restriction enzyme will cut the DNA from any organism into a set of **restriction fragments** determined by the locations of restriction sites in the DNA, and will produce the same pattern of fragments every time that DNA is cut.

TABLE 10-1 Restriction enzymes

Restriction enzyme	Source bacterium	Restriction site	Length (bp)	Staggered (S) Blunt (B)
BamHI	<i>B. amyloliquefaciens</i>		6	S
EcoRI	<i>E. coli</i>		6	S
MspI	<i>Moraxella sp.</i>		4	B
NorI	<i>N. otitidis</i>		8	S
NsiI	<i>N. sicca</i>		6	S
PvuII	<i>P. vulgaris</i>		6	B
XbaI	<i>X. holcicola</i>		6	S
Arrows indicate sites of cleavage.				
Th				

KEY CONCEPT Restriction enzymes cut DNA at specific sequences, producing fragments with staggered or blunt ends.

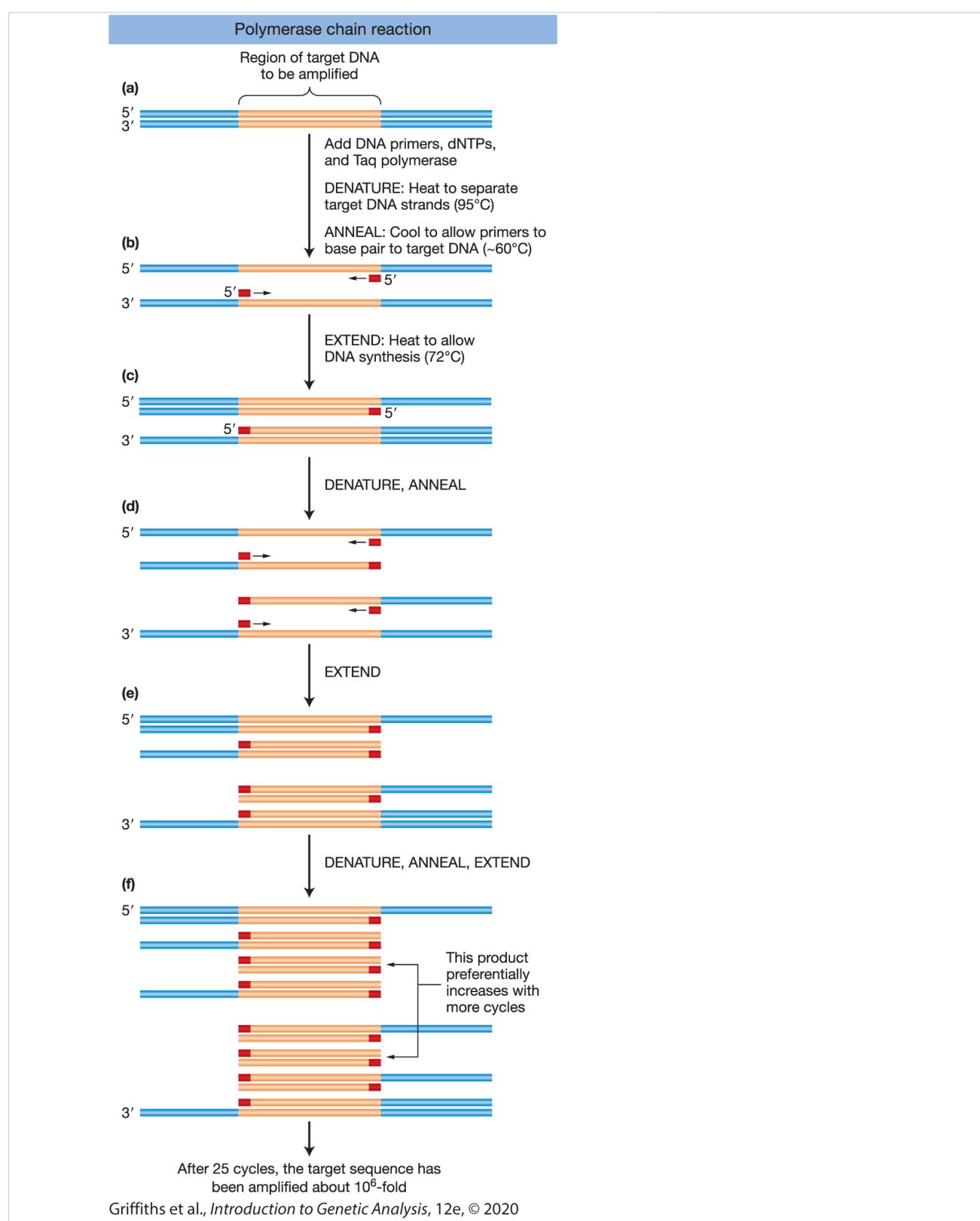
The expected results of Southern blot analysis for the insulin gene are shown in [Figure 10-6c](#). Each sample on the blot contains genomic DNA that was digested with restriction enzymes. If the sequence of the 3×10^9 base pair human genome was completely random, the 6-base pair restriction site for EcoRI should occur every 4096 base pairs (4^6 , the number of possible base pairs at each position in the restriction site^{the number of base pairs in the restriction site}), which means that digestion of the human genome would produce about 730,000 EcoRI fragments ($3 \times 10^9/4096$). Based on [restriction maps](#) of genomic DNA, as shown in [Figure 10-3b](#), the probe complementary to the insulin gene should detect a single EcoRI band out of the estimated 730,000 EcoRI bands ([Figure 10-6c](#), lane 3). The same probe should detect a smaller band when the DNA is digested with the restriction enzyme XbaI or with both EcoRI and XbaI ([Figure 10-6c](#), lanes 2 and 4), because the XbaI fragment is located within the EcoRI fragment. Southern blot analysis of DNA from β cells and kidney cells produces the same result ([Figure 10-6c](#), lanes 4 and 5) because all cells in an organism have the same genomic DNA. Whereas Northern and Western blot analysis of β cells and kidney cells produces different results ([Figures 10-6a](#) and [10-6b](#)), because insulin is expressed only in β cells.

Detecting and amplifying DNA by the polymerase chain reaction

In 1985, the ability of researchers to analyze and manipulate DNA was transformed by the invention of the **polymerase chain reaction (PCR)** by Kary Mullis. PCR makes it possible to produce billions of copies of a specific DNA sequence starting with only one copy. This is called **DNA amplification**. To develop PCR, Mullis brought together several pieces of information obtained through basic research. First, he knew from studies of DNA replication that DNA polymerases copy a single-stranded DNA template by extending off the 3' end of an annealed primer. Second, he knew from techniques such as Southern blotting that, in solution, a short piece of single-stranded DNA called an oligonucleotide will specifically anneal to DNA sequences that are perfectly complementary. Third, he knew that high temperatures disrupt hydrogen bonds between bases in double-stranded DNA to produce single-stranded DNA. Lastly, he knew that about 20 years earlier, the microbiologist Thomas Brock was able to culture a bacterium *Thermus aquaticus* that grows at high temperatures in the hot springs of Yellowstone National Park. The DNA polymerase from this bacterium, called *Taq* polymerase, not only is active at high temperatures but also remains active over many cycles of heating and cooling. By putting these pieces of information together, Mullis constructed a simple method to amplify any DNA sequence *in vitro*.

The basic strategy of PCR is outlined in [Figure 10-7](#). The process uses a pair of chemically synthesized oligonucleotide DNA primers that are each about 20 nucleotides long. Each primer is designed to base pair to one end of the target gene or region to be amplified, such that the primers base pair to opposite DNA strands with their 3' ends pointing toward each other. The primers are added to a solution containing the DNA template (e.g., genomic DNA), the four deoxyribonucleoside triphosphates (dATP, dCTP, dGTP, and dTTP) required for DNA synthesis ([Figure 7-5](#)), and the heat-stable *Taq* DNA polymerase. The DNA template is denatured by heat (95°C), resulting in single-stranded DNA molecules. Upon cooling the reaction to between 50°C and 65°C, the primers anneal to their complementary sequences in the single-stranded DNA template. After the temperature is raised to 72°C, *Taq* polymerase replicates the single-stranded DNA segments by extending from the annealed primers. Complementary new strands are synthesized, as in DNA replication in cells, forming two double-stranded DNA molecules identical to the single parental double-stranded DNA molecule. Thus, one cycle of PCR consists

of three main steps (denaturing, annealing, and extending) and results in doubling the starting amount of the target sequence. Subsequent cycles of denaturing, annealing, and extending also double the amount of the target sequence, resulting in an exponential increase ($2^{\text{number of cycles}}$) in the number of copies of DNA. Thus, a typical PCR with 30 cycles of five minutes each will amplify the target DNA about one billion-fold (2^{30}) in 2.5 hours.



Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020

W. H. Freeman and Company

FIGURE 10-7 The polymerase chain reaction rapidly synthesizes many copies of a target DNA sequence. (a) Double-stranded DNA (blue) containing the target sequence (orange). (b) Addition of *Taq* polymerase, deoxyribonucleotides, and two DNA primers that have sequences complementary to the 3' ends of the two strands of the target DNA. The strands are denatured (separated) by heating and then cooled to allow the primers to anneal to the target DNA. (c) After the temperature is raised, *Taq* polymerase synthesizes the first set of complementary strands. These first two strands are of varying length because the template extends beyond the site of binding by the other primer. (d) The two duplexes are heated again, exposing four binding sites, two for each primer. After cooling, the two primers again bind to their respective strands at the 3' ends of the target region. (e) After the temperature is raised, *Taq* polymerase synthesizes four complementary strands. Although the template strands at this stage are variable in length, two of the four strands just synthesized from them are precisely the length of the target sequence desired. This precise length is achieved because each of these strands begins at the primer-binding site, at one end of the target sequence, and proceeds until it runs out of template, at the other end of the sequence. (f) The process is repeated for many cycles, each time creating twice as many double-stranded DNA molecules that are identical to the target sequence.

Pa ANIMATED ART  Sapling Plus

Polymerase chain reaction

PCR is a powerful technique that is routinely used to isolate specific regions of DNA when there is prior knowledge of the sequence to be amplified. What makes PCR so powerful is that only small amounts of starting material are needed, which makes it possible to work with DNA samples that are difficult to obtain, such as from a small number of tumor cells. There are applications for PCR in genotyping, sequencing, cloning, paternity testing, forensics, molecular archeology, detection of infectious diseases, and many other clinical and basic research efforts that involve DNA. In recognition of the importance of PCR, Kary Mullis was awarded the Nobel Prize in Chemistry in 1993.

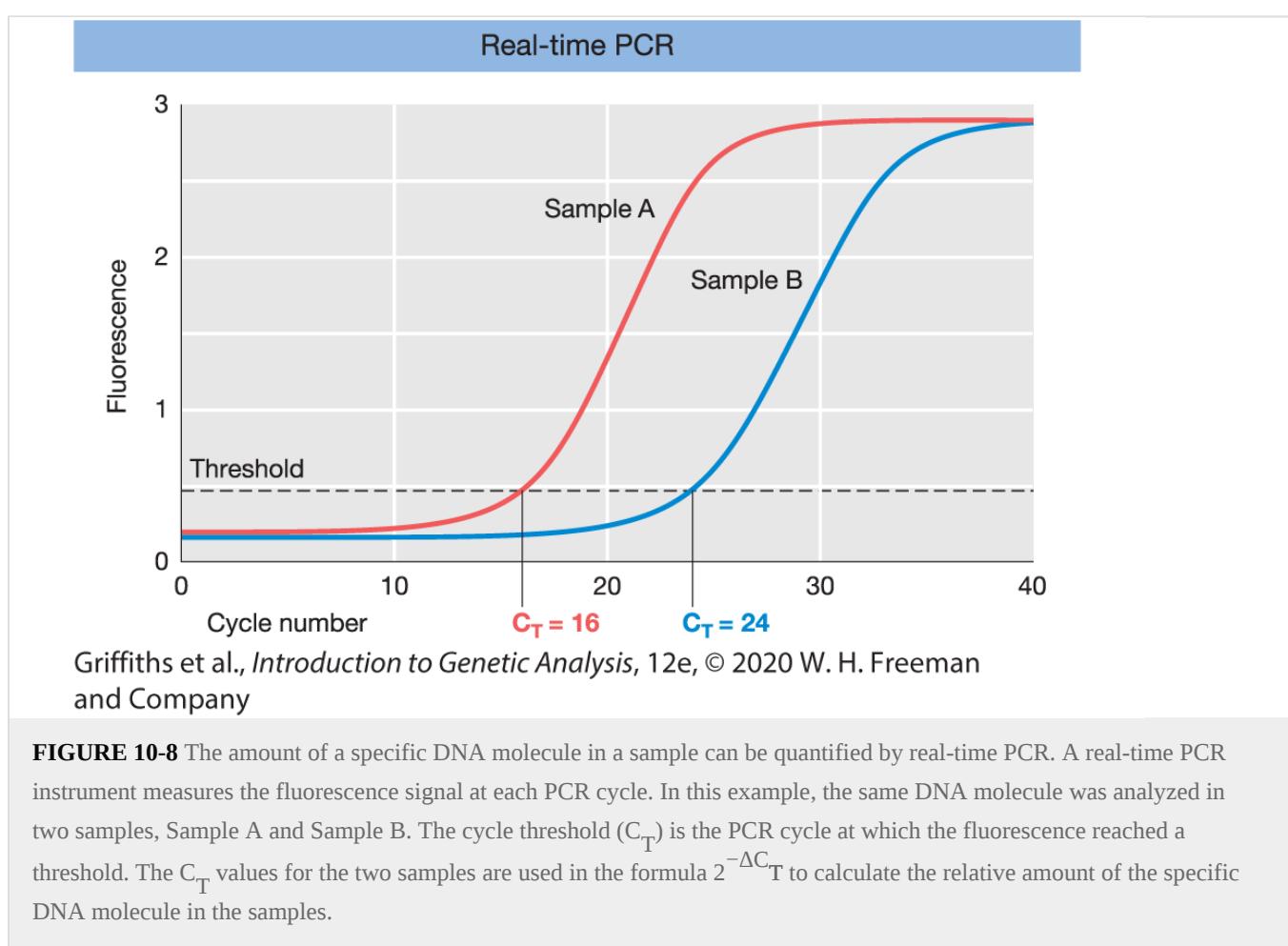
KEY CONCEPT The polymerase chain reaction uses specially designed primers to amplify specific regions of DNA in a test tube.

Quantifying DNA by real-time PCR

Since the amount of DNA produced by PCR doubles with each cycle, it is possible to calculate the amount of DNA in a sample based on the amount of DNA produced after a given number of PCR cycles. Quantification of DNA by PCR, called **quantitative PCR (qPCR)**, is automated by real-time PCR instruments that measure the amount of DNA product in “real-time” during each PCR cycle. To be more exact, these instruments measure the intensity of a fluorescent signal generated

by a dye called SYBR green, which, like ethidium bromide, intercalates into double-stranded DNA. As the amount of DNA increases with each PCR cycle, the fluorescent signal increases (**Figure 10-8**). The number of PCR cycles that it takes for the fluorescent signal to be detected above background is called the C_T (cycle threshold) value. If two samples have C_T values of 16 and 24, it would mean that the second sample had 256-fold less DNA than the first sample because eight more cycles were required to attain a fluorescent signal above background. In other words, the fold difference in the amount of DNA in the two samples is equal to $2^{-\Delta C_T}$. So, in this case,

$$2^{-(16-24)} = 2^{8} = 256$$



THE ANIMATED ART Sapling Plus

Real-time qPCR

qPCR is often used to compare the relative amounts of different DNA molecules in a single sample. For example, to diagnose cancer due a somatic mutation, qPCR is used to determine the

fraction of cells in a tumor sample that contain the mutant gene versus the wild-type gene. qPCR is also used to compare the relative amount of the same DNA molecule in different samples. For example, to determine the rate of progression of a viral infection, qPCR is used to compare the amount of viral DNA in blood samples that are collected at different times.

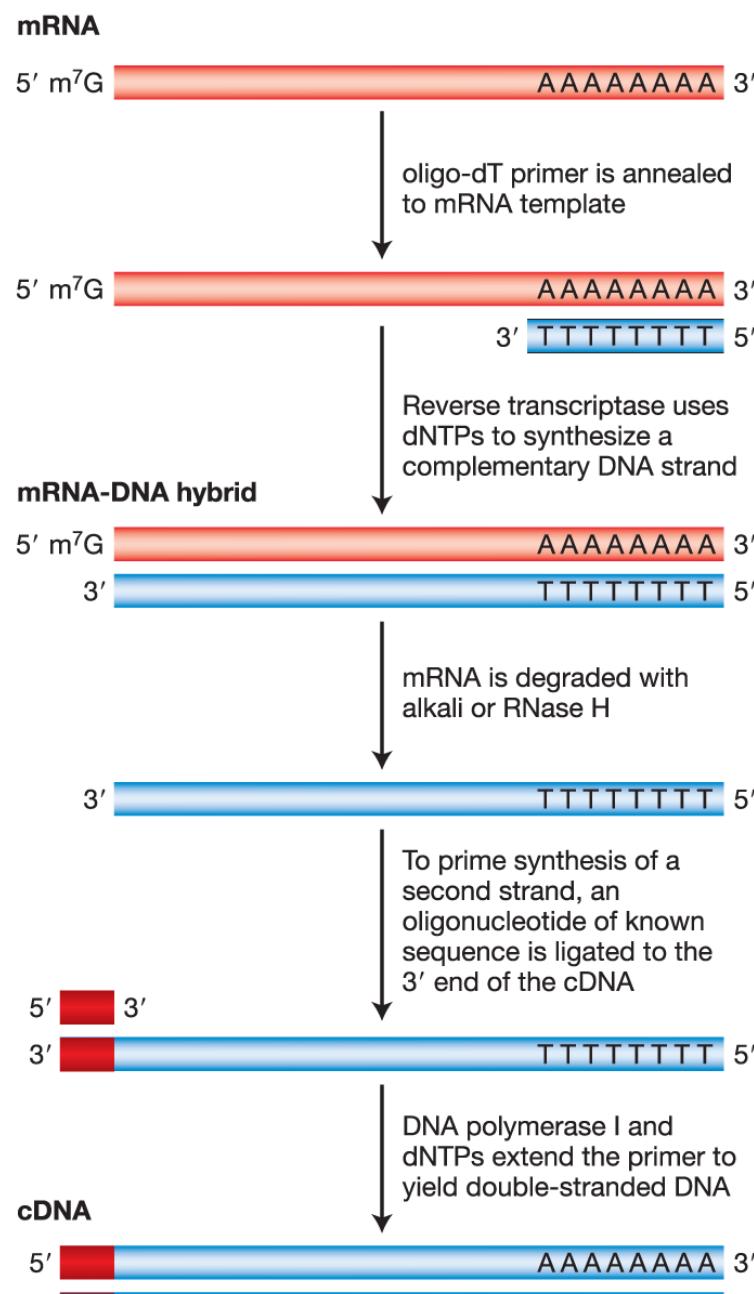
KEY CONCEPT Quantitative PCR (qPCR) is a method that uses a real-time PCR instrument to determine the amount of a specific DNA molecule in a sample.

Detecting and quantifying mRNA by reverse-transcription PCR

PCR can also be used to detect, amplify, and quantify mRNA; however, single-stranded RNA must first be converted into double-stranded DNA. **Complementary DNA (cDNA)** is a double-stranded DNA version of an mRNA molecule. cDNA is made from mRNA *in vitro* by a special enzyme called **reverse transcriptase**, originally isolated from retroviruses (see [Chapter 16](#)). Reverse transcriptase is a type of DNA polymerase that synthesizes a DNA strand complementary to an RNA template. Retroviruses such as human immunodeficiency virus (HIV) use reverse transcriptase to convert their RNA genomes into DNA as part of their replication cycle.

cDNA synthesis begins with the purification of mRNA from a tissue or specific cells such as the pancreas or β cells. Purification is necessary because mRNA accounts for only about 5 percent of the total amount of cellular RNA, with rRNA accounting for about 80 percent and tRNA for 15 percent. mRNAs from eukaryotic cells are commonly purified using affinity methods that target the unique features of mRNA relative to other types of RNA, that is, the 5'-m⁷G cap and the 3' poly(A) tail. Next, purified mRNA is incubated with reverse transcriptase, the four dNTPs, and an oligo-dT primer (an oligonucleotide of about 20 T residues) ([Figure 10-9](#)). The oligo-dT primer anneals to the poly(A) tail of the mRNA, and, using the mRNA as a template, reverse transcriptase synthesizes single-stranded DNA starting from the oligo-dT primer and ending at the m⁷G cap. This is called first-strand cDNA synthesis. The RNA strand of the RNA-DNA hybrid is then removed by alkaline hydrolysis (as described in [Chapter 8](#), RNA is susceptible to base-catalyzed hydrolysis because of the hydroxyl group at the ribose sugar 2' position) or by RNase H, an enzyme that cuts the RNA strand of an RNA-DNA hybrid. Second-strand cDNA synthesis is carried out by *E. coli* DNA polymerase I and primed by ligation of an oligonucleotide of known sequence to the 3' end of the first-strand cDNA.

cDNA synthesis by reverse transcription



Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020
W. H. Freeman and Company

FIGURE 10-9 mRNA is converted to cDNA by the enzymes reverse transcriptase and DNA polymerase I. Reverse transcriptase first synthesizes a single-stranded DNA molecule using the mRNA as a template. Then DNA polymerase I synthesizes a double-stranded DNA molecule (cDNA) using the single-stranded DNA as a template.

KEY CONCEPT Reverse transcriptase synthesizes DNA using an RNA template and can be used to create cDNA, a double-stranded DNA copy of an mRNA molecule.

This method produces double-stranded cDNA copies of all of the mRNAs that were in the source cells, tissue, or organism. Thus, the collection of cDNAs can be used to amplify by PCR any gene that is transcribed in the cells, tissue, or organism. This process is called **reverse transcription**.

PCR (RT-PCR). Furthermore, for mRNA transcribed from a given gene, the number of copies of cDNA is equal to the number of copies of mRNA. Thus, using cDNA as a template, a real-time PCR instrument can quantify mRNA levels in tissues and cells. For example, to determine if reduced transcription of the insulin gene is the cause of type I diabetes in an individual, insulin mRNA can be quantified by real-time PCR analysis of cDNA generated from β cells of the individual.

KEY CONCEPT Conversion of mRNA into cDNA makes it possible to use PCR approaches to amplify and quantify specific mRNAs.

10.2 GENERATING RECOMBINANT DNA

LO 10.2 Describe the functional components of vectors that are useful for cloning DNA.

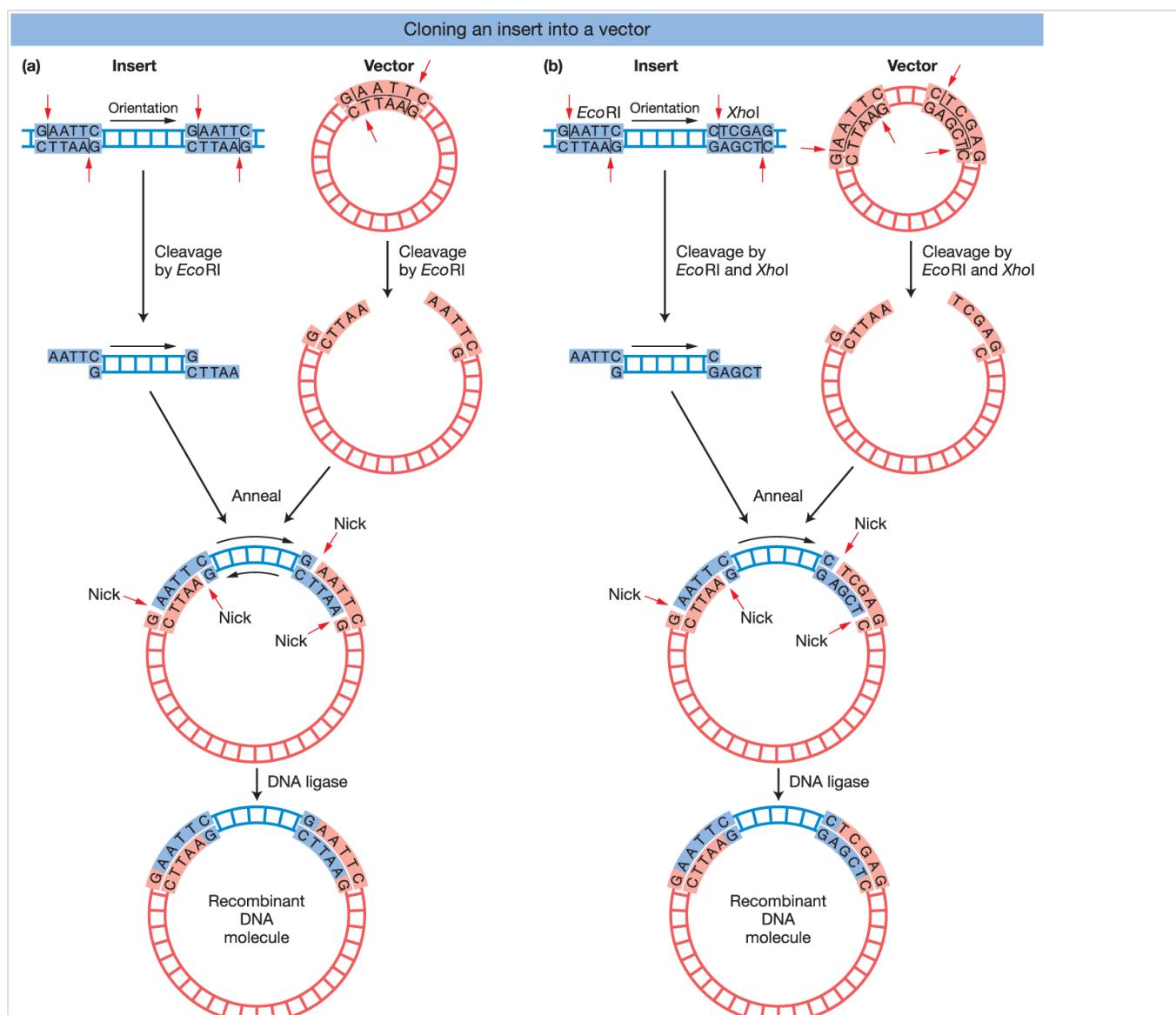
LO 10.3 Describe methods for generating and isolating recombinant DNA molecules.

To investigate the function of genes and their products, it is helpful to be able to manipulate DNA sequences. For example, manipulation of the insulin gene has made it possible to carry out experiments that explain the effect of heritable mutations in the insulin gene on insulin protein expression and function in type I diabetes. A general approach to manipulate DNA is **DNA cloning**, which involves isolating a specific piece of DNA called **donor DNA**, or, more informally, **insert DNA**, and combining it with **vector DNA** to form a **recombinant DNA** molecule. Cloning vectors are naturally occurring DNA molecules that serve as vehicles to carry foreign DNA into a cell. Host cells increase the amount of a recombinant DNA molecule by DNA replication. Thus, the term DNA cloning refers to the process by which many identical copies of a piece of DNA, a clone, are produced.

DNA cloning

In DNA cloning, restriction enzymes and DNA ligase are used to combine insert and vector DNAs into a single molecule. Recall that restriction enzymes cut DNA at specific sequences, producing DNA fragments with staggered or blunt ends (see [Table 10-1](#)). On the other hand, **DNA ligase** joins two DNA fragments together by catalyzing the formation of phosphodiester bonds. [**Figure 10-10**](#) illustrates the basic steps in producing a recombinant DNA molecule. In this example, the restriction enzyme *Eco*RI is used to make a staggered double-strand cut at a single site in a circular vector such as a bacterial plasmid, converting the circular DNA into a single linear molecule with half of an *Eco*RI site at each end. Plasmids can be engineered to contain a **multiple cloning site (MCS)** or **polylinker** that contains restriction enzyme recognition sites that do not occur elsewhere in the plasmid. Thus, cleavage at any one of these sites linearizes the plasmid rather than cutting it into multiple pieces. In the case of the insert, *Eco*RI digestion of a linear piece of DNA at two sites produces a DNA fragment with half of an *Eco*RI site at each end. Mixing the linearized vector with the linear insert allows the “sticky” ends of the vector and insert to hybridize and form a recombinant molecule. DNA ligase finishes the job by creating phosphodiester linkages at the junctions between vector and insert sequences. If a single

restriction enzyme is used for cloning, or if two restriction enzymes are used that both create blunt ends, the insert can hybridize in two orientations relative to the vector ([Figure 10-10a](#)); whereas cloning with two different restriction enzymes (at least one of which produces a staggered cut) limits hybridization of the insert and vector to one orientation ([Figure 10-10b](#)).



Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020 W. H. Freeman and Company

FIGURE 10-10 To form a recombinant DNA molecule, restriction enzymes are used to cut a vector and an insert. Because of sequence complementarity at the ends of the vector and insert, the vector and insert anneal. DNA ligase then permanently links the vector and insert together. Depending on the restriction enzymes used, the insert can hybridize with the vector in (a) two orientations or (b) only one orientation.



ANIMATED ART Sapling Plus

Plasmid cloning

KEY CONCEPT Insert and vector DNAs with the same sticky ends or with blunt ends can be joined efficiently and ligated.

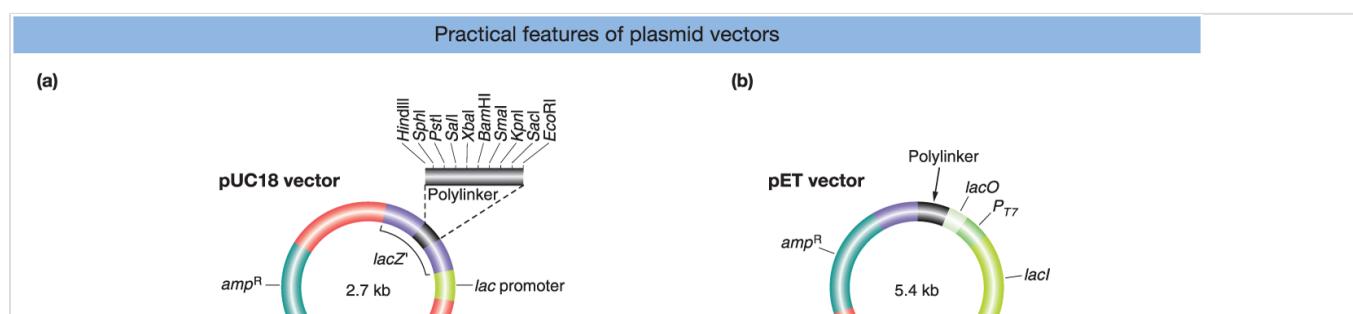
Choice of cloning vector

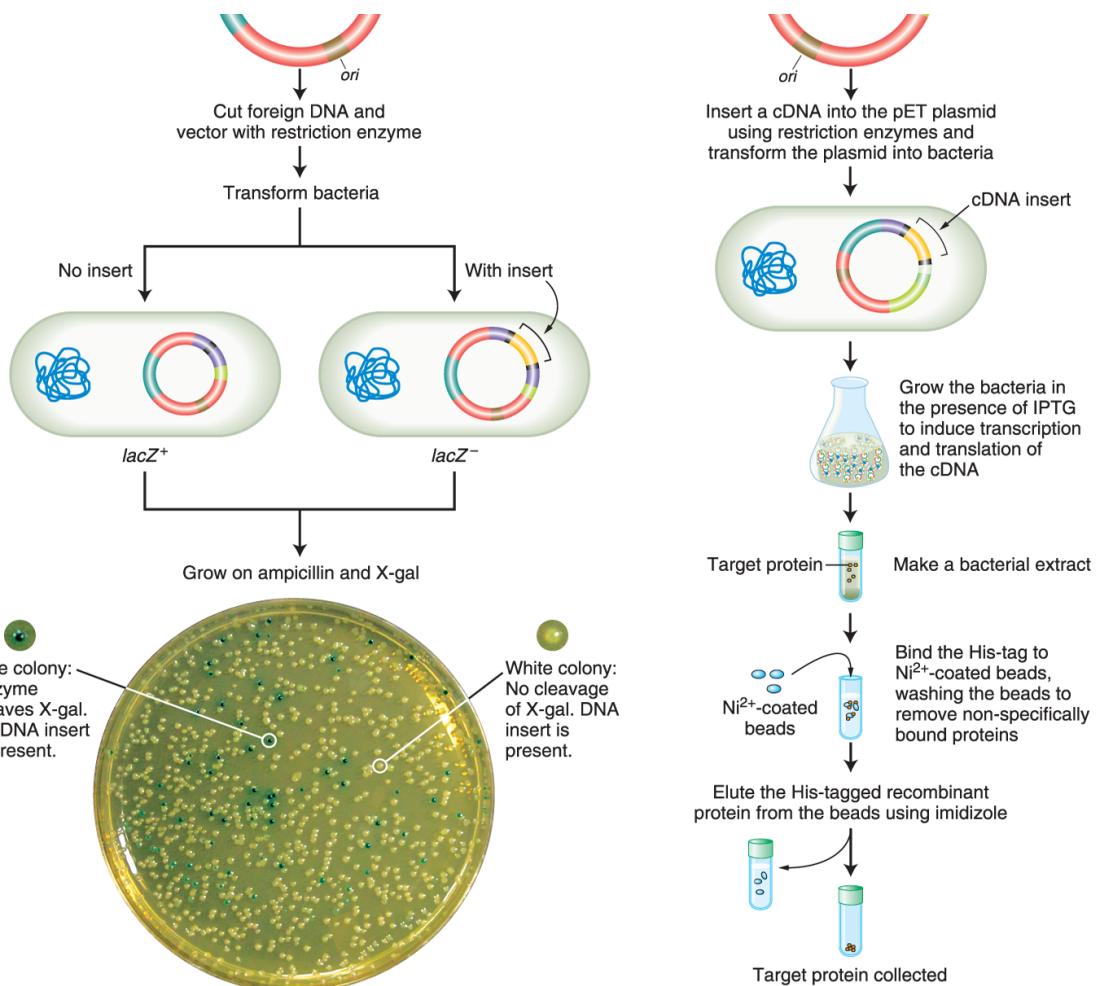
Numerous cloning vectors that meet a wide range of experimental needs are in current use.

Vectors are used mainly to express a specific RNA or protein or to increase the amount of a specific DNA molecule so that the DNA can be sequenced or further cloned. As described earlier, all vectors must have convenient restriction sites at which the DNA to be cloned can be inserted. Other important features of vectors are ways to quickly identify the desired recombinant vector as well as to express and purify the recombinant protein. Some general classes of cloning vectors follow.

Plasmid vectors

Bacterial **plasmids**, which we first encountered in [Chapter 6](#), are small circular DNA molecules that are replicated independently of the bacterial chromosome because they have an origin of replication (ori). Plasmids routinely used as vectors carry a gene for drug resistance and a gene to distinguish plasmids with and without DNA inserts. Genes that confer resistance to antibiotics such as ampicillin (amp^R gene), tetracycline (tet^R gene), and chloramphenicol (cam^R gene) provide a convenient way to select for bacterial cells transformed by plasmids: those cells still alive after exposure to the drug must carry the plasmid vector. However, because not all plasmids in transformed cells will contain DNA inserts, some plasmid vectors also have a system that allows researchers to identify bacterial colonies with plasmids containing DNA inserts. Such a feature is part of the pUC18 plasmid vector shown in [Figure 10-11a](#); DNA inserts disrupt a gene ($lacZ$) in the plasmid that encodes an enzyme (β -galactosidase) capable of cleaving a compound added to the bacterial culture plate (X-gal) so that it produces a blue pigment. Thus, colonies that contain plasmids with an insert will be white rather than blue (i.e., they cannot cleave X-gal because they cannot produce β -galactosidase).





Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020 W. H. Freeman and Company
(a) Dr. James M. Burnette III and Dr. Leslie Bañuelos.

FIGURE 10-11 (a) The pUC18 plasmid has been designed for use as a vector for DNA cloning. The polylinker has multiple restriction sites into which donor DNA can be inserted. Insertion of DNA into pUC18 is detected by inactivation of the β -galactosidase function of *lacZ*, resulting in an inability to convert the artificial substrate X-gal into a blue dye. (b) The pET plasmid has been designed for expression and purification of recombinant proteins. Inducible expression of a recombinant protein is controlled by three pET plasmid elements, the lac operator (*lacO*), the *lacI* gene that encodes the Lac repressor protein, and the T7 polymerase promoter. The plasmid also contains a sequence that encodes a His-tag that is translated in-frame with the recombinant protein. IPTG induces expression of the His-tagged recombinant protein, which is then purified based on the affinity of the His-tag for Ni²⁺-coated beads.

Pal Plasmids called expression plasmids contain sequence that control the transcription and translation of the inserted DNA, often a gene in the form of a cDNA. Expression plasmids can drive transcription of the inserted gene constitutively (i.e., all the time) or inducibly (i.e., only in response to a signal). For example, some pET plasmids are used for constitutive expression of recombinant proteins in *E. coli* (Figure 10-11b). These plasmids contain the bacteriophage T7 promoter that drives transcription of the inserted gene in *E. coli* that express bacteriophage T7 RNA polymerase. Another type of the pET plasmid uses components of the lac operon (described in Chapter 11) to inducibly express the inserted gene in *E. coli*. This type of plasmid contains the

lac operator site near the T7 promoter and also contains the *lacI* gene that encodes the Lac repressor protein. In uninduced cells, the Lac repressor binds the *lac* operator and represses transcription of the inserted gene by T7 polymerase. However, when the compound IPTG (isopropyl β-D-1 thiogalactopyranoside) is added to the growth media, the Lac repressor is inactivated and T7 polymerase can transcribe the inserted gene. Therefore, the recombinant protein is expressed only in the presence of IPTG. Inducible expression is helpful in cases where constitutive expression of the recombinant protein produces a large amount of protein that is toxic to *E. coli*, or it makes the protein insoluble.

A final feature of bacterial expression plasmids, including pET vectors, is a sequence that encodes an epitope tag that can be used to purify recombinant proteins. Epitope tags are short protein sequences that are translated in-frame, often at the N- or C-terminus of a recombinant protein. pET vectors contain an epitope tag called a His-tag that consists of six histidine amino acids (6X-His-tag). Since the tag is small, it typically does not affect the structure or function of the recombinant protein. Purification of His-tagged proteins is based on the affinity of histidine for metal ions such as nickel Ni²⁺. As shown in [Figure 10-11b](#), His-tagged recombinant proteins are purified by Ni²⁺ affinity chromatography from an *E. coli* extract, a solution of *E. coli* cells that are broken open to release the recombinant protein as well as *E. coli* proteins. The *E. coli* extract is mixed with inert beads that have Ni²⁺ immobilized on their surface. The His-tagged recombinant protein binds tightly to the beads. The beads are washed several times to remove non-specifically bound *E. coli* proteins, leaving only His-tagged recombinant proteins bound to the beads. The bound proteins are then released (i.e., eluted) from the beads in a pure form by adding a chemical called imidazole that competes with the His-tag for binding to the beads. Bacterial expression plasmids and methods of this kind are used to synthesize and purify human insulin protein.

KEY CONCEPT The essential features of plasmids for cloning are an origin of replication so that the plasmid is replicated when bacteria divide, a drug resistance gene so that bacteria containing the plasmid can be identified, and a polylinker so that DNA can be inserted with restriction enzymes.

KEY CONCEPT Non-essential but useful features of plasmids include sequence elements for the identification of plasmids that contain inserts, the constitutive or inducible expression of inserted genes, and the addition of an epitope tag onto a recombinant protein.

Bacteriophage vectors

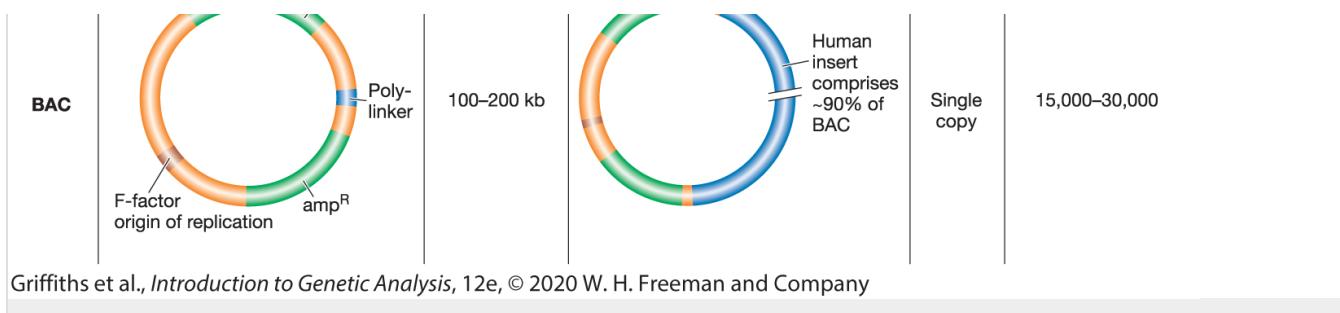
A bacteriophage vector harbors DNA as an insert packaged inside the phage particle. Different classes of bacteriophage vectors can carry different sizes of insert DNA. Bacteriophage λ (lambda; discussed in [Chapters 6](#) and [11](#)) is an effective cloning vector for double-stranded DNA inserts as long as 15 kb. The central part of the phage genome is not required for replication or packaging of λ DNA molecules in *E. coli* and so can be cut out by restriction enzymes and discarded. The deleted central part is then replaced by insert DNA.

Vectors for larger DNA inserts

The standard plasmid and phage λ vectors just described can accept inserts as large as 15 kb. However, many experiments require inserts well in excess of this size. To meet these needs, special vectors that require more sophisticated methods for transferring DNA into the host cell have been engineered. In each case, the recombinant DNAs replicate as large plasmids after they have been delivered into the bacterium.

Fosmids are vectors that can carry 35- to 45-kb inserts ([Figure 10-12](#)). They are engineered hybrids of λ phage DNA and bacterial F plasmid DNA ([Chapter 6](#)). Because of their cos sites from λ phage, fosmids are packaged into λ phage particles that introduce these big pieces of recombinant DNA into recipient *E. coli* cells. Cos (an abbreviation for cohesive) sites are 12 base pair overlapping sticky ends that circularize the linear phage DNA through complementary base pairing. Once introduced into the bacterium, fosmids form circular molecules that replicate extrachromosomally in a manner similar to plasmids. However, because of the presence of an F plasmid origin of replication that couples plasmid replication to host cell chromosome duplication, very few copies accumulate in a cell.

Fosmid and BAC vectors carry large inserts					
	Vector	Size of genomic DNA insert	Vector-insert clone	Copy number	Number of clones for 1× human genome coverage
Fosmid		35–45 kb		Single copy	~75,000



Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020 W. H. Freeman and Company

FIGURE 10-12 Features of some large-insert cloning vectors. The number of clones needed to cover the human genome once ($1 \times$) is based on a genome size of 3000 Mb (3 billion base pairs).

F_o **Bacterial artificial chromosomes (BACs)** are another type of vector for carrying large inserts.

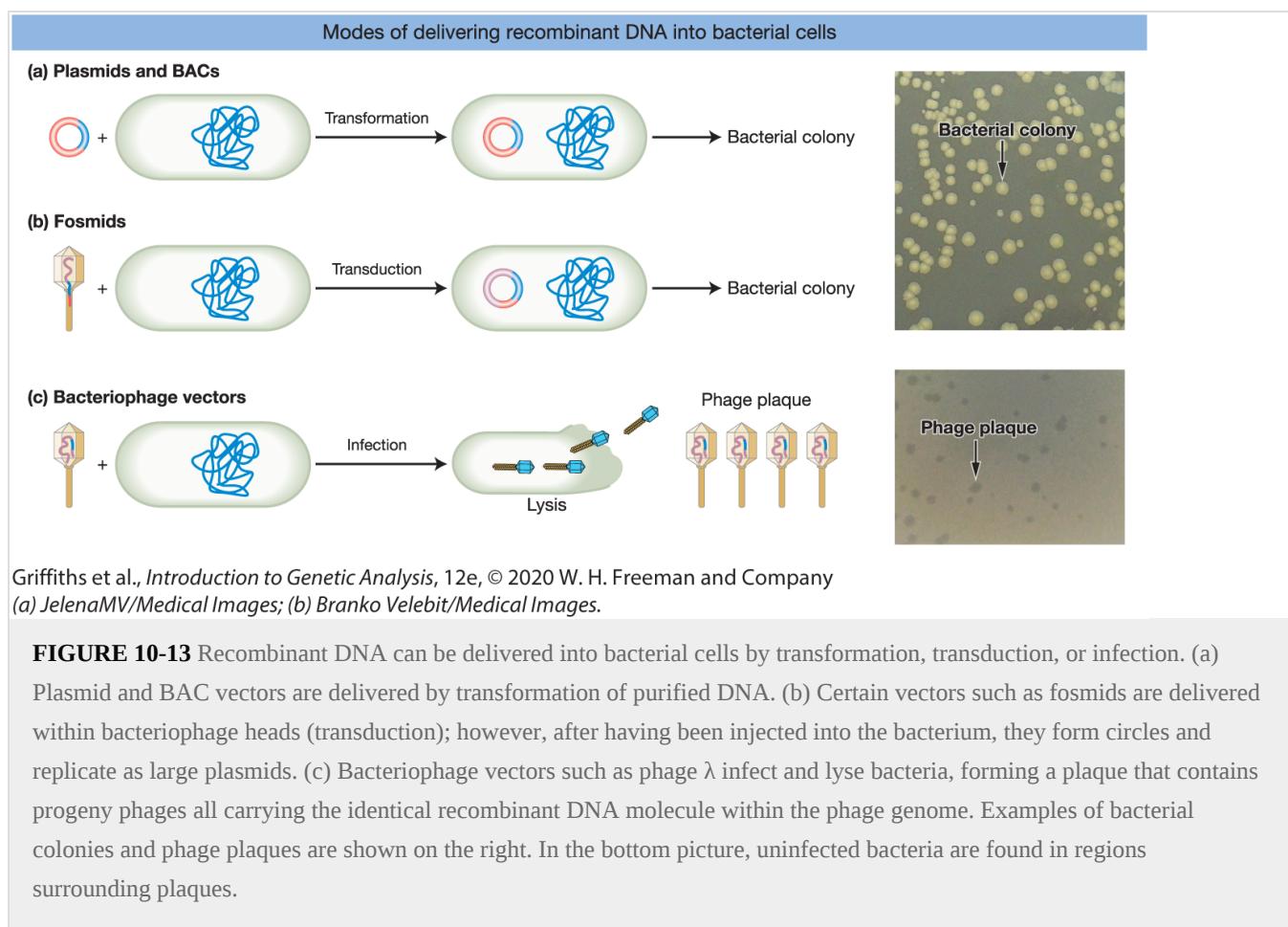
Derived from the F plasmid, BACs can carry inserts ranging from 100 to 200 kb, although the vector itself is only about 7 kb (see [Figure 10-12](#)). The DNA to be cloned is inserted into the plasmid, and this large circular recombinant DNA is introduced into the bacterium. BACs were the workhorse vectors for the extensive cloning required by large-scale genome-sequencing projects, including the public project to sequence the human genome ([Chapter 14](#)).

KEY CONCEPT Cloning vectors accept inserts of small sizes for plasmids, to medium sizes for bacteriophage, to large sizes for fosmids and BACs.

Entry of recombinant DNA molecules into bacterial cells

Three methods are used to introduce recombinant DNA molecules into bacterial cells: transformation, transduction, and infection ([Figure 10-13](#); see also [Sections 6.3, 6.4, and 6.5](#)). In **transformation**, bacteria are incubated in a solution containing the recombinant DNA molecule. Because bacterial cells used in research do not naturally take up plasmids, they must be made *competent* (that is, able to take up the DNA from the surrounding media) by either incubation in a calcium solution (*calcium chloride transformation*) or exposure to a high-voltage electrical pulse (*electroporation*). After entering a competent cell through membrane pores, the recombinant molecule becomes a plasmid chromosome ([Figure 10-13a](#)). Electroporation is the method of choice for introducing especially large DNAs such as BACs into bacterial cells. In **transduction**, the recombinant DNA is combined with phage proteins to produce a virus that contains largely non-viral DNA. These engineered phages inject their DNA into the bacterial cells, but new phages cannot form because they do not carry the viral genes necessary for phage replication. Fosmids are introduced into cells by transduction ([Figure 10-13b](#)). In contrast to transduction, which produces plasmids and bacterial colonies but not new viruses, **infection** of bacteria produces

recombinant phage particles ([Figure 10-13c](#)). Through repeated rounds of infection, a plaque full of λ phage particles forms from each initial bacterium that was infected. Each phage particle in a plaque contains not only the recombinant DNA, but also viral genes needed to create new infective phage particles.



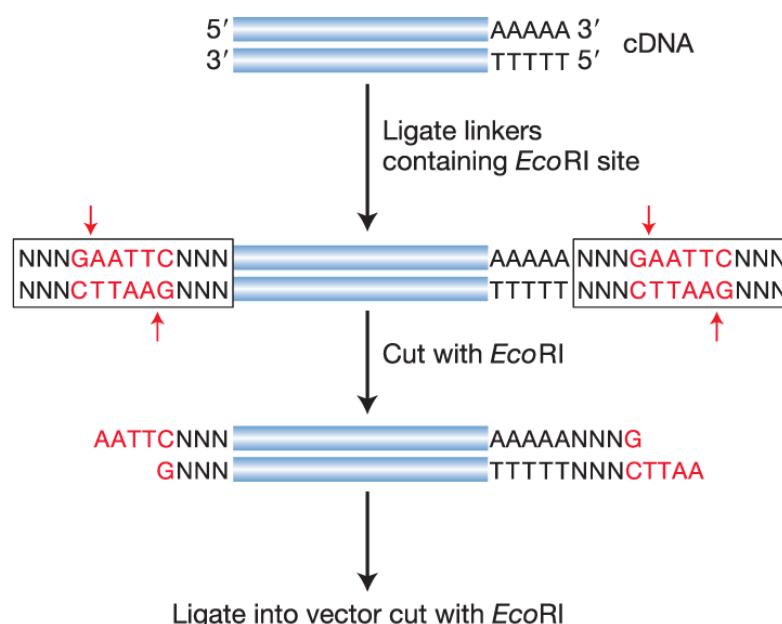
DNA libraries

To clone specific genes and mRNAs such as those for insulin, researchers have put into practice the information presented thus far in this chapter. A common cloning approach is to generate a collection of recombinant DNA molecules called a library, and to fish out the molecule of interest. For example, in 1982 the human insulin gene was identified from a library of human genome fragments. To create a [genomic library](#), restriction enzymes or physical methods are used to break human genomic DNA into fragments of appropriate size for a cloning vector, and each fragment is inserted into a different copy of the vector. If fosmids that accept ~ 40 kb inserts are used as the cloning vector, $\sim 75,000$ independent clones would be required to represent one human genome's worth of DNA (3×10^9 bp in the human genome / 4×10^4 bp per fosmid). To ensure that

all regions of the genome are included, genomic libraries aim to have each DNA fragment represented an average of five times. So, in this example there would need to be 375,000 independent clones in the genomic library ($5 \times 75,000$).

To create a **cDNA library**, mRNA is purified from a cellular source, converted into cDNA, and inserted into a vector. One method for preparing cDNAs for insertion into a cloning vector is to add restriction sites to both ends of each cDNA (Figure 10-14). To do this, DNA ligase is used to link short double-stranded oligonucleotides called **DNA linkers** or **DNA adapters**, which contain a restriction site, to cDNAs. After ligation, the cDNAs are digested with the selected restriction enzyme to generate staggered ends for cloning into a vector that is digested with the same restriction enzyme. cDNA libraries require tens or hundreds of thousands of independent cDNA clones to completely represent the set of expressed genes in a particular cellular source. Suppose we want to identify cDNAs corresponding to insulin mRNAs. Since β cells of the pancreas are the most abundant source of insulin, mRNAs from pancreas are the appropriate source for a cDNA library. To completely represent all the mRNAs expressed by an organism, many cDNA libraries from sources such as different tissues, developmental stages, and environmental conditions are needed.

Producing cDNA molecules with sticky ends



Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020
W. H. Freeman and Company

FIGURE 10-14 Adding EcoRI sites to the ends of cDNA molecules. The cDNA molecules come from the last step in Figure 10-9. Adapters (boxed regions) are added at both ends of the cDNA molecules. These adapters are double-stranded oligonucleotides that contain a restriction site (EcoRI is shown in red) and random DNA sequence at both ends (represented

by N). Note that in the example shown, any cDNAs that contain an internal EcoRI site will be cut into pieces, so some clones will not contain full-length cDNAs.

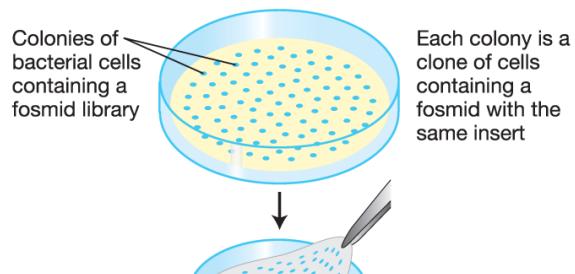
KEY CONCEPT The task of isolating a clone of a specific gene can begin with making a library of genomic DNA or cDNA.

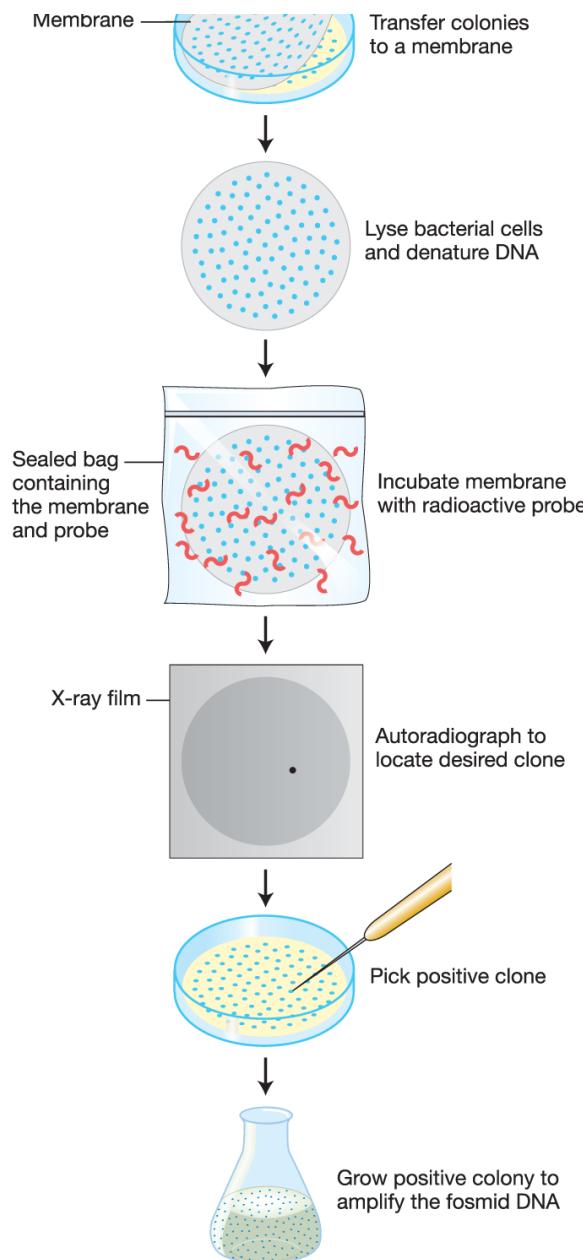
KEY CONCEPT Genomic libraries represent all of the genes in an organism, while cDNA libraries represent only those genes that were expressed in the cells that were the source of mRNA.

Identifying a clone of interest from a genomic or cDNA library

After generating a genomic or cDNA library, the next task in finding a particular clone is to screen the library. Such screening is accomplished by a procedure called colony or plaque hybridization, which is similar to Southern blotting; but in this case, the DNA being analyzed comes from bacterial colonies or phage plaques. The procedure shown in [Figure 10-15](#) is for a library cloned into a fosmid vector, but the steps are similar for libraries of plasmids, BACs, or phages. First, colonies of the library on a petri dish are transferred to a membrane by laying the membrane onto the colonies. The membrane is peeled off, colonies clinging to the surface are lysed in place on the membrane, and the DNA is simultaneously denatured so that it is single-stranded. Second, the membrane is incubated in a solution of a single-stranded probe that is specific for the DNA sequence being sought. Generally, the probe is itself a cloned piece of DNA whose sequence is complementary to that of the desired gene. Since the probe is labeled with either a radioactive isotope or a fluorescent dye, the position of the radioactive or fluorescent label will indicate the position of positive clones. Radioactive probes are detected by autoradiography using X-ray film, and fluorescent probes are detected by photographing the membrane after exposure to a wavelength of light that activates the dye's fluorescence.

Finding a clone of interest by screening a library





Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020
W. H. Freeman and Company

FIGURE 10-15 The clone carrying a gene of interest is identified by probing a genomic library, in this case made by cloning genes in a fosmid vector, with DNA or RNA that has sequence complementary to the desired gene. A radioactive probe hybridizes with any recombinant DNA molecule containing a matching DNA sequence, and the position of the clone having the DNA is revealed by autoradiography. Now the desired clone can be selected from the corresponding spot on the petri dish and grown to high levels in a liquid bacterial culture.



Genomic and cDNA clones are used in different ways

Genomic and cDNA clones of a given eukaryotic gene contain different sequences, which dictates how the clones can be used for gene expression. For illustrative purposes, we will use genomic and cDNA clones of the mouse *Ins1* insulin gene as examples (see [Figure 10-3a](#)). A genomic

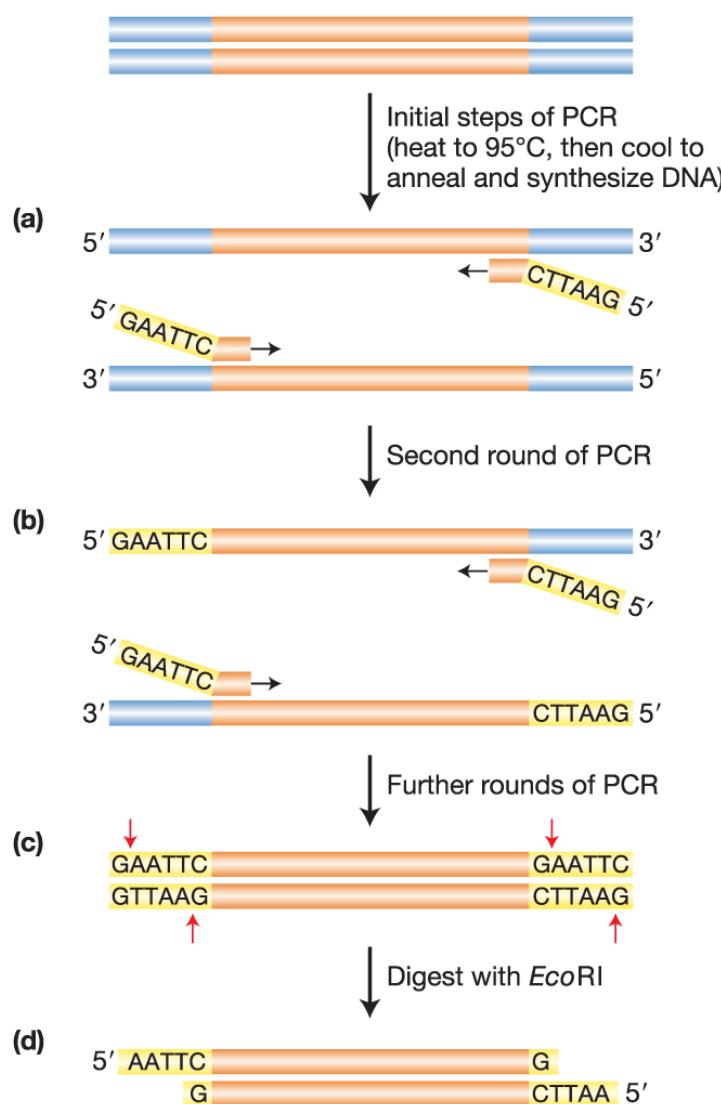
clone of *Ins1* can contain all of the regulatory sequences that are needed to direct the normal expression pattern of *Ins1* in mice. This includes transcriptional regulatory sequences as well as RNA processing sequences for splicing of mRNA introns and polyadenylation of mRNA 3' ends. However, the genomic clone of *Ins1* cannot be expressed in bacteria because bacterial proteins do not recognize eukaryotic transcriptional regulatory sequences and bacteria do not carry out splicing. In contrast, a cDNA clone of *Ins1* cannot be expressed in mice because it lacks regulatory sequences for transcription and polyadenylation, since they are not transcribed into mRNA. Nevertheless, it is possible to express the *Ins1* cDNA in mice using vectors that contain these regulatory sequences from either the *Ins1* gene or another mouse gene. Likewise, the *Ins1* cDNA cannot be expressed in bacteria because it lacks bacterial transcriptional regulatory sequences; but these sequences can be provided by the vector, as described in [Figure 10-11b](#). Furthermore, the lack of splicing in bacteria is not an issue because splicing has already taken place. The same considerations apply to genomic and cDNA clones of the human insulin gene, which is why a cDNA clone is used to produce recombinant human insulin in bacteria (see [Figure 10-1](#)).

KEY CONCEPT Genomic and cDNA clones of a gene are not functionally interchangeable. Both can be used for gene expression, but under different conditions.

Cloning by PCR

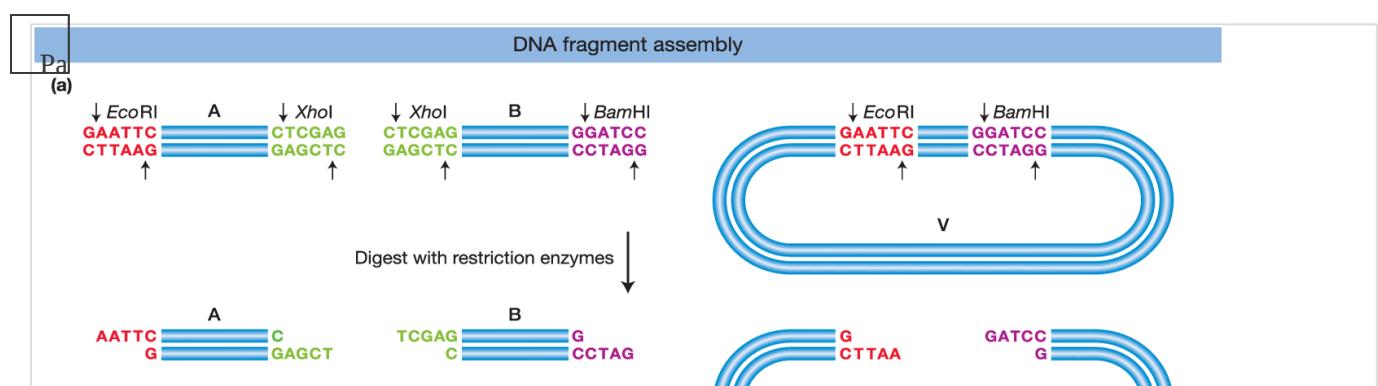
Since the widespread use of PCR in the 1990s, PCR, rather than screening a library, is routinely used to construct a particular genomic or cDNA clone. For example, to clone the human insulin cDNA, oligonucleotide primers are designed that are complementary to the 5' and 3' ends of the insulin cDNA, and PCR is carried out using cDNA generated from pancreas mRNA as the template. To enable cloning of a PCR product, a common approach is to use PCR primers that have restriction sites at their 5' end ([Figure 10-16](#)). Thus, after digestion with a restriction enzyme, the PCR product can be ligated into a vector that is linearized with the same restriction enzyme. A problem with this approach is that the length of PCR products is limited to about 2 kb. To circumvent this problem, cDNAs and genes larger than 2 kb can be cloned by stitching together multiple PCR products with restriction sites at their ends that direct the order in which they are assembled ([Figure 10-17a](#)).

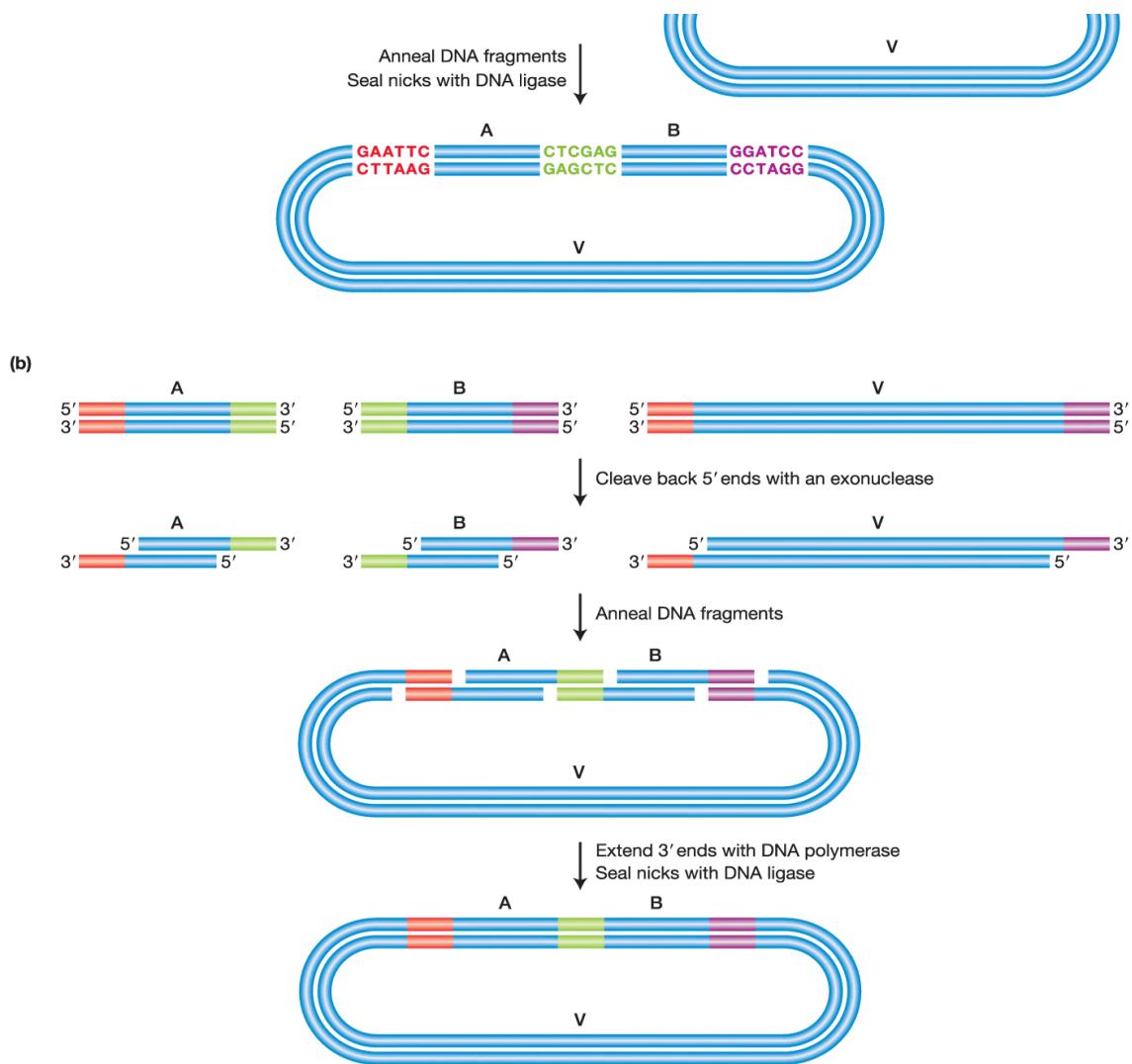
Producing PCR products with sticky ends



Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020
W. H. Freeman and Company

FIGURE 10-16 Adding EcoRI sites to the ends of PCR products. (a) A pair of PCR primers is designed so that their 3' ends anneal to the target sequence, while their 5' ends contain sequences encoding a restriction site (EcoRI in this case). The target DNA is denatured, and 5' ends with the restriction sites remain single stranded while the rest of each primer anneals and is extended by *Taq* DNA polymerase. (b) In the second round of PCR—only the newly synthesized strands are shown—the DNA primers anneal again, and this time DNA synthesis produces double-stranded DNA molecules with restriction sites at one end. (c) The products of all subsequent rounds have EcoRI sites at both ends. (d) Sticky ends are produced when these PCR products are cut with EcoRI.





Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020 W. H. Freeman and Company

FIGURE 10-17 Recombinant DNA molecules with large inserts can be constructed by assembly methods involving multiple DNA fragments (A and B) and a vector (V) with (a) compatible restriction sites at their ends or, as in Gibson assembly, (b) regions of sequence similarity (homology regions) at their ends. Both restriction sites and regions of sequence similarity can be built into the DNA fragments by the appropriate design of PCR primers.

As an alternative to restriction enzyme cloning, researchers have developed other **DNA assembly** methods that can be used to construct large genomic regions and cDNAs and even whole chromosomes and genomes. As an example, *Gibson assembly* can piece together multiple linear DNA fragments that have 15- to 40-bp regions of sequence similarity, often referred to as homology regions, at their ends (Figure 10-17b). The fragments can be produced by PCR or by chemical synthesis, which can generate oligonucleotides of up to 200 nucleotides in length. Assembly is achieved by incubating the fragments with three enzymes: (1) an exonuclease that chews back the 5' ends of each fragment, producing long single-stranded regions with sequence complementarity between fragments, (2) a DNA polymerase to fill in the gaps between annealed fragments, and (3) a DNA ligase to form the final phosphodiester bonds between the fragments. The main advantage of assembly over standard restriction enzyme-based cloning is that assembly

allows the joining of any two DNA fragments at any position, whereas restriction enzyme-based cloning is limited to positions of natural or engineered restriction sites. In addition, assembly is faster than standard restriction enzyme-based cloning because a greater number of DNA fragments can be efficiently joined in a single reaction.

Assembly methods can also be used to put together parts from different genes. For example, researchers commonly combine the transcriptional regulatory region of one gene with the cDNA of another gene. Returning to the insulin gene, this approach could be used to identify the β cells in a mouse pancreas. A researcher could put together the transcriptional regulatory region of a mouse insulin gene, which is only expressed in β cells, and a cDNA for a reporter gene that encodes a protein that is easy to detect. The green fluorescent protein (GFP) gene from a jellyfish is commonly used as a reporter because it encodes a small protein (238 amino acids) that exhibits bright green fluorescence when exposed to ultraviolet light of a particular wavelength. In mice that have this engineered gene inserted into their genome (using techniques described later in this chapter), β cells in the pancreas can be identified because they will be the only cells that express GFP and glow green. Alternatively, a researcher could construct a gene that expresses a single protein, called a *fusion protein*, that is comprised of both insulin and GFP amino acid sequences. This could be done by assembling an insulin gene along with the GFP cDNA such that the insulin and GFP protein-coding regions are translationally in-frame. In this scenario, the insulin protein is said to be “tagged” with GFP. Another way to tag a protein is to append to it a few amino acids that can be recognized by an antibody in vitro by Western blot analysis and in vivo by immunofluorescence microscopy. Like His-tags that are used for protein purification, such tags are called **epitope tags**. Commonly used epitope tags for this purpose include the 7-amino-acid FLAG tag (DYKDDDK), the 9-amino-acid HA tag (YPYDVPDYA), and the 10-amino-acid Myc tag (NNKLISEEDL). The advantage of epitope tags over GFP is that, because of their small size, epitope tags are less likely to alter the structure and function of the protein to which they are fused.

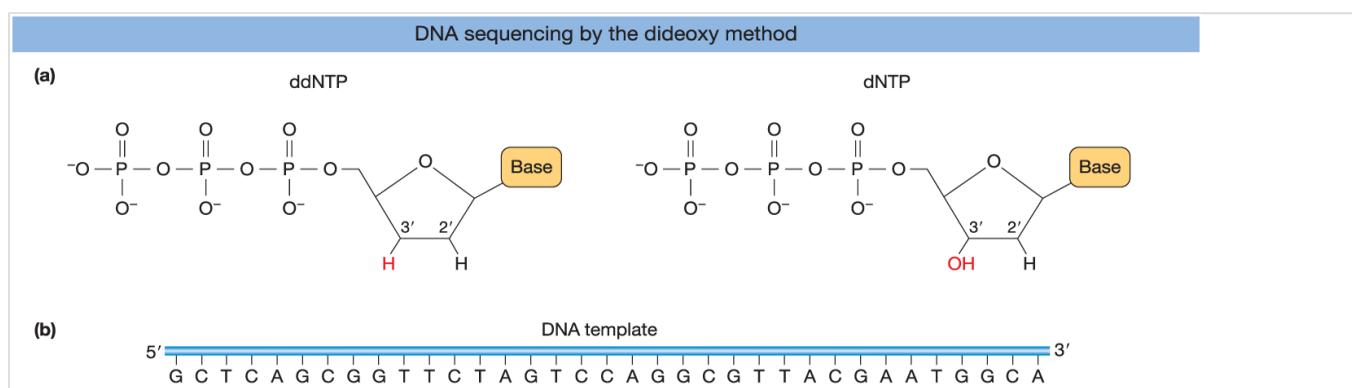
KEY CONCEPT Assembly methods make it relatively easy to modify the sequence of genes to create tools that are useful for research, including reporter genes that reveal the expression pattern of transcriptional regulatory elements and epitope-tagged genes that enable the detection and purification of recombinant fusion proteins.

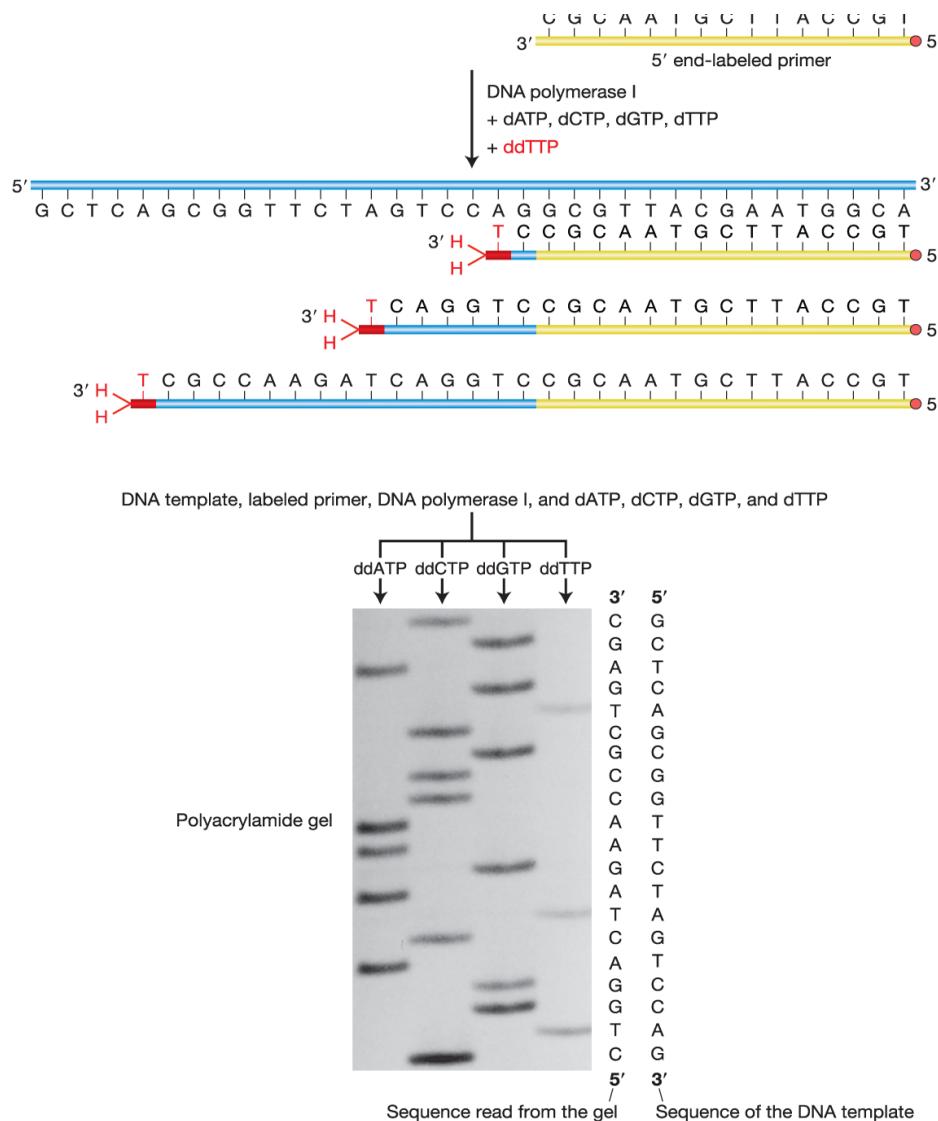
10.3 SEQUENCING DNA

LO 10.4 Diagram the steps of dideoxy DNA sequencing.

As described in many chapters in this text, the regulatory and coding information in DNA is determined by its nucleotide sequence. To reveal this information, public and private institutions have invested heavily in the sequencing of DNA genomes, including the human genome, which was completed in 2001. Furthermore, since mRNA can be converted to cDNA (see [Figure 10-9](#)), the same technologies used to sequence genomes have been extensively applied to sequence cDNAs and to discover the information that flows from DNA to RNA by transcription in cells, tissues, and organisms. Lastly, sequencing of DNA is a common activity in individual laboratories for purposes such as identifying specific DNA lesions in mutant alleles and confirming the sequence of recombinant DNA molecules and PCR products.

Since the late 1970s, researchers have put considerable effort into developing techniques to sequence DNA. Currently, the most commonly used technique for small scale sequencing is called [dideoxy sequencing](#) or [Sanger sequencing](#), after its inventor Fred Sanger. However, as we will see in [Chapter 14](#), other sequencing technologies have largely supplanted this technique when the goal is to determine the sequence of an entire genome. The term *dideoxy* comes from a modified nucleotide, called a dideoxynucleoside triphosphate (ddNTP). This modified nucleotide is key to the Sanger technique because of its ability to be added to a growing DNA chain but to block continued DNA synthesis. A dideoxynucleotide lacks the ribose sugar 3'-hydroxyl group as well as the 2'-hydroxyl group that is absent in a regular deoxynucleotide ([Figure 10-18a](#)). For DNA synthesis to take place, DNA polymerase must catalyze phosphodiester bond formation between the 3'-hydroxyl group of the last nucleotide in the growing chain and the α -phosphate of the nucleotide to be added. Because a dideoxynucleotide lacks the 3'-hydroxyl group, this reaction cannot take place, and DNA synthesis terminates.





Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020 W. H. Freeman and Company

(c) Loida Escote-Carlson, Ph.D.

FIGURE 10-18 (a) 2', 3'-Dideoxynucleotides, which are employed in the Sanger DNA-sequencing method, are missing the 3' ribose hydroxyl group normally present in DNA. DNA is efficiently sequenced by including dideoxynucleotides among the nucleotides used to copy a DNA segment. (b) A labeled primer (designed from the sequence next to the region to be sequenced) is used to initiate DNA synthesis. The addition of four different dideoxynucleotides (ddTTP is shown here) randomly arrests synthesis. (c) Products of the four sequencing reactions are separated by polyacrylamide gel electrophoresis and subjected to autoradiography.

Pa ANIMATED ART Sapling Plus

Dideoxy sequencing of DNA

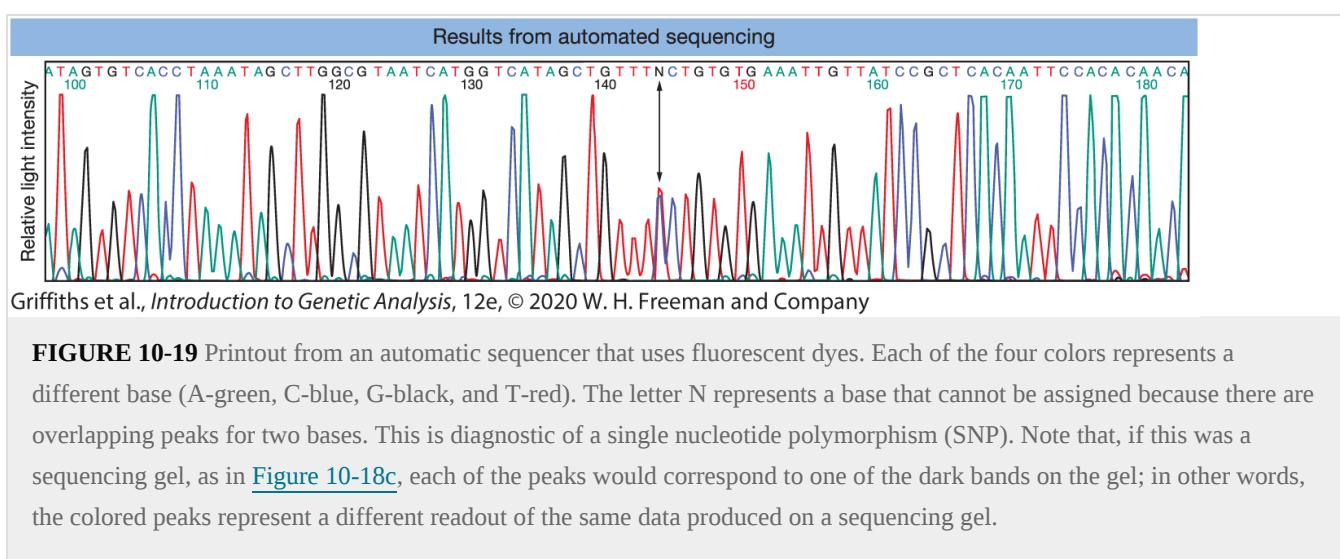
KEY CONCEPT Dideoxynucleotides (ddNTPs) cause chain termination because they lack the 3'-hydroxyl group on the ribose sugar that is essential for phosphodiester bond formation.

DNA sequencing requires four separate reactions, each containing the DNA segment (e.g., a cloned plasmid insert or a PCR product), a radioactive DNA primer that will hybridize to exactly one location on the DNA segment, DNA polymerase, and the four deoxynucleoside triphosphates (dNTPs: dATP, dCTP, dGTP, and dTTP). In addition, each reaction receives a small amount of a different dideoxynucleoside triphosphate (ddNTP: ddATP, ddCTP, ddGTP, or ddTTP). As in DNA replication, the DNA polymerase will add deoxynucleotides to the 3' end of the primer, with the identity of the added deoxynucleotide being determined by base pairing complementarity to the template strand ([Figure 10-18b](#)). Since the dNTPs and each ddNTP are present in a ratio of about 300:1, most of the time, DNA polymerase will add a dNTP and continue synthesis. However, every once in a while, DNA polymerase will incorporate a ddNTP into the new strand, which will terminate synthesis of that DNA strand. Hence, in the tube that contains ddATP, new DNA being synthesized will terminate when a ddATP is added to the strand, thus marking locations of T nucleotides in the DNA segment being sequenced. When complete, the reaction in the ddATP tube results in a collection of radiolabeled single-stranded DNA fragments of different length, each ending with an A residue. This process is repeated for the reactions with ddCTP, ddGTP, and ddTTP. The DNA fragments in the four reactions are separated by polyacrylamide gel electrophoresis and visualized by autoradiography ([Figure 10-18c](#)). Since polyacrylamide gels can resolve fragments of DNA that vary by only one nucleotide in length, the fragments in the gel are separated and ordered by size, with the lengths increasing by one base at a time. Shorter DNA fragments migrate fastest in the gel, so bands at the bottom of the gel represent the sequence closest to the primer. Therefore, the sequence is read in the 5'-to-3' direction from the bottom to the top of the gel and is complementary to the DNA strand being sequenced.

KEY CONCEPT DNA sequencing by the dideoxy (Sanger) method uses dideoxynucleotides to terminate synthesis by DNA polymerase from a DNA template, producing DNA fragments of different lengths that end at each nucleotide position in the template.

In 1986, a modified dideoxy sequencing method was developed that uses an automated electrophoresis system and labels synthesized DNA strands with fluorescent dideoxynucleotides rather than a radioactive primer. The automated method is superior to the prior method because more samples can be sequenced at the same time, and the length of sequence reads is increased from about 200 to 1000 base pairs. Automated sequencing is carried out in a single reaction containing an unlabeled primer and all four dideoxynucleotides, each labeled with a differently colored fluorescent dye. So, synthesized fragments are not fluorescently labeled until they terminate, but, once they are labeled, the color of the fluorescence indicates the nucleotide at the 3' end of the fragment. The synthesized fragments are separated by size by capillary gel

electrophoresis, in which the gel matrix is contained within a thin tube rather than between glass plates. As in polyacrylamide gel electrophoresis, all of the fragments of the same size migrate as a single band. As the bands reach the bottom of the capillary tube, the fluorescence is detected with a laser beam. The intensity of light in each band is depicted as peak in the computer output, as shown in [Figure 10-19](#). This figure shows the sequence of a DNA fragment that was amplified by PCR from the genomic DNA of an individual. At most positions, there is a single peak with green = A, blue = C, black = G, and red = T. The sequence also contains a single nucleotide polymorphism (SNP) at position 144. This nucleotide is read as both a T and a C, which means that the diploid individual has one allele with a T–A base pair and the other with a C–G base pair. This automated technology was used to sequence the human genome as well as the genomes of many other organisms, but more cost-effective and faster technologies are currently in use for large sequencing projects (discussed in [Chapter 14](#)).



ANIMATED ART Sapling Plus

Dideoxy sequencing using fluorescent nucleotides

KEY CONCEPT Automated sequencing is superior to the original sequencing method that uses a radioactive primer because it involves only one reaction, not four, and it produces longer sequencing reads.

10.4 ENGINEERING GENOMES

LO 10.5 Describe methods for generating transgenic organisms.

LO 10.6 Describe the CRISPR-Cas9 technique for precise engineering of genomes.

Thanks to recombinant DNA technologies, genes can be isolated and characterized as specific nucleotide sequences. But even this achievement is not the end of the story. We will see next that knowledge of a sequence is often the beginning of a fresh round of genetic manipulation. When characterized, a sequence can be manipulated to alter an organism's genotype. The introduction of an altered gene into an organism has become central to basic genetic research, but it also finds wide commercial application. Three examples of the latter are (1) goats that are modified to secrete into their milk human antithrombin protein, which is used to treat a rare human blood clotting disorder, (2) rice that is modified to produce beta-carotene, the precursor to vitamin A, the deficiency of which causes health problems for millions of people around the world, and (3) plants that are modified to keep from freezing by incorporation of arctic-fish "antifreeze" genes into their genomes. The use of recombinant DNA techniques to alter an organism's genotype and phenotype is termed *genetic engineering*, and its application for practical purposes is called *biotechnology*.

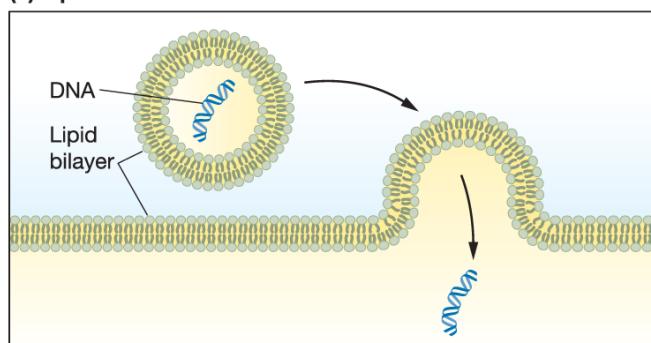
The techniques of genetic engineering described in the first part of this chapter were originally developed in bacteria. These techniques have been extended to model eukaryotes, which constitute a large proportion of organisms used for research. Eukaryotic genes are still typically cloned into bacterial vectors, but eventually they are introduced into a eukaryote, either the original donor species or a completely different one. The gene transferred is called a **transgene**, and the engineered product is called a **transgenic organism**.

Transgenes are introduced into eukaryotic cells by chemical, physical, and biological methods (**Figure 10-20**). Chemical methods are based on the principle that DNA co-precipitated with minerals such as calcium phosphate or packaged inside tiny phospholipid vesicles can be taken up into cells by endocytosis, a natural process by which cells take in molecules from the environment by engulfing them. Physical methods include electroporation, biolistic particle delivery, and microinjection. Electroporation involves applying an electrical field to cells for a short period of time to create microscopic holes in the plasma membrane through which DNA can enter. Biolistic particle delivery systems, also known as gene guns, bombard cells with DNA-coated metal

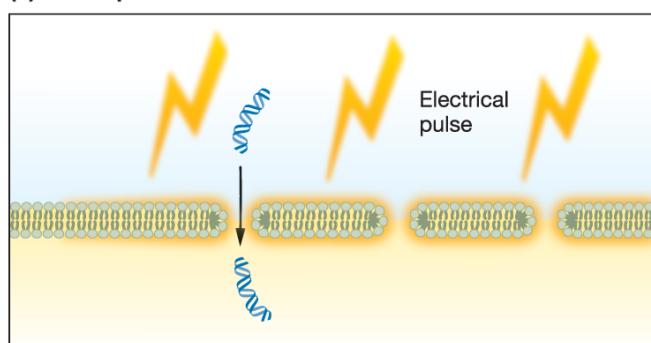
particles that are small enough to enter cells but not destroy them. The last physical method, microinjection, directly delivers DNA into cells through a fine-point needle. Biological methods use bacteria or viruses to transfer DNA into cells. For example, as described shortly, the bacterium *Agrobacterium tumefaciens* can transfer to a plant genome part of its own genome that carries a gene of interest, and viruses can transfer into animal cells their genome that is engineered to include a gene of interest.

Methods of introducing a transgene

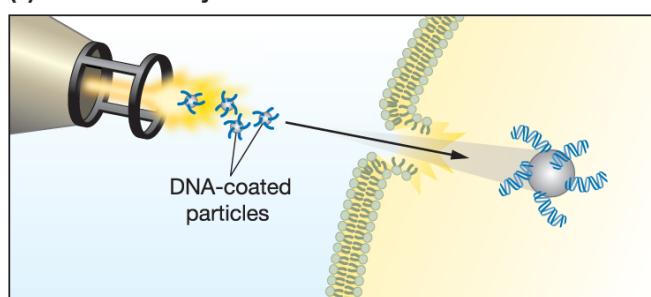
(a) Lipid vesicle



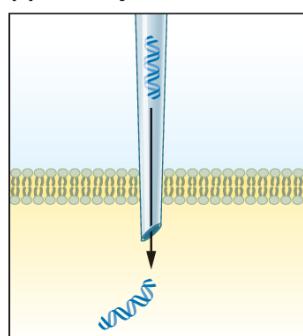
(b) Electroporation



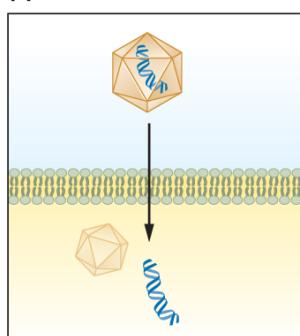
(c) Biostatic delivery



(d) Microinjection



(e) Virus infection



Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020
W. H. Freeman and Company

FIGURE 10-20 Chemical, physical, and biological methods are used to deliver transgenes (i.e., recombinant DNA molecules) into eukaryotic cells. Examples of chemical methods include (a) lipid vesicles; physical methods include (b) electroporation, (c) biolistic delivery, and (d) microinjection; and biological methods include (e) virus infection.

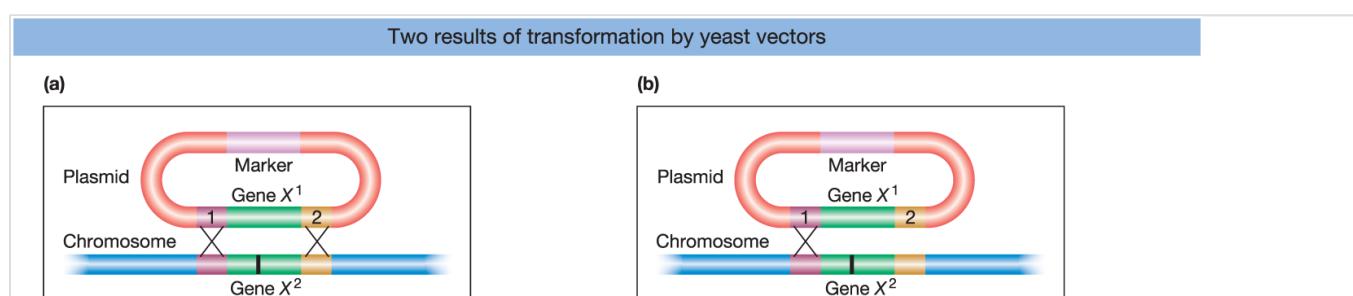
When a transgene enters a cell, it travels to the nucleus, where it becomes a stable part of the genome by either inserting into a chromosome or (in a few species) replicating as part of a plasmid. If insertion occurs, the transgene can either replace the resident gene by homologous recombination or insert **ectopically**—that is, at other locations in the genome. Transgenes from other species typically insert ectopically. We now turn to some examples in fungi, plants, and animals.

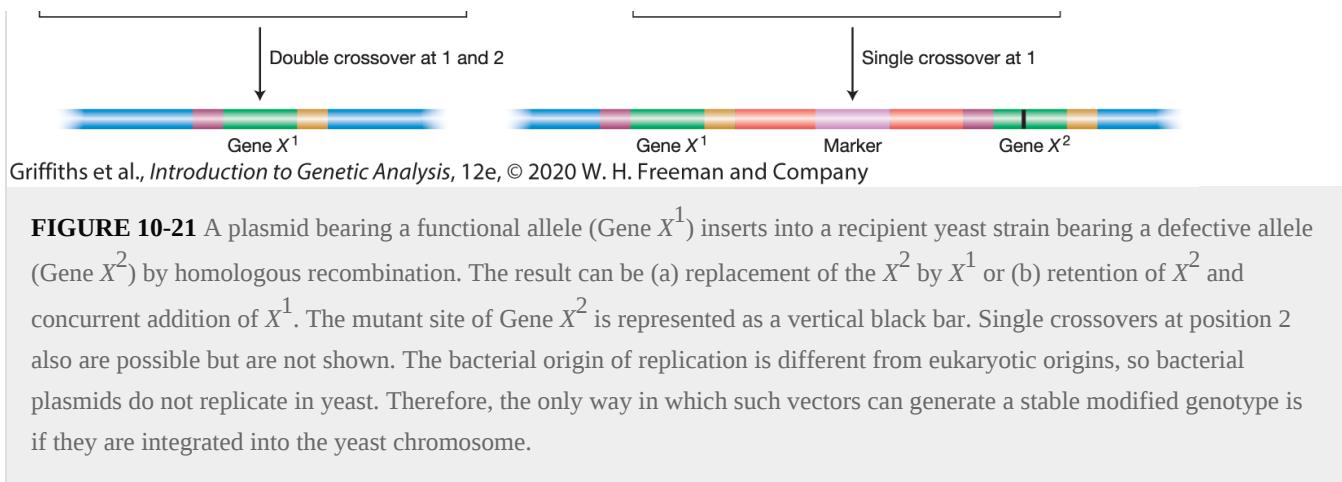
KEY CONCEPT Transgenesis introduces new or modified genetic material into eukaryotic cells.

Genetic engineering in *Saccharomyces cerevisiae*

It is fair to say that *S. cerevisiae* is the most easily manipulated eukaryotic genetic model. Most of the techniques typically used for eukaryotic genetic engineering were developed in yeast; so let's consider the general routes for transgenesis in yeast.

The simplest yeast vectors are yeast integrative plasmids (YIs), derivatives of bacterial plasmids into which yeast DNA of interest has been inserted. When transformed into yeast cells, these plasmids insert into yeast chromosomes, generally by homologous recombination with the resident gene, by either a single or a double crossover ([Figure 10-21](#)). As a result, either the entire plasmid is inserted, or the targeted allele is replaced by the allele on the plasmid. The latter is an example of *gene replacement*—in this case, the substitution of an engineered gene for the gene originally in the yeast cell. Gene replacement can be used to delete a gene or substitute a mutant allele for its wild-type counterpart or, conversely, to substitute a wild-type allele for a mutant.





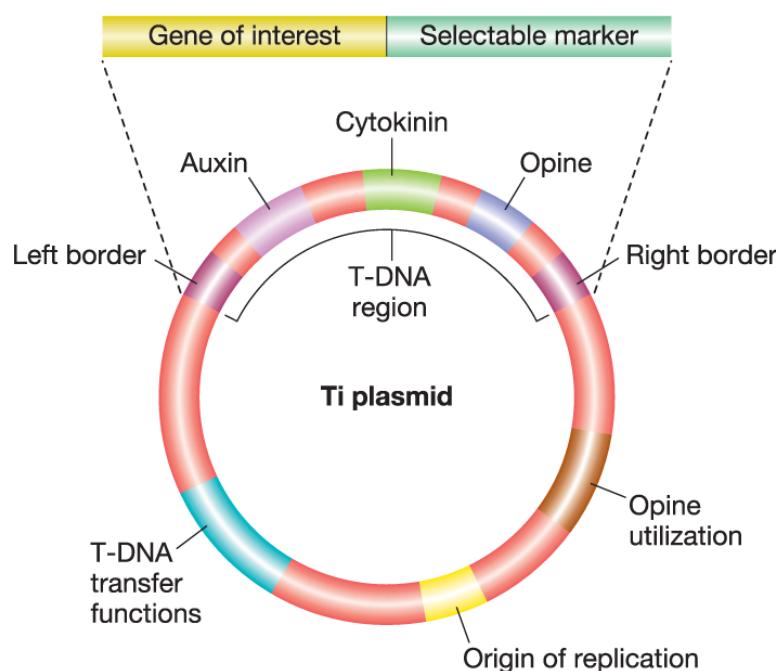
KEY CONCEPT Transgenic yeast cells are generated by homologous recombination between a yeast chromosome and a plasmid that is transformed into yeast and carries a gene of interest.

Genetic engineering in plants

Recombinant DNA technologies have introduced a new dimension to the effort to develop improved crop varieties. No longer is genetic diversity achieved solely by selecting variants within a given species. DNA can now be introduced from other species of plants, animals, or even bacteria, producing **genetically modified organisms (GMOs)**. Genome modifications made possible by this technology are almost limitless. In response to new possibilities, a sector of the public has expressed concern that introduction of GMOs into the food supply may produce unexpected health problems. The concern about GMOs is one facet of an ongoing public debate about complex public health, safety, ethical, and educational issues raised by new genetic technologies.

A vector routinely used to produce transgenic plants is derived from the **Ti plasmid**, a natural plasmid from a soil bacterium called *Agrobacterium tumefaciens*. This bacterium causes what is known as *crown gall disease*, in which the infected plant produces uncontrolled growths called tumors or galls. The key to tumor production is a large (200 kb) circular DNA plasmid—the *Ti (tumor-inducing) plasmid* (**Figure 10-22**). When the bacterium infects a plant cell, a part of the Ti plasmid is transferred and inserted, apparently more or less at random, into the genome of the host plant. The region of the Ti plasmid that inserts into the host plant is called *T-DNA*, for transfer DNA. The genes whose products catalyze this T-DNA transfer reside in a region of the Ti plasmid separate from the T-DNA region itself.

The Ti plasmid vector

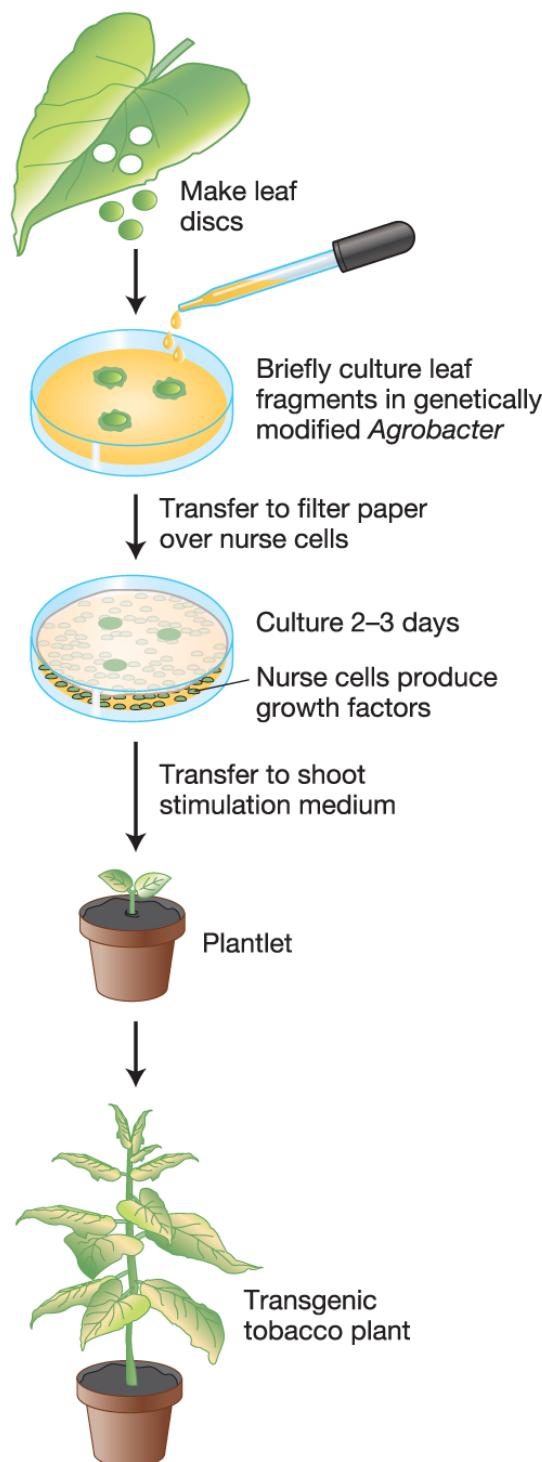


Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020
W. H. Freeman and Company

FIGURE 10-22 Simplified representation of the major regions of the Ti plasmid of *A. tumefaciens* containing an engineered T-DNA.

The natural behavior of the Ti plasmid makes it well suited to the role of a vector for plant genetic engineering. In particular, any DNA that is inserted between the left and right T-DNA border sequences (24 base-pair ends) can be mobilized by other functions provided by the Ti plasmid and inserted into plant chromosomes. Thus, scientists are able to eliminate all of the T-DNA sequence between the borders (including the tumor-causing genes) and replace it with a gene of interest and a selectable marker (for example, kanamycin resistance). One method of introducing the T-DNA into the plant genome is shown in [Figure 10-23](#). Bacteria containing an engineered Ti plasmid are used to infect cut segments of plant tissue, such as punched-out leaf discs. If the leaf disks are placed on a medium containing kanamycin, only the plant cells that have acquired the kan^R gene engineered into the T-DNA will undergo cell division. Transformed cells grow into a clump, or callus, that can be induced to form shoots and roots. These calli are transferred to soil, where they develop into transgenic plants. Typically, only a single copy of the T-DNA region inserts into a given plant genome, where it segregates at meiosis like a regular Mendelian allele. Therefore, one-quarter of the progeny from crossing of the original transgenic plants will get two copies of the T-DNA. The presence of the insert can be verified by Southern blot analysis of purified DNA with a T-DNA probe or by PCR using primers specific for the T-DNA.

Generation of a transgenic plant



Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020
W. H. Freeman and Company

FIGURE 10-23 Insertion of T-DNA into plant chromosomes. Incubation of tobacco leaf discs with the bacterium *A. tumefaciens* containing an engineered T-DNA leads to leaf cells with the T-DNA in their genome, which are able to grow on plates with growth factors and can be coaxed to differentiate into transgenic tobacco plants.



Transgenic plants carrying any one of a variety of foreign genes are in current use, including crop plants carrying genes that confer resistance to certain bacterial or fungal pests, and many more are in development. Not only are the qualities of plants themselves being manipulated, but, like microorganisms, plants are also being used as convenient “factories” to produce proteins encoded by foreign genes.

KEY CONCEPT Transgenic plants are generated by random insertion into a chromosome of a Ti plasmid that carries a gene of interest and is delivered by *Agrobacterium tumefaciens*.

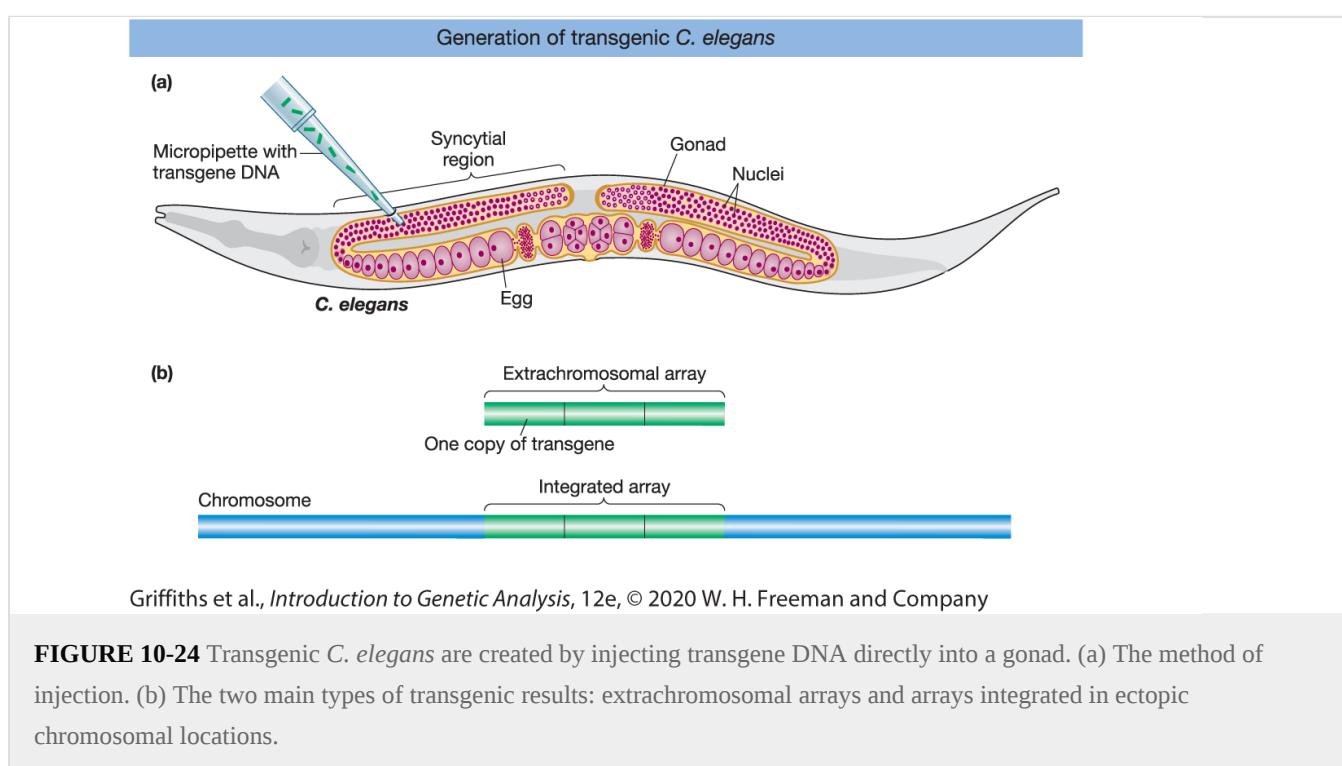
Genetic engineering in animals

Transgenic technologies are now being employed with many animal model systems. We will focus on two animal models heavily used for basic genetic research: the nematode *Caenorhabditis elegans* and the mouse *Mus musculus*. A commonly used method to transform a third model organism, the fruit fly *Drosophila melanogaster*, is described in [Chapter 16](#). Versions of many of the techniques considered so far can also be applied in these animal systems.

Transgenesis in *C. elegans*

Microinjection is used to introduce transgenes into *C. elegans*. Transgenic DNAs are injected directly into the organism, typically as plasmids, fosmids, or other DNAs cloned in bacteria. The injection strategy is determined by the worm’s reproductive biology of the hermaphrodite gonad. The gonads are syncytial, meaning that there are many nuclei within the same gonadal cell. One syncytial cell is a large proportion of one arm of the gonad, and the other syncytial cell is the bulk of the other arm ([Figure 10-24a](#)). These nuclei do not form individual cells until meiosis, when they begin their transformation into individual eggs or sperm. A solution of DNA is injected into the syncytial region of one of the arms, thereby exposing more than 100 nuclei to the transgenic DNA. By chance, a few of these nuclei will incorporate the DNA (remember, the nuclear membrane breaks down in the course of division, and so the cytoplasm into which the DNA is injected becomes continuous with the nucleoplasm). Typically, the transgenic DNA forms multicopy *extrachromosomal arrays* ([Figure 10-24b](#)) that exist as independent units outside the

chromosomes. The arrays are stably inherited, but not with the same efficiency as chromosomes. More rarely, the transgenes will become integrated into an ectopic position in a chromosome, still as a multicopy array.



KEY CONCEPT Transgenic worms are generated by injection of a plasmid containing a gene of interest into the gonad. The plasmid is typically stably inherited as a multicopy extrachromosomal array.

Transgenesis in *M. musculus*

Mice are a very important model for mammalian genetics because they are relatively easy to breed and genetically manipulate. Furthermore, many of the technologies developed in mice and biological insights gained from studies of mice are potentially applicable to humans. There are two strategies for transgenesis in mice, each having its advantages and disadvantages:

- **Ectopic insertions.** Transgenes are inserted randomly in the genome, usually as multicopy arrays. Mice generated with an ectopic insertion are called *transgenic* mice.
- **Gene targeting.** Transgenes are inserted into a location occupied by a homologous sequence in the genome. That is, the transgene replaces its normal homologous counterpart. Mice generated by gene targeting are called *knock-in* or *knockout* mice. For knock-in mice, new DNA either is added to the targeted gene or is substituted for DNA sequences at the targeted

gene. For knockout mice, part or all of the targeted gene is deleted, or a DNA sequence is inserted into the targeted gene to disrupt its expression, thereby creating a loss-of-function mutation.

Ectopic insertions

To insert transgenes in random locations, a solution of bacterially cloned DNA is injected into either the male or female pronucleus of a fertilized egg ([Figure 10-25](#)). Several injected eggs are inserted into the oviduct of a recipient mouse. Progeny are analyzed for integration of the transgene. Typically, DNA extracted from a piece of the tail is used for Southern blot analysis or PCR analysis for the transgene. Occasionally, mice are mosaic; that is, not every cell contains the transgene because DNA integration occurred at a two-cell or later stage of embryogenesis. Positive mice are subsequently mated, their offspring are analyzed for transgene expression, and positive mice are used to establish transgenic mouse lines with stable integration and expression of the transgene. The technique gives rise to some problems: (1) the expression pattern of the randomly inserted genes may be abnormal due to [position effects](#) from the local chromatin environment (see [Chapter 12](#) for more on position effects), and (2) DNA rearrangements can occur inside the multicopy arrays (in essence, mutating the sequences). Nonetheless, this technique is much more efficient and less laborious than gene targeting. Because of the ease of generating mice with ectopic insertions, it has been used to produce human antibodies for use as therapeutics.

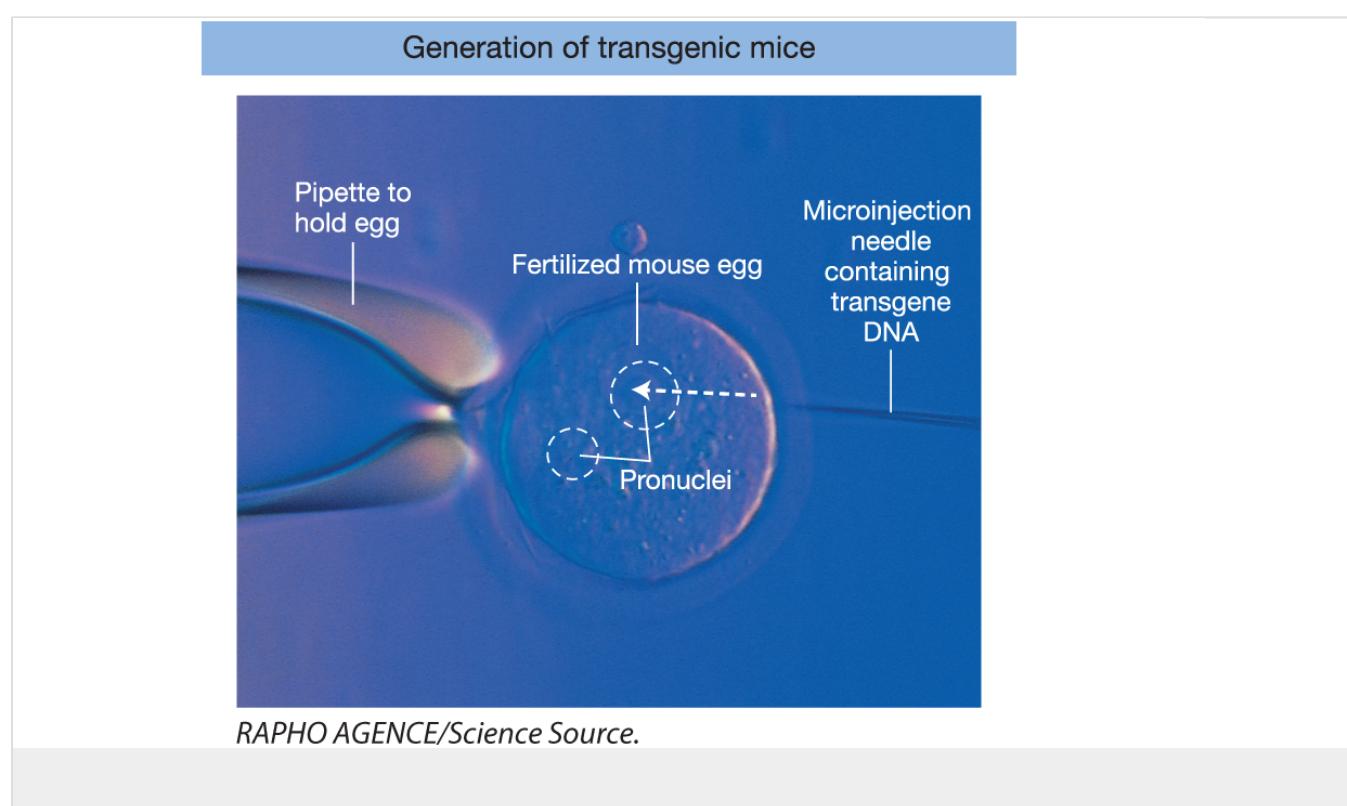


FIGURE 10-25 Transgenic mice are created by injection of cloned DNA into fertilized eggs and subsequent insertion in ectopic chromosomal locations.

ANIMATED ART  Sapling Plus

Creating a transgenic mouse

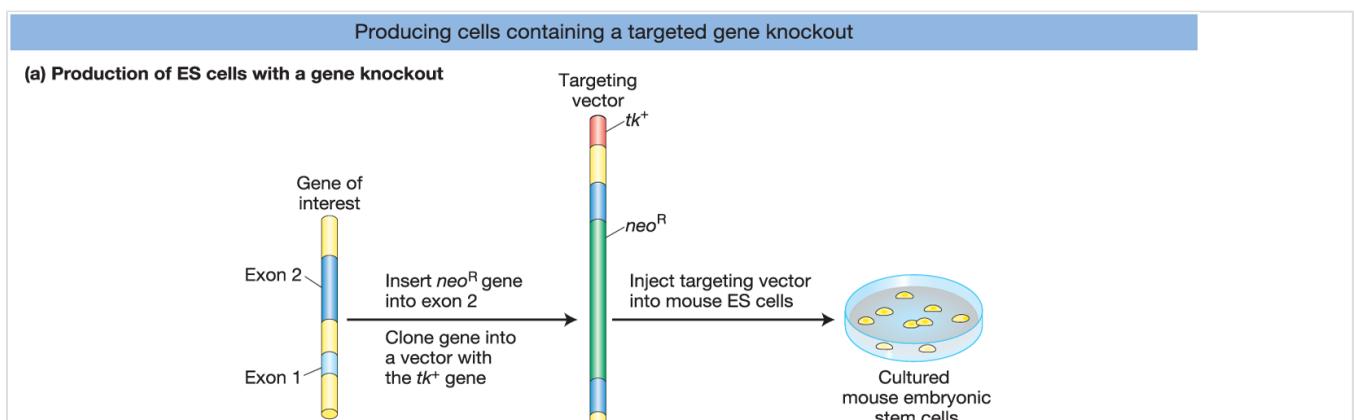
Gene targeting

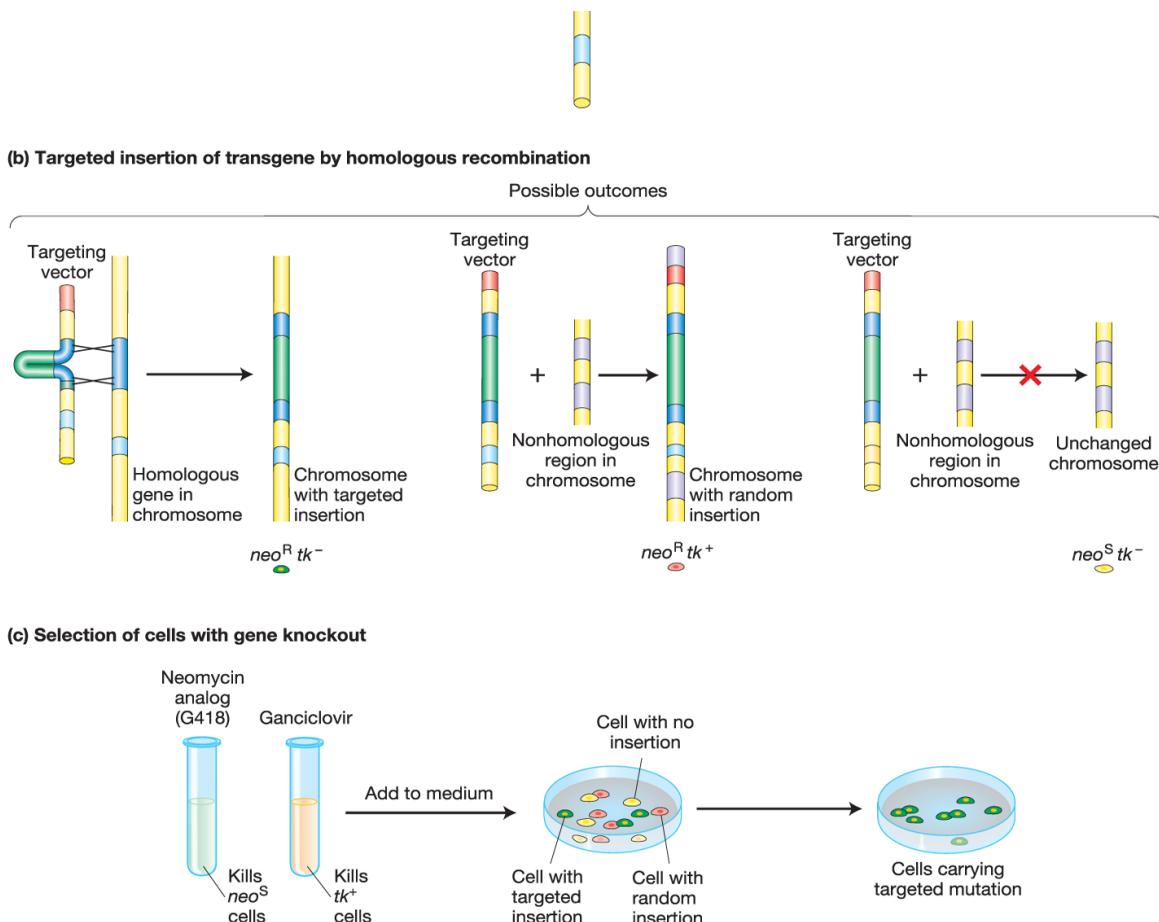
Gene targeting enables researchers to eliminate a gene or modify its function. In one application, called **gene replacement**, a mutant allele can be repaired by substituting a wild-type allele in its normal chromosomal location. Gene replacement avoids both position effects and DNA rearrangements associated with ectopic insertion, because a single copy of the gene is inserted in its normal chromosomal environment. Conversely, a gene may be inactivated by substituting an inactive gene for the normal gene. Such a targeted inactivation is called a **gene knockout**.

Gene targeting in mice is carried out in cultured embryonic stem cells (ES cells). In general, a stem cell is an undifferentiated cell in a given tissue or organ that divides asymmetrically to produce a progeny stem cell and a cell that will differentiate into a terminal cell type. ES cells are special stem cells called pluripotent stem cells that can differentiate to form any cell type in the body—including, most important, the germ line.

To illustrate the process of gene targeting, we look at how it achieves one of its typical outcomes—namely, the substitution of an inactive gene for the normal gene, or gene knockout. The process requires three stages:

1. An inactive gene is targeted to replace the functioning gene in a culture of ES cells, producing ES cells containing a gene knockout ([Figure 10-26](#)).

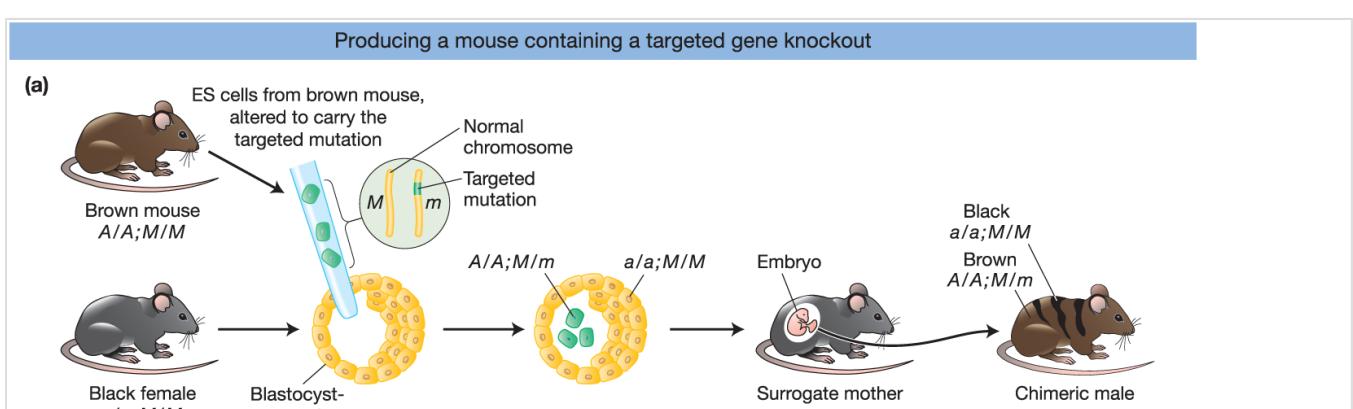




Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020 W. H. Freeman and Company

FIGURE 10-26 Producing cells that contain a mutation in one specific gene, known as a targeted mutation or a gene knockout. (a) The gene of interest (cloned gene) is inactivated by insertion of the neo^R gene in exon 2 and cloned into a targeting vector containing the tk gene. Copies of a cloned gene are altered in vitro to produce the targeting vector. The vector is then injected into ES cells. (b) When homologous recombination occurs (*left*), the homologous regions on the vector, together with any DNA in between but excluding the marker at the tip, take the place of the original gene. This event is important because the vector sequences serve as a useful tag for detecting the presence of this mutant gene. In many cells, though, the full vector (complete with the extra marker at the tip) inserts ectopically (*middle*) or does not become integrated at all (*right*). (c) To isolate cells carrying a targeted mutation, all of the cells are cultured in media containing drugs to select for cells containing the targeted insertion.

- 2. ES cells containing the inactive gene are transferred to mice embryos ([Figure 10-27a](#)).
- 3. Knockout mice are identified and bred to produce mice of known genotype ([Figure 10-27b](#)).



a/a, m/m

stage embryo

(b)

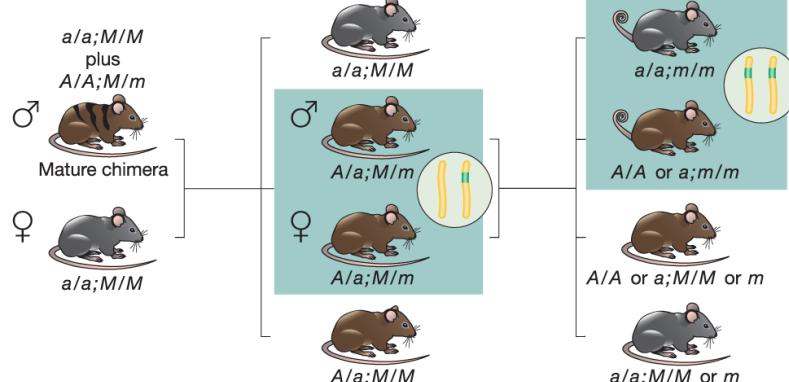
Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020 W. H. Freeman and Company

FIGURE 10-27 A knockout mouse is produced by inserting embryonic stem (ES) cells carrying the targeted mutation into an embryo. (a) ES cells are isolated from an agouti (brown) mouse strain (*A/A*) and altered to carry a targeted mutation (*m*) in one chromosome. The ES cells are then inserted into young embryos, one of which is shown. The coat color of the future newborns is a guide to whether the ES cells have survived in the embryo. Hence, ES cells are typically put into embryos that, in the absence of the ES cells, would acquire a totally black coat. Such embryos are obtained from a black strain that lacks the dominant agouti allele (*a/a*). Embryos containing the ES cells grow to term in surrogate mothers. Agouti shading intermixed with black indicates those newborns in which the ES cells have survived and proliferated. Such mice are called *chimeras* because they contain cells derived from two different strains of mice. Solid black coloring, in contrast, indicates that the ES cells have perished, and these mice are excluded. *A* represents agouti; *a*, black; *m* is the targeted mutation; and *M* is its wild-type allele. (b) Chimeric males are mated with black (nonagouti) females. Progeny are screened for evidence of the targeted mutation (green in inset) in the gene of interest. Direct examination of the genes in the agouti mice reveals which of those animals (*boxed*) inherited the targeted mutation. Males and females carrying the mutation are mated with one another to produce mice whose cells carry the chosen mutation in both copies of the target gene (*inset*) and thus lack a functional gene. Such animals (*boxed*) are identified definitively by direct analysis of their DNA. The knockout in this case results in a curly-tail phenotype.

Stage 1: A recombinant DNA molecule is generated that disrupts a gene of interest. In the example shown in [Figure 10-26a](#), the gene was inactivated by insertion of the neomycin-resistance gene (neo^R) into a protein-coding region (exon 2) of the gene. The mutant gene was then cloned into a vector containing the herpes virus *thymidine kinase* (*tk*) gene. In later steps, the neo^R gene will serve as a marker to indicate that the transgene inserted in a chromosome, and loss of the *tk* gene will ensure that the transgene is inserted at the homologous locus rather than randomly in a chromosome ([Figure 10-26b](#)). These markers are standard, but others could be used instead. The cloned DNA is microinjected into the nucleus of cultured ES cells. The defective gene inserts far more frequently into nonhomologous (ectopic) sites than into homologous sites, so the next step is to select the rare cells in which the defective gene has replaced the functioning gene as desired ([Figure 10-26c](#)). To isolate cells carrying a targeted mutation, the cells are cultured in medium containing drugs—here, a neomycin analog (G418) and ganciclovir. G418 is

lethal to cells unless they carry a functional *neo^R* gene, and so it eliminates cells in which no integration of vector DNA has taken place (yellow cells). Meanwhile, ganciclovir kills any cells that harbor the *tk* gene, thereby eliminating cells bearing a randomly integrated vector (red cells). Consequently, the only cells that survive and proliferate are those harboring the targeted insertion (green cells).

Stage 2: ES cells that contain one copy of the disrupted gene of interest, that is, a gene knockout, are injected into a blastocyst-stage embryo, which is then implanted in a surrogate mother ([Figure 10-27a](#)). Some of the ES cells may become incorporated into the host embryo, and if that happens, the mouse that develops will be **chimeric**—that is, it will contain cells from two different mouse strains.

Stage 3: When the chimeric mouse reaches adulthood, it is mated with a normal mouse. If the chimeric mouse contained germ-line cells that were derived from the ES cells (with the knockout gene), some of the resulting offspring will inherit the gene knockout in all their cells. Sibling mice that are identified by Southern blot or PCR analysis as being heterozygous for the knockout version of the gene of interest are then mated to produce mice that are homozygous for the knockout allele. If the gene is essential, homozygotes will be lethal, and none will be obtained from this cross ([Figure 10-27b](#)).

KEY CONCEPT Knock-in and knockout mice are generated by homologous recombination in ES cells between a mouse chromosome and a plasmid. Gene-targeted ES are then injected into embryos to generate chimeric mice that are crossed to assess heritable germ-line transmission.

KEY CONCEPT Germ-line transgenic techniques have been developed for all well-studied eukaryotic species. These techniques depend on an understanding of the reproductive biology of the recipient species.

CRISPR-Cas9 genome engineering

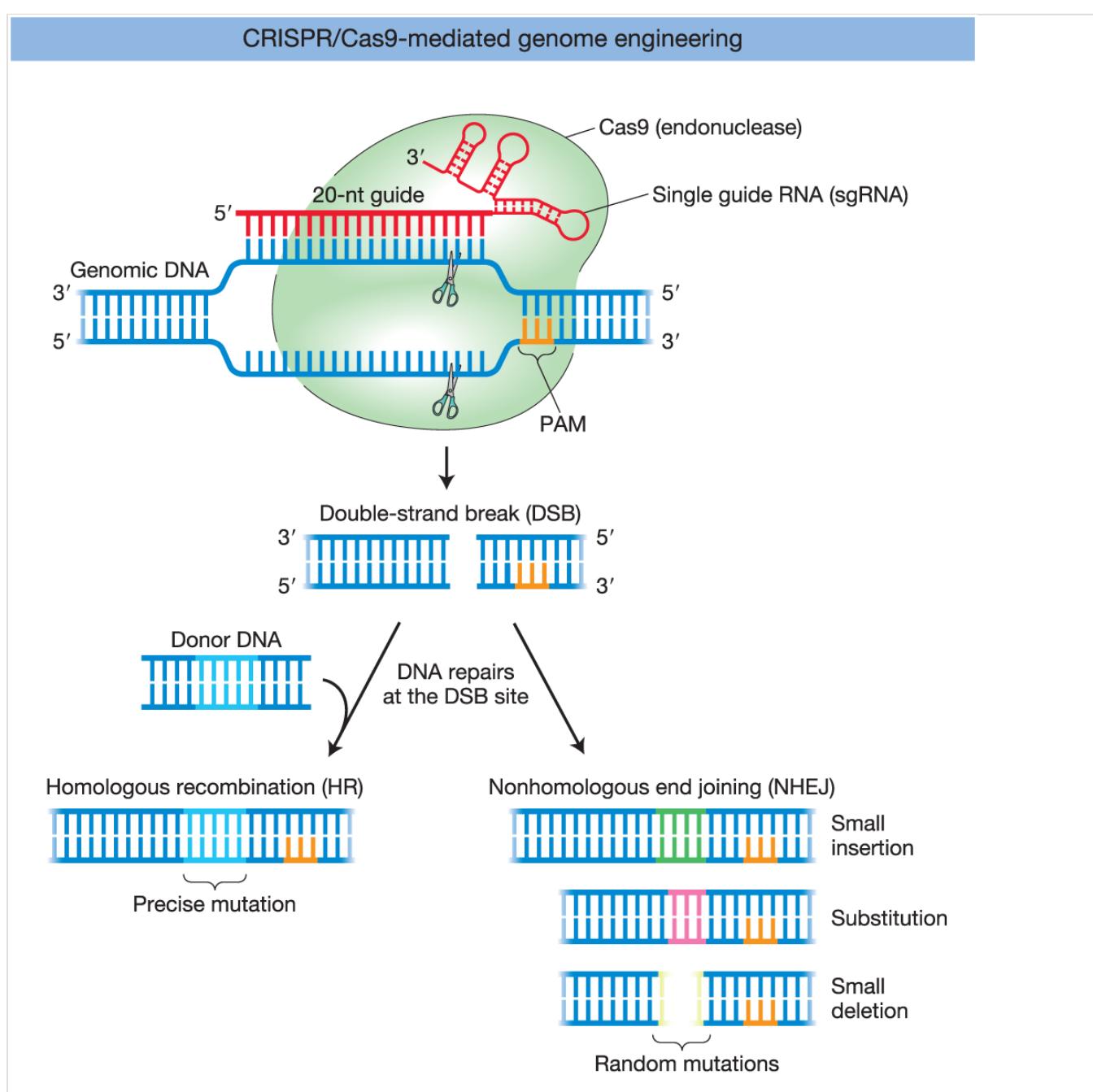
An alternative approach to transgenesis for engineering genomes takes advantage of the natural ability of cells to repair DNA double-strand breaks (DSBs), a topic covered in detail in [Chapter 15](#). In brief, DSBs in eukaryotes are usually repaired by **nonhomologous end joining (NHEJ)**, a mechanism that reattaches the two chromosomal pieces but in a sloppy fashion, causing nucleotides to be inserted or deleted at the site of the DSB. Alternatively, DSBs are repaired by

homologous recombination (HR), which fixes the break without errors using a homologous donor DNA (e.g., a sister chromatid or a plasmid). Therefore, if DSBs could somehow be directed to occur at a particular place in the genome, repair by NHEJ would create mutations that are likely to result in inactivation of the targeted gene, while repair by HR using a homologous donor DNA with an altered sequence would create mutations in the gene. To date, three technologies have been developed that create site-specific DSBs: zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and CRISPR RNA-guided Cas nucleases (CRISPR-Cas). ZFNs and TALENs are proteins that contain two functional domains: a domain with DNA-binding activity that is designed to bind a specific DNA sequence, and a domain with non-specific DNA endonuclease activity that produces DSBs. When expressed in cells, ZFNs and TALENs bind their targeted sequence in genomic DNA and generate a DSB at a nearby sequence that is then repaired by NHEJ or HR. In contrast, in the CRISPR-Cas system, base pairing between a noncoding RNA and genomic DNA targets the Cas endonuclease to generate a DSB at a specific place in the genome. A major technical advantage of CRISPR-Cas over ZFNs and TALENs is that it is much easier to produce RNAs than DNA-binding domains that bind a unique sequence in a DNA genome.

The **CRISPR-Cas (clustered, regularly interspaced short palindromic repeats-CRISPR-associated protein)** technology is derived from a bacterial immune system that protects bacteria against foreign plasmid and bacteriophage DNA (as discussed in [Chapter 6](#)). For instance, the bacterium *Streptococcus pyogenes*, which contains a relatively simple CRISPR-Cas system, stores the memory of encounters with foreign DNA by integrating 20-nucleotide sequences from the foreign DNA into a particular place in the bacterial genome called a CRISPR array. When *S. pyogenes* is attacked for a second time by the same foreign DNA, guide RNAs (gRNAs) that contain the 20-nucleotide sequences are produced by cutting apart a long RNA transcript from the CRISPR array. A sequence common to all gRNAs base pairs with another noncoding RNA called a trans-activating CRISPR RNA (tracrRNA), and the dual RNAs form a complex with a Cas protein called Cas9. The gRNA then directs Cas9 to produce a DSB at a location in the foreign DNA that contains the complementary 20-nucleotide target sequence located next to a trinucleotide NGG protospacer adjacent motif (PAM). Cas9 has two separate endonuclease domains that cut three nucleotides upstream of the PAM; one domain cuts the target strand that base pairs to the gRNA, and the other cuts the nontarget strand.

In 2012, the laboratories of Emmanuelle Charpentier and Jennifer Doudna demonstrated that when a gRNA and a tracrRNA are modified to be in the same transcript, called a **single guide RNA (sgRNA)**, the sgRNA retains the ability of the separate RNAs to assemble with Cas9, base

pair to targeted sequences in DNA, and activate the endonuclease activities of Cas9. This simplification made it practical to employ the CRISPR-Cas9 system in the laboratory to make modifications to eukaryotic genomes with high efficiency and specificity. Application of the CRISPR-Cas9 technology involves two plasmids: one expresses Cas9 protein, and the other expresses an sgRNA with a 20-nucleotide guide sequence that is designed by a researcher to be complementary to a specific genomic site adjacent to a PAM ([Figure 10-28](#)). After both plasmids are introduced into cells or organisms, Cas9 protein and sgRNA are expressed and form a complex that produces a DSB in the targeted gene. Inaccurate repair of the DSB by NHEJ causes gene inactivation. In contrast, specific mutations in a gene can be introduced by inclusion of a third plasmid, a donor plasmid, that is used for repair of the DSB by HR because it contains sequences identical to those that flank the site of cleavage in addition to the specified mutations.



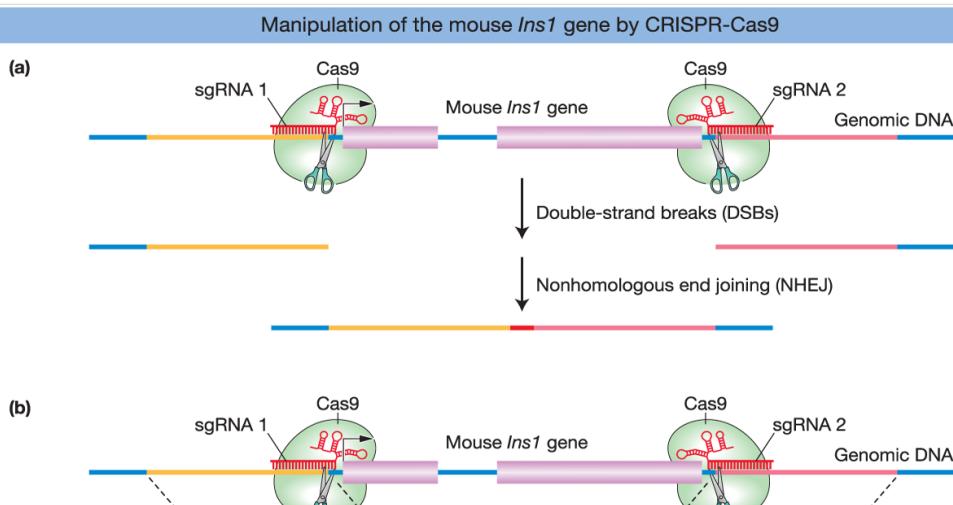
Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020 W. H. Freeman and Company

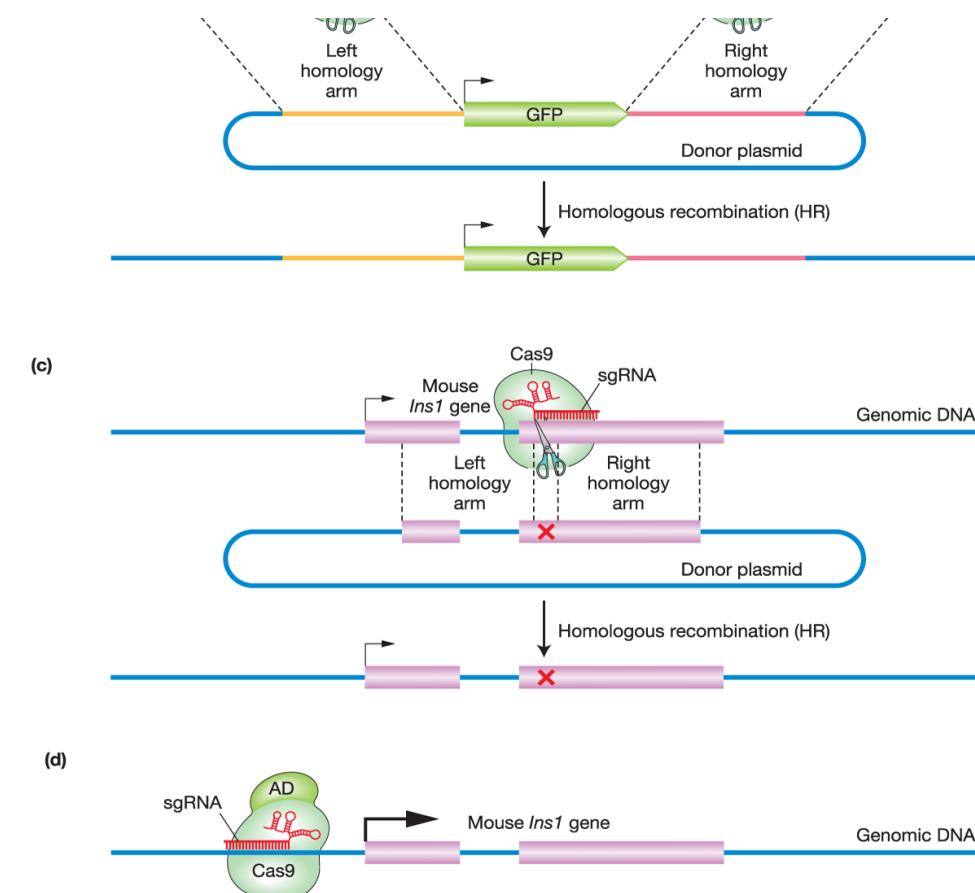
FIGURE 10-28 To target a DSB in a genome, the sequence of an sgRNA is designed to base pair to a target site, which is chosen in part because of its proximity to a PAM (orange DNA). A complex consisting of the sgRNA and Cas9 binds the target site, and the endonuclease domains of Cas9 produce a DSB (scissors). The DSB is then repaired either by homologous recombination using a supplied donor DNA, which creates precise mutations, or by nonhomologous end joining, which generates small insertions, base substitutions, or small deletions.

ANIMATED ART SaplingPlus

CRISPR

CRISPR-Cas9 gene editing occurs in reproductive cells of many organisms, making targeted gene modifications heritable. Therefore, CRISPR-Cas9 can be used for essentially all of the same reverse genetic purposes as transgenesis and even more. For example, to study the phenotypic consequences of loss of one of the two insulin genes in mice, a null allele of *Ins1* can be generated by CRISPR-Cas9-mediated deletion of the *Ins1* gene. This is accomplished by expressing two sgRNAs that create DSBs at the 5' and 3' ends of the *Ins1* gene, which will cut out the *Ins* gene and trigger repair of the broken chromosome by NHEJ ([Figure 10-29a](#)). Alternatively, the *Ins1* gene can be replaced by a reporter gene such as GFP. This is done by expressing the same two sgRNAs as well as a donor plasmid containing the GFP gene flanked by homology arms, sequences identical to those found upstream and downstream of the *Ins1* gene ([Figure 10-29b](#)). Repair of the DSBs by HR using the donor plasmid will replace the *Ins1* gene with the GFP gene. An advantage of this approach is that mice carrying null alleles of *Ins1* can be easily identified because they will express GFP in β cells under control of *Ins1* transcriptional regulatory sequences. Lastly, a donor plasmid could contain a missense or nonsense mutation that inactivates or alters the function of the *Ins1* gene ([Figure 10-29c](#)).





Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020 W. H. Freeman and Company

FIGURE 10-29 Four examples of the use of CRISPR-Cas9 to manipulate the mouse *Ins1* gene. (a) Two sgRNAs are designed to target DSBs to the ends of the *Ins1* gene. The chromosome break is then repaired by nonhomologous end joining, which deletes the *Ins1* gene and may introduce other sequence changes, indicated by the red line. (b) Addition of a donor plasmid triggers repair by homologous recombination, which, in this case, replaces the *Ins1* gene with the GFP gene. (c) Repair by homologous recombination of a single DSB in the *Ins1* gene using an altered *Ins1* donor DNA can generate missense and nonsense mutations as well as insertions and deletions in the *Ins1* gene. (d) sgRNA-mediated targeting of a Cas9 protein with inactivated endonuclease domains fused to a transcriptional activation domain (AD) yields enhanced transcription of the *Ins1* gene (bold arrow).

ANIMATED ART  SaplingPlus

CRISPR

Researchers have also modified the CRISPR-Cas9 system to manipulate gene expression in specific ways. The basic idea behind these technologies is that, when complexed with an sgRNA, a mutant Cas9 protein lacking its endonuclease activity can transport any protein or protein domain to a specific place in the genome. As shown in [Figure 10-29d](#), transcription of the mouse *Ins1* gene can be activated by sgRNA-mediated targeting to the *Ins1* gene of a Cas9 protein that is converted into a transcription factor by the addition of a transcriptional activation domain.

Conversely, *Ins1* transcription can be turned off by fusion of a transcriptional repression domain to Cas9. One can imagine that someday CRISPR-Cas9 technologies will be used to treat individuals with type I diabetes by manipulating the sequence or expression of genes.

KEY CONCEPT The CRISPR-Cas9 system efficiently and specifically changes the sequence of targeted genes in an organism, and modified versions of the system alter gene expression without changing gene sequences.

SUMMARY

Recombinant DNA is constructed in the laboratory to allow researchers to manipulate and analyze DNA segments (donor DNA) from any genome or a DNA copy of mRNA. Three sources of donor DNA are (1) genomes digested with restriction enzymes, (2) PCR products of specific DNA regions, and (3) cDNA copies of mRNAs. Sequencing of DNA by the dideoxy (Sanger) method is used to confirm the accuracy of recombinant DNA molecules and also to discover information stored in genomic DNA and mutant genes.

The polymerase chain reaction is a powerful method for direct amplification of a small sequence of DNA from within a complex mixture of DNA, without the need for a host cell or very much starting material. The key is to have primers that are complementary to flanking regions on each of the two DNA strands. These regions act as sites for polymerization. Multiple rounds of denaturation, annealing, and extension amplify the sequence of interest exponentially.

To insert donor DNA into vectors, donor and vector DNA are cut by the same restriction endonuclease, joined by annealing the sticky ends that result from digestion, and ligated to covalently join the molecules. PCR products and cDNA molecules are inserted into vectors by first adding restriction sites to the 5' end of PCR primers or by ligating short adapters containing restriction sites to their ends before insertion into the vector. Assembly methods that do not require restriction sites have made the construction of recombinant DNA molecules more flexible and efficient.

There are a wide variety of bacterial vectors. The choice of vector depends largely on the size of DNA fragment to be cloned. Plasmids are used to clone small genomic DNA fragments, PCR products, or cDNAs. Intermediate-size fragments, such as those resulting from digestion of genomic DNA, can be cloned into modified versions of λ bacteriophage (for inserts of 10–15 kb) or into phage–plasmid hybrids called fosmids (for inserts of 35–45 kb). Finally, bacterial artificial chromosomes (BACs) are used routinely to clone very large genomic fragments (~100–200 kb). A variety of plasmids have been developed that contain features that make it easier to clone DNA fragments and to control the expression of constituent genes in different organisms.

The vector-donor DNA construct is amplified inside bacterial host cells as extrachromosomal molecules that are replicated when the host is replicating its genome. Amplification of plasmids, phages, and BACs results in clones containing multiple copies of each recombinant DNA construct. In contrast, only a single fosmid is present in each bacterial cell.

Often, finding a specific clone with a gene of interest requires the screening of a genomic library, a set of clones, ligated in the same vector, that together represent all regions of the genome of the organism in question. The number of clones that constitute a genomic library depends on (1) the size of the genome in question, and (2) the insert size tolerated by the particular cloning-vector system. Similarly, a cDNA library is a representation of the total mRNA set produced by a tissue or developmental stage in a given organism.

Hybridization with single-stranded nucleic acid probes is fundamental to both *in vitro* and *in vivo* methods for identifying DNA fragments or RNAs of interest. These methods include Southern blotting for DNA, Northern blotting for RNA, and screening of genomic and cDNA libraries. In contrast, labeled antibodies are probes for identifying specific proteins from complex mixtures in Western blotting or immunofluorescence.

Transgenes are engineered DNA molecules that are introduced and expressed in eukaryotic cells. They can be used to engineer a novel mutation or to study the regulatory sequences that constitute part of a gene. Transgenes can be introduced as extrachromosomal molecules, or they can be integrated into a chromosome, either in random (ectopic) locations or in place of the homologous gene, depending on the system. Typically, the mechanisms used to introduce a transgene depend on an understanding and exploitation of the reproductive biology of the organism. New genome engineering methods like the CRISPR-Cas9 system are being developed whose defining features are the creation and repair of site-specific DNA double-strand breaks. These methods have opened the door to new and exciting reverse genetic studies in a wide variety of eukaryotic organisms, and they could potentially be used for gene-editing therapies for patients with serious diseases.

KEY TERMS

[antibody](#)

[autoradiography](#)

[bacterial artificial chromosome \(BAC\)](#)

[cDNA library](#)

[chimera \(chimeric\)](#)

[clustered, regularly interspaced short palindromic repeats \(CRISPR\)](#)

[complementary DNA \(cDNA\)](#)

[CRISPR-associated protein \(Cas\)](#)

[dideoxy \(Sanger\) sequencing](#)

DNA amplification
DNA assembly
DNA cloning
DNA ligase
DNA linker (DNA adapter)
DNA technologies
donor DNA (insert DNA)
ectopic (ectopically)
epitope tag
fluorescence in situ hybridization (FISH)
fosmid
gel electrophoresis
gene knockout
gene replacement
genetically modified organism (GMO)
genetic engineering
genomic library
genomics
homologous recombination (HR)
hybridization
immunofluorescence
infection
in situ hybridization (ISH)
multiple cloning site (MCS) (polylinker)
nonhomologous end joining (NHEJ)
Northern blotting
palindrome (palindromic)
plasmid
polymerase chain reaction (PCR)
position effect
probe
quantitative PCR (qPCR)
recombinant DNA
restriction enzyme
restriction fragment
restriction map
restriction site
reverse transcriptase

[reverse transcription-PCR \(RT-PCR\)](#)

[single guide RNA \(sgRNA\)](#)

[Southern blotting](#)

[Ti plasmid](#)

[transduction](#)

[transformation](#)

[transgene](#)

[transgenic organism](#)

[vector](#)

[Western blotting](#)

SOLVED PROBLEMS

SOLVED PROBLEM 1

In [Chapter 9](#), we studied the structure of tRNA molecules. Suppose that you want to clone a fungal gene that encodes a certain tRNA. You have a sample of the purified tRNA and an *E. coli* plasmid that contains a single *Eco*RI cutting site in a *tet*^R (tetracycline-resistance) gene, as well as a gene for resistance to ampicillin (*amp*^R). How can you clone the gene of interest?

SOLUTION

You can use the tRNA itself or a cloned cDNA copy of it to probe for the DNA containing the gene. One method is to digest the genomic DNA with *Eco*RI and then mix it with the plasmid, which you also have cut with *Eco*RI. After transformation of an *amp*^S *tet*^S recipient, select *Amp*^R colonies, indicating successful transformation. Of these *Amp*^R colonies, select the colonies that are *Tet*^S. These *Tet*^S colonies will contain vectors with inserts in the *tet*^R gene, and a great number of them are needed to make the library. Test the library by using the tRNA as the probe. Those clones that hybridize to the probe will contain the gene of interest. Alternatively, you can subject *Eco*RI-digested genomic DNA to gel electrophoresis and then identify the correct band by probing with the tRNA. This region of the gel can be cut out and used as a source of enriched DNA to clone into the plasmid cut with *Eco*RI. You then probe these clones with the tRNA to confirm that these clones contain the gene of interest.

PROBLEMS

Visit SaplingPlus for supplemental content. Problems with the  icon are available for review /grading.

WORKING WITH THE FIGURES

(The first 33 questions require inspection of text figures.)

1. In the opening figure, what would happen if the cathode and anode were switched during gel electrophoresis?
2. In [Figure 10-1](#), by what methods could the plasmid be introduced into bacteria?
3. In [Figure 10-2](#), why can both DNA and RNA be used as probes for both Southern and Northern blot analysis?
4. In [Figure 10-3a](#), what changes occur to the human *Ins* pre-mRNA to produce the mature *Ins* mRNA?
5. In [Figure 10-3b](#), what size fragment would probe 2 detect in Southern blot analysis of mouse genomic DNA digested with *Pvu*II, and what size fragment would probe 3 detect in an *Nsi*I digest?

6. In [Figure 10-4](#), which of the seven lanes contains the smallest piece of DNA?
7. In [Figure 10-5](#), what is the purpose behind transferring the nucleic acid from a gel to a membrane?
8. In [Figure 10-6c](#), what size band would be detected by probe 1 in Southern blot analysis of human smooth muscle cell genomic DNA digested with *Eco*RI?
9. Examine [Table 10-1](#). Draw the staggered ends produced by digestion with *Not*I and the blunt ends produced by *Msp*I.
10. In [Figure 10-7](#), why would PCR not work if *Taq* DNA polymerase was replaced with a DNA polymerase from human cells?
11. In [Figure 10-8](#), if Sample A had a C_T value of 24 and Sample B had a C_T value of 27, which sample had more DNA and how much more?

12. In [Figure 10-9](#), explain why the polymerase used for synthesizing DNA from an RNA template is called reverse transcriptase.

13. In [Figure 10-10a](#), in the second step of the procedure, label the 5' and 3' ends of the linearized vector and the *Eco*RI insert.
14. In [Figure 10-11a](#), which colonies (blue or white) contain plasmids with a DNA insert?
15. In [Figure 10-12](#), determine approximately how many BAC clones are needed to provide 1 \times coverage of
- the yeast genome (12 Mbp).
 - the *E. coli* genome (4.1 Mbp).
 - the fruit fly genome (130 Mbp). 
16. In [Figure 10-13](#), what is the difference between plasmid transformation and fosmid transduction?
17. In [Figure 10-14](#), is it possible for more than one insert to ligate into a single vector? Why or why not?
18. In [Figure 10-15](#), how is screening a genomic library similar to Southern blot analysis?
19. In [Figure 10-16](#), how would you modify the restriction sites in the primers so that PCR products only insert into a vector in one orientation?
20. In [Figure 10-17a](#), what enzymes would you use to cut a full-length insert back out of the vector?
21. In [Figure 10-18a](#), draw a ribonucleotide that would act as a chain terminator during transcription.
22. In [Figure 10-18b](#), analogous to the drawing for the sequencing reaction that contains ddTTP, write the sequence of the first three termination products of the sequencing reaction that contains ddCTP. 
23. In [Figure 10-19](#), what would happen to the height of the peak at the SNP position in an individual that did not contain a SNP?
24. In [Figure 10-20](#), why are multiple procedures needed for introducing DNA into cells?
25. In [Figure 10-21](#), how can the marker be used to determine if a single or double crossover event occurred?
26. In [Figure 10-22](#), what is the purpose of the selectable marker? Provide two examples of selectable markers. 
27. In [Figure 10-23](#), do all of the cells of a transgenic plant grown from one clump of cells contain T-DNA? Justify your answer.

28. In [Figure 10-24](#), what is distinctive about the syncytial region that makes it a good place to inject DNA?
29. In [Figure 10-25](#), why do the fertilized eggs have two nuclei? What is the ploidy of each nucleus?
30. In [Figure 10-26c](#), does the selection procedure distinguish whether the targeting vector inserted in one or both copies of the homologous gene in the diploid ES cells?
31. In [Figure 10-27a](#), why are chimeric males, rather than females, used in the mating crosses to generate a homozygous mutant mouse line?
32. In [Figure 10-28](#), how would you determine the sequence change resulting from nonhomologous end joining (NHEJ) at a double-strand break site created by CRISPR-Cas9? 
33. In [Figure 10-29b](#), draw an analogous figure that shows a scheme for replacing the second exon in *Ins1* with the GFP gene.

BASIC PROBLEMS

34. Why is a range of temperatures indicated in the annealing step of PCR?
35. In the PCR process, if we assume that each cycle takes 5 minutes, what fold amplification would be accomplished in 1 hour? 
36. Would the blot be called a Southern, Northern, or Western blot if RNA was on the membrane and the probe was single-stranded DNA?
37. How can genomic and cDNA sequences be used to determine where introns are located in genes? Use [Figure 10-3a](#) to illustrate your answer.
38. In [Figure 10-10](#), it is possible that two or more copies, rather than one copy, of the insert ligated into the vector. How would you test this possibility **(a)** using restriction enzymes, **(b)** by PCR, and **(c)** by Southern blot? **Hints:** The complete sequence of the vector is known, and the *Eco*RI site is one of many restriction sites in the polylinker of the vector.
39. Write out the sequence of 20-nucleotide primers to be used for PCR to amplify the region of interest in the following piece of DNA: 

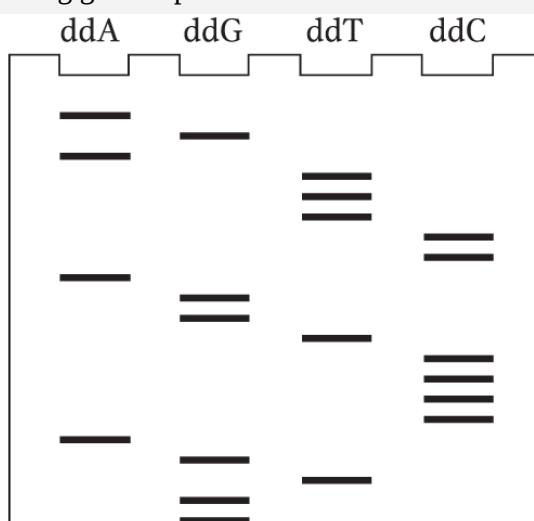
5'-CCGTAACACGTCAGGGCTAACAGG	Region of interest	-TTGACAATGCCTGGAATTCTGTAAC-3'
3'-GGCATTGTGCAGTCCCGGATTGTCC		-AACTGTTACGGACCTTAAGACATTG-5'

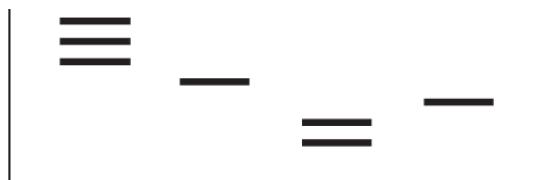
- To 40. Draw a diagram that explains how automated dideoxy sequencing reactions are analyzed by capillary gel electrophoresis and laser detection.

41. Explain why dideoxynucleosides, which are converted to dideoxynucleotides in human cells, are effective drugs to block replication of the human immunodeficiency virus (HIV) genome by the HIV reverse transcriptase.
42. Compare and contrast the use of the word *recombinant* as used in the phrases **(a)** “recombinant DNA” and **(b)** “recombinant frequency.”
43. Why is DNA ligase needed to make recombinant DNA? What would be the immediate consequence in the cloning process if DNA ligase was not included in the reaction?
44. In [Figure 10-26](#), describe how positive-negative selection is used to find rare homologous recombination-mediated gene targeting events.
45. In T-DNA transformation of a plant with a transgene from a fungus (not found in plants), the presumptive transgenic plant does not display the expected phenotype of the transgene. How would you determine whether the transgene is in fact inserted in the plant genome? How would you determine whether the transgene mRNA and protein are expressed in the plant? 
46. Why was cDNA and not genomic DNA used to express human insulin in *E. coli*?
47. Based on the information presented in [Figures 10-26](#) and [10-29](#), explain how CRISPR-Cas9 could be used to knock out the *Ins1* gene in mice. In particular, what RNAs and proteins would be expressed from plasmids that are injected into fertilized mouse eggs?

CHALLENGING PROBLEMS

48. Using the information in [Figures 10-3](#) and [10-26a](#), draw a targeting vector that could be used to tag the human insulin protein with GFP at the N-terminus.
49. Diagram how Gibson assembly could be used to construct the targeting vector for question 48.
50. A cloned fragment of DNA was sequenced by using the dideoxy method. A part of the autoradiogram of the sequencing gel is represented here.





Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020 W. H. Freeman and Company

Th

- a. Write out the nucleotide sequence of the DNA molecule synthesized from the primer. Label the 5' and 3' ends.
 - b. Write out the nucleotide sequence of the DNA molecule used as the template strand. Label the 5' and 3' ends.
 - c. Write out the nucleotide sequence of the DNA double helix. Label the 5' and 3' ends.
51. Transgenic tobacco plants were obtained in which the vector Ti plasmid was designed to insert the gene of interest plus an adjacent kanamycin-resistance gene. The inheritance of chromosomal insertion was followed by testing progeny for kanamycin resistance. Two plants typified the results obtained generally. When plant 1 was backcrossed with wild-type tobacco, 50 percent of the progeny were kanamycin resistant and 50 percent were sensitive. When plant 2 was backcrossed with the wild type, 75 percent of the progeny were kanamycin resistant and 25 percent were sensitive. What must have been the difference between the two transgenic plants? What would you predict about the situation regarding the gene of interest?
52. The sequence of the *actin* gene in the haploid fungus *Neurospora* is known from the complete genome sequence. If you had a slow-growing mutant that you suspected of being an *actin* mutant, how would you use (a) restriction enzyme cloning and sequencing, (b) PCR and sequencing, and (c) restriction enzyme cloning and functional complementation (rescue) to determine whether your suspicion is correct?
53. Bacterial glucuronidase converts a colorless substance called X-Gluc into a bright blue indigo pigment. The gene for glucuronidase also works in plants if given a plant promoter region. How would you use this gene as a reporter gene to find the tissues in which a plant gene that you have just cloned is normally active? (Assume that X-Gluc is easily taken up by the plant tissues.)
54. The plant *Arabidopsis thaliana* was transformed using the Ti plasmid into which a kanamycin-resistance gene had been inserted in the T-DNA region. Two kanamycin-resistant colonies (A and B) were selected, and plants were regenerated from them. The plants were allowed to self-pollinate, and the results were as follows:

Plant A selfed → 34 progeny resistant to kanamycin
14 progeny sensitive to kanamycin
Plant B selfed → 15 progeny resistant to kanamycin
16 progeny sensitive to kanamycin

Plant A selfed →	$\frac{3}{4}$	progeny resistant to kanamycin
	$\frac{1}{4}$	progeny sensitive to kanamycin
Plant B selfed →	$\frac{15}{16}$	progeny resistant to kanamycin
	$\frac{1}{16}$	progeny sensitive to kanamycin

- Draw the relevant plant chromosomes in both plants.
- Explain the two different ratios. 

GENETICS AND SOCIETY

In 2018, a researcher claimed to have used the CRISPR-Cas9 genome editing technique to produce the world's first gene-edited babies. The researcher announced that they edited the *CCR5* gene in two embryos, which were then implanted in a woman. *CCR5* encodes a receptor that is expressed on the surface of white blood cells and other cells, where it coordinates immune responses. *CCR5* is also the main receptor used by the human immunodeficiency virus (HIV) to gain entry into cells, which is necessary for its replication. Genetic variations in the *CCR5* gene have been identified in the human population that confer natural resistance to HIV infection. This includes a *CCR5* allele called delta-32 that is missing 32 base pairs from the coding region of the gene, causing a deletion and a frameshift in the encoded *CCR5* protein that blocks its expression on the cell surface. Therefore, to create babies that were resistant to HIV infection, the researcher used the CRISPR-Cas9 editing technique to produce a deletion in the *CCR5* gene that was similar to delta-32. This use of the CRISPR-Cas9 technique has prompted a great deal of discussion about the scientific and ethical implications of making heritable changes to the human genome. Given your newfound insights into the CRISPR-Cas9 technique and genetic phenomena, what are your concerns?