CHAPTER **TEN**

10

Analyzing the Structure and Function of Genes

Since the turn of the century, biologists have amassed an unprecedented wealth of information about the genes that direct the development and behavior of living things. Thanks to advances in our ability to rapidly determine the nucleotide sequence of entire genomes, we now have access to the complete molecular blueprints for thousands of different organisms, from the platypus to the plague bacterium, and for thousands of different people from all over the world.

This information explosion was ignited by technological advances that allowed the isolation and manipulation of a selected piece of DNA from among the many millions of nucleotide pairs in a typical chromosome. Investigators then developed powerful techniques for replicating, sequencing, and modifying this DNA—and even introducing it into other organisms that can then be studied in the laboratory.

These technical breakthroughs have had a dramatic impact on all aspects of cell biology. They have advanced our understanding of the organization and evolutionary history of complex eukaryotic genomes (as discussed in Chapter 9) and have led to the discovery of whole new classes of genes, RNAs, and proteins. They continue to generate new ways of determining the functions of genes and proteins in living organisms, and they provide an important set of tools for unraveling the mechanisms—still poorly understood—by which a complex organism can develop from a single fertilized egg.

At the same time, our ability to manipulate DNA has had a profound influence on our understanding and treatment of disease: using these techniques, we can now detect the mutations in human genes that are responsible for inherited disorders or that predispose us to a variety of

QUESTION 10–1

DNA sequencing of your own two β-globin genes (one from each of your two Chromosome 11s) reveals a mutation in one of the genes. Given this information alone, should you worry about being a carrier of an inherited disease that could be passed on to your children? What other information would you like to have to assess your risk?

common diseases, including cancer. We can also produce an increasing number of pharmaceuticals, such as insulin for diabetics and blood-clotting proteins for hemophiliacs.

In this chapter, we present a brief overview of how we can manipulate DNA, identify genes, and produce many copies of any given nucleotide sequence in the laboratory. We discuss several ways to explore gene function, including recent approaches to DNA sequencing and to modifying or inactivating genes in cells, animals, and plants. These methods—which are continuously being improved and made more powerful—are not only revolutionizing the way we do science, but are transforming our understanding of cell biology and human disease.

ISOLATING AND CLONING DNA MOLECULES

Humans have been experimenting with DNA, albeit without realizing it, for millennia. The roses in our gardens, the corn on our plate, and the dogs in our yards are all the product of selective breeding that has taken place over many, many generations (**Figure 10–1**). But it wasn't until the 1970s that we could begin to engineer organisms with desired properties by directly tinkering with their genes.

Isolating and manipulating individual genes is not a trivial matter. Unlike a protein, a gene does not exist as a discrete entity in cells; it is a small part of a much larger DNA molecule. Even bacterial genomes, which are much less expansive and complex than the chromosomes of eukaryotes, are still enormously long. The *E. coli* genome, for example, contains 4.6 million nucleotide pairs.

How, then, can we go about separating a single gene from a eukaryotic genome—which is considerably larger than that of a bacterium—so that it can be handled in the laboratory? The solution to this problem emerged, in large part, with the discovery of a class of bacterial enzymes that cut double-stranded DNA at particular sequences. These enzymes can be used to produce a reproducible set of specific DNA fragments from any genome—including fragments that harbor genes. The desired fragment is then amplified, producing many identical copies, by a process called **DNA cloning**. It is this amplification that makes it possible to separate a gene of interest from the rest of the genome.

In this section, we describe how specific DNA fragments can be generated, isolated, and produced in large quantities in bacteria—the classical approach to DNA cloning. In the next section of the chapter, we present

(A) (B)

Figure 10–1 Selective breeding is, in essence, a form of genetic manipulation. (A) The oldest known depiction of a rose in Western art, from the palace of Knossos in Crete, around 2000 BC. Modern roses are the result of centuries of breeding between such wild roses. (B) Dogs have been bred to exhibit a wide variety of characteristics, including different head shapes, coat colors, and of course size. All dogs, regardless of breed, belong to a single species that was domesticated from the gray wolf some 10,000 to 15,000 years ago. (B, from A.L. Shearin & E.A. Ostrander, *PLoS Biol.* 8:e1000310, 2010.)

an alternative approach to cloning DNA: this method, which is carried out in a test tube, uses a special form of DNA polymerase to make copies of the desired nucleotide sequence.

Restriction Enzymes Cut DNA Molecules at Specific Sites

Like many of the tools of DNA technology, the enzymes used to prepare DNA fragments for cloning were discovered by researchers trying to understand an intriguing biological phenomenon. It had been observed that certain bacteria always degraded "foreign" DNA that was introduced into them experimentally. A search for the underlying mechanism revealed a novel class of enzymes that cleave DNA at specific nucleotide sequences. Because these enzymes function to restrict the transfer of DNA between strains of bacteria, they were called **restriction enzymes**, or *restriction nucleases*. The pursuit of this seemingly arcane biological puzzle set off the development of technologies that have forever changed the way cell and molecular biologists study living things.

Different bacterial species produce different restriction enzymes, each cutting at a different, specific nucleotide sequence (**Figure 10–2**). The bacteria's own DNA is protected from cleavage by chemical modification of these specific sequences. Because these target sequences are short—generally four to eight nucleotide pairs—many sites of cleavage will occur, purely by chance, in any long DNA molecule. The reason restriction enzymes are so useful in the laboratory is that each enzyme will cut a particular DNA molecule at the same sites. Thus for a given sample of DNA, a particular restriction enzyme will reliably generate the same set of DNA fragments.

The size of the resulting fragments depends on the target sequences of the restriction enzymes. As shown in Figure 10–2, the enzyme HaeIII cuts at a sequence of four nucleotide pairs; a sequence this long would be expected to occur purely by chance approximately once every 256 nucleotide pairs (1 in 4^4). In comparison, a restriction enzyme with a target sequence that is eight nucleotides long would be expected to cleave DNA on average once every 65,536 nucleotide pairs (1 in 4^8). This difference in sequence selectivity makes it possible to cleave a long DNA molecule into the fragment sizes that are most suitable for a given application.

Gel Electrophoresis Separates DNA Fragments of Different Sizes

After a large DNA molecule is cleaved into smaller pieces with a restriction enzyme, the DNA fragments can be separated from one another on the basis of their length by gel electrophoresis—the same method used to separate mixtures of proteins (see Panel 4–5, p. 167). A mixture of DNA

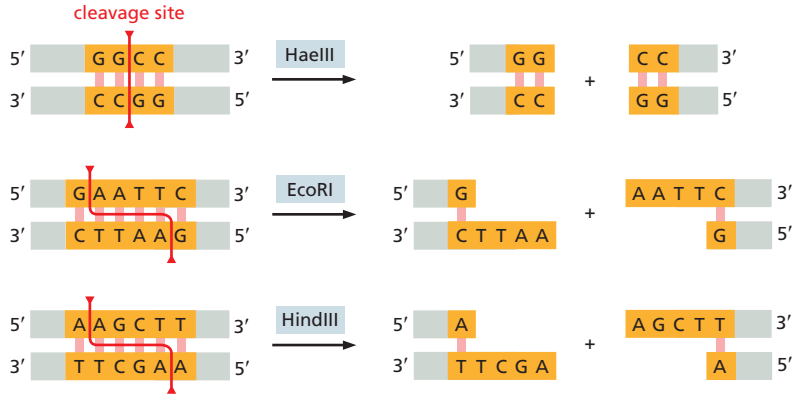


Figure 10–2 Restriction enzymes cleave both strands of the DNA double helix at specific nucleotide sequences. Target sequences (*orange*) are often palindromic—that is, the nucleotide sequence is symmetrical around a central point. Some enzymes, such as HaeIII, cut straight across the double helix and leave two blunt-ended DNA molecules; with others, such as EcoRI and HindIII, the cuts on each strand are staggered. These staggered cuts generate "sticky ends"—short, single-stranded overhangs that help the cut DNA molecules join back together through complementary base-pairing. This rejoining of DNA molecules becomes important for DNA cloning, as we discuss shortly. Restriction enzymes are usually obtained from bacteria, and their names reflect their origins: for example, the enzyme EcoRI comes from *E. coli*.

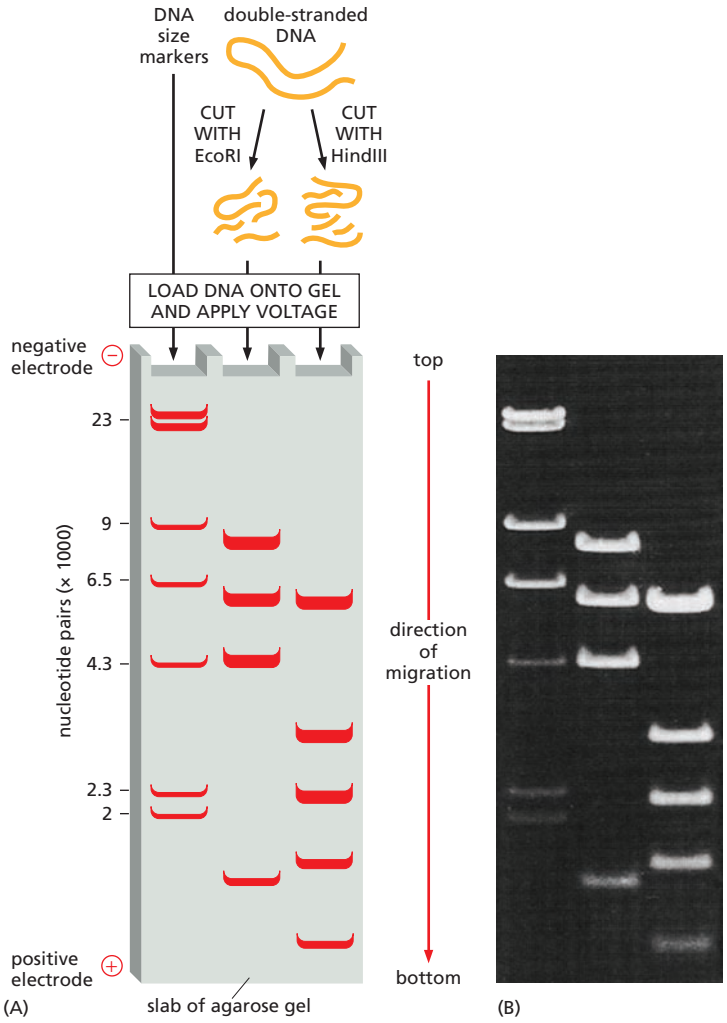


Figure 10–3 DNA molecules can be separated by size using gel electrophoresis. (A) Schematic illustration compares the results of cutting the same DNA molecule (in this case, the genome of a virus that infects parasitic wasps) with two different restriction enzymes, EcoRI (*middle*) and HindIII (*right*). The fragments are then separated by gel electrophoresis. Because larger fragments migrate more slowly than smaller ones, the lowermost bands on the gel contain the smallest DNA fragments. The sizes of the fragments can be estimated by comparing them to a set of DNA fragments of known sizes (*left*). (B) Photograph of an actual gel shows the positions of DNA bands that have been labeled with a fluorescent dye. (B, from U. Albrecht et al., *J. Gen. Virol.* 75: 3353–3363, 1994. With permission from the Microbiology Society.)

fragments is loaded at one end of a slab of agarose or polyacrylamide gel, which contains a microscopic network of pores. When a voltage is applied across the gel, the negatively charged DNA fragments migrate toward the positive electrode; larger fragments migrate more slowly because their progress is impeded to a greater extent by the gel matrix. Over several hours, the DNA fragments become spread out across the gel according to size, forming a ladder of discrete bands, each composed of a collection of DNA molecules of identical length (**Figure 10–3**).

The separated DNA bands on an agarose or polyacrylamide gel are not, by themselves, visible. To see these bands, the DNA must be labeled or stained in some way. One sensitive method involves exposing the gel to a dye that fluoresces under ultraviolet (UV) light when it is bound to DNA. When the gel is placed on a UV light box, the individual bands glow bright orange—or bright white when the gel is photographed in black and white

QUESTION 10–2

Which products result when the double-stranded DNA molecule *below* is digested with (A) EcoRI, (B) HaeIII, (C) HindIII, or (D) all three of these enzymes together? (See Figure 10–2 for the target sequences of these enzymes.)

5′-AAGAATTGCGGAATTCGGGCCTTAAGCGCCGCGTCGAGGCCTTAAA-3′
3′-TTCTTAACGCCTTAAGCCCGGAATTCGCGGCGCAGCTCCGGAATTT-5′

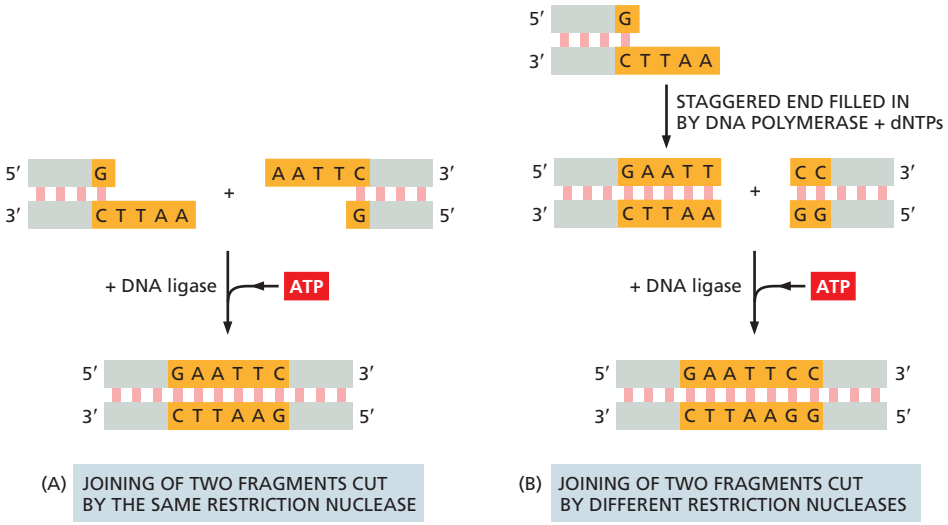


Figure 10–4 DNA ligase can join together any two DNA fragments *in vitro* to produce recombinant DNA molecules. The fragments joined by DNA ligase can be from different cells, tissues, or even different organisms. ATP provides the energy necessary to reseal the sugar–phosphate backbone of the DNA. (A) DNA ligase can readily join two DNA fragments produced by the same restriction enzyme, in this case EcoRI. Note that the staggered ends produced by this enzyme enable the ends of the two fragments to base-pair correctly with each other, greatly facilitating their rejoining. (B) DNA ligase can also be used to join DNA fragments produced by different restriction enzymes—for example, EcoRI and HaeIII. In this case, before the fragments undergo ligation, DNA polymerase plus a mixture of deoxyribonucleoside triphosphates (dNTPs) are used to fill in the staggered cut produced by EcoRI prior to ligation.

(see Figure 10–3B). To isolate a desired DNA fragment, the small section of the gel that contains the band is excised with a scalpel, and the DNA is then extracted.

DNA Cloning Begins with the Production of Recombinant DNA

Once a genome has been broken into smaller, more manageable pieces, the resulting fragments must then be prepared for cloning. This process involves inserting the DNA fragments into a carrier, or *vector*—another piece of DNA that can be copied inside cells. Because this union involves "recombining" DNA from different sources, the resulting molecules are called **recombinant DNA**. The production of recombinant DNA molecules in this way is a key step in the classical approach to DNA cloning.

Like the cutting of DNA by restriction enzymes, the joining together of DNA fragments to produce recombinant DNA molecules is made possible by an enzyme produced by cells. In this case, the enzyme is **DNA ligase**. In cells, DNA ligase reseals the nicks that arise in the DNA backbone during DNA replication and DNA repair (see Figure 6–19). In the laboratory, DNA ligase can be used to link together any two pieces of DNA in a test tube, producing recombinant DNA molecules that are not found in nature (**Figure 10–4**).

Recombinant DNA Can Be Copied Inside Bacterial Cells

The vectors used to carry the DNA that is to be cloned are small, circular DNA molecules called **plasmids** (**Figure 10–5**). Each plasmid contains its own replication origin, which enables it to replicate in a bacterial cell independently of the bacterial chromosome. This feature allows the DNA of interest to be produced in large amounts, even within a single bacterial cell. The plasmid also has cleavage sites for common restriction enzymes, so that it can be conveniently opened and a foreign DNA fragment inserted.

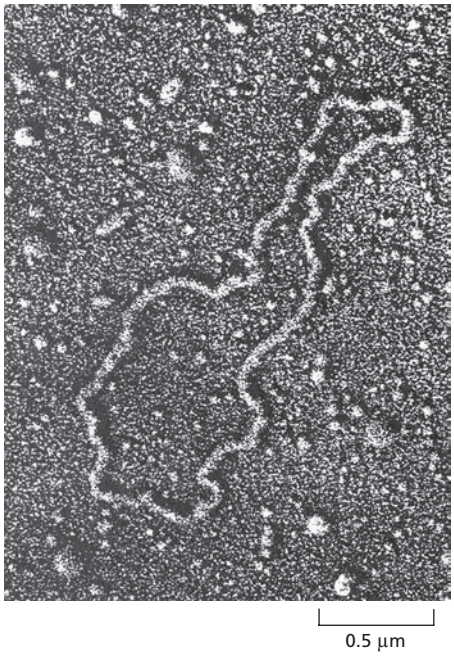


Figure 10–5 Bacterial plasmids are commonly used as cloning vectors. This circular, double-stranded DNA molecule was the first plasmid for DNA cloning; it contains about nine thousand nucleotide pairs. The staining procedure used to make the DNA visible in this electron micrograph causes the DNA to appear much thicker than it actually is. (Courtesy of Stanley N. Cohen, Stanford University.)

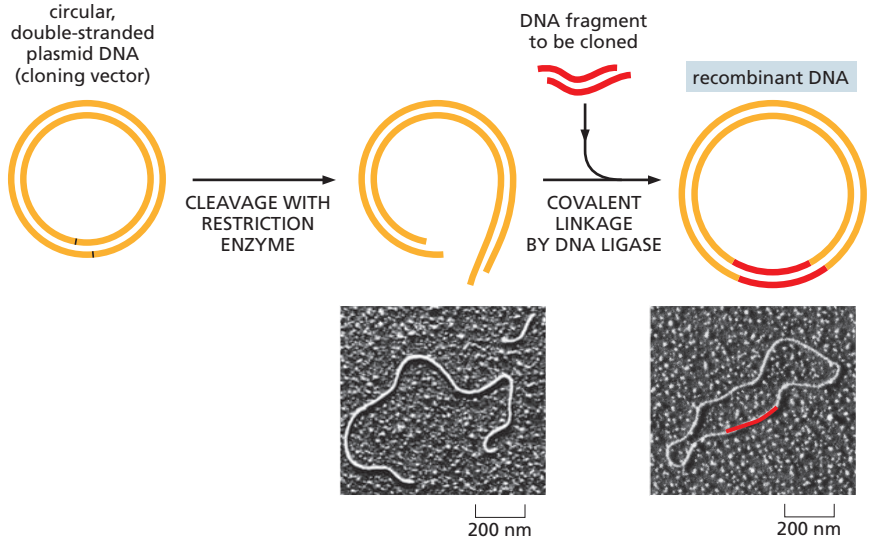


Figure 10–6 A DNA fragment is inserted into a bacterial plasmid using the enzyme DNA ligase. The plasmid is first cut open at a single site with a restriction enzyme (in this case, one that produces staggered ends). It is then mixed with the DNA fragment to be cloned, which has been cut with the same restriction enzyme. The staggered ends base-pair, and when DNA ligase and ATP are added, the nicks in the DNA backbone are sealed to produce a complete recombinant DNA molecule. In the accompanying micrographs, we have colored the DNA fragment *red* to make it easier to see. (Micrographs courtesy of Huntington Potter and David Dressler.)

The vectors used for cloning are streamlined versions of plasmids that occur naturally in many bacteria. Bacterial plasmids were first recognized by physicians and scientists because they often carry genes that render their microbial host resistant to one or more antibiotics. Indeed, historically potent antibiotics—penicillin, for example—are no longer effective against many of today's bacterial infections because plasmids that confer resistance to the antibiotic have spread among bacterial species by horizontal gene transfer (see Figure 9–15).

To insert a piece of DNA into a plasmid vector, the purified plasmid DNA is opened up by a restriction enzyme that cleaves it at a single site, and the DNA fragment to be cloned is then spliced into that site using DNA ligase (**Figure 10–6**). This recombinant DNA molecule is now ready to be introduced into a bacterium, where it will be copied and amplified.

To accomplish this feat, investigators take advantage of the fact that some bacteria naturally take up DNA molecules present in their surroundings. The mechanism that controls this uptake is called **transformation**, because early observations suggested it could "transform" one bacterial strain into another. Indeed, the first proof that genes are made of DNA came from an experiment in which DNA purified from a pathogenic strain of pneumococcus was used to transform a harmless bacterium into a deadly one (see How We Know, pp. 192–194).

In a natural bacterial population, a source of DNA for transformation is provided by bacteria that have died and released their contents, including DNA, into the environment. In a test tube, however, bacteria such as *E. coli* can be coaxed to take up recombinant DNA that has been created in the laboratory. These bacteria are then suspended in a nutrient-rich broth and allowed to proliferate.

Each time the bacterial population doubles—every 30 minutes or so—the number of copies of the recombinant DNA molecule also doubles. Thus, in 24 hours, the engineered cells will produce hundreds of millions of copies of the plasmid, along with the DNA fragment it contains. The bacteria can then be split open (lysed) and the plasmid DNA purified from the rest of the cell contents, including the large bacterial chromosome (**Figure 10–7**).

The DNA fragment can be readily recovered by cutting it out of the plasmid DNA with the same restriction enzyme that was used to insert it, and then separating it from the plasmid DNA by gel electrophoresis (see Figure 10–3). Together, these steps allow the amplification and purification of any segment of DNA from the genome of any organism.

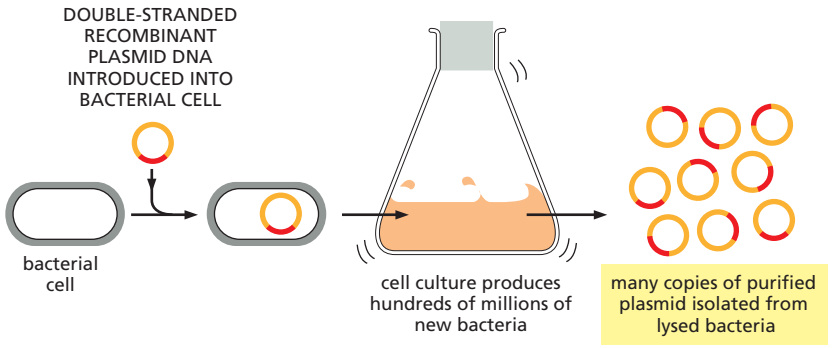


Figure 10–7 A DNA fragment can be replicated inside a bacterial cell. To clone a particular fragment of DNA, it is first inserted into a plasmid vector, as shown in Figure 10–6. The resulting recombinant plasmid DNA is then introduced into a bacterium, where it is replicated many millions of times as the bacterium multiplies. For simplicity, the genome of the bacterial cell is not shown.

An Entire Genome Can Be Represented in a DNA Library

When a whole genome is cut by a restriction enzyme, a large number of different DNA fragments is generated. This collection of DNA fragments can be ligated into plasmid vectors, under conditions that favor the insertion of a single DNA fragment into each plasmid molecule. These recombinant plasmids are then introduced into *E. coli* at a concentration that ensures that no more than one plasmid molecule is taken up by each bacterium. The resulting collection of cloned DNA fragments, present in the bacterial culture, is known as a **DNA library**. Because the DNA fragments were derived by digesting chromosomal DNA directly from an organism, the resulting collection—called a **genomic library**—should represent the entire genome of that organism (**Figure 10–8**). Such genomic libraries often provide the starting material for determining the complete nucleotide sequence of an organism's genome.

For other applications, however, it can be advantageous to work with a different type of library—one that includes only the coding sequences of genes; that is, a library that lacks intronic and other noncoding sequences that make up most eukaryotic DNA. For some genes, the complete genomic clone—including introns and exons—is too large and unwieldy to handle conveniently in the laboratory (see, for example, Figure 7–19B). What's more, the bacterial cells typically used to amplify cloned DNA are unable to remove introns from mammalian RNA transcripts. So if the goal is to use a cloned mammalian gene to produce a large amount of the protein it encodes, for example, it is essential to use only the coding sequence of the gene.

In this case, investigators generate a **cDNA library**. A cDNA library is similar to a genomic library in that it also contains numerous clones containing many different DNA sequences. But it differs in one important respect. The DNA that goes into a cDNA library is not genomic DNA; it is DNA copied from the mRNAs present in a particular type of cell. To prepare a cDNA library, all of the mRNAs are extracted, and double-stranded DNA copies of these mRNAs are produced by the enzymes *reverse transcriptase* and DNA polymerase (**Figure 10–9**). The resulting **complementary DNA**—or **cDNA**—molecules are then introduced into bacteria and amplified, as described for genomic DNA fragments (see Figure 10–8).

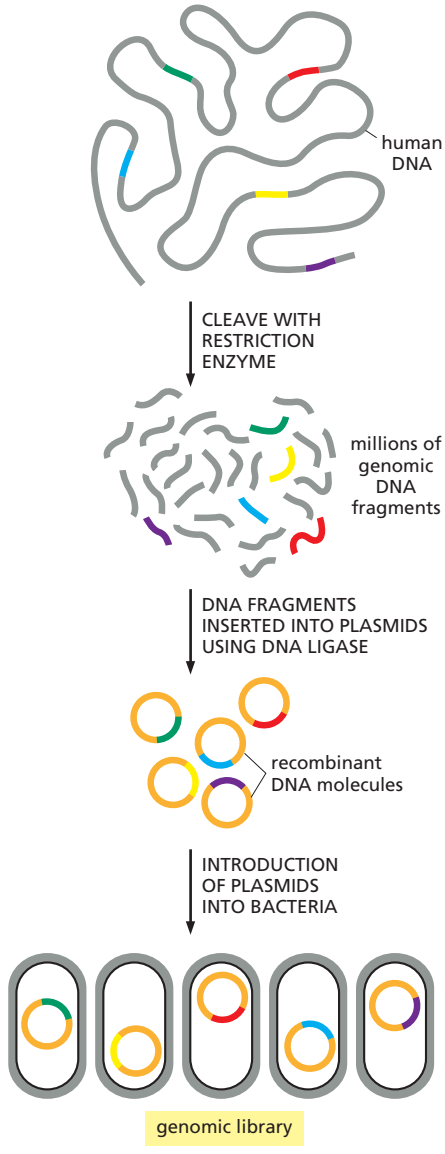


Figure 10–8 Human genomic libraries containing DNA fragments representing the whole human genome can be constructed using restriction enzymes and DNA ligase. Such a genomic library consists of a set of bacteria, each carrying a different small fragment of human DNA. For simplicity, only the *colored* DNA fragments are shown in the library; in reality, all of the different *gray* fragments will also be represented.

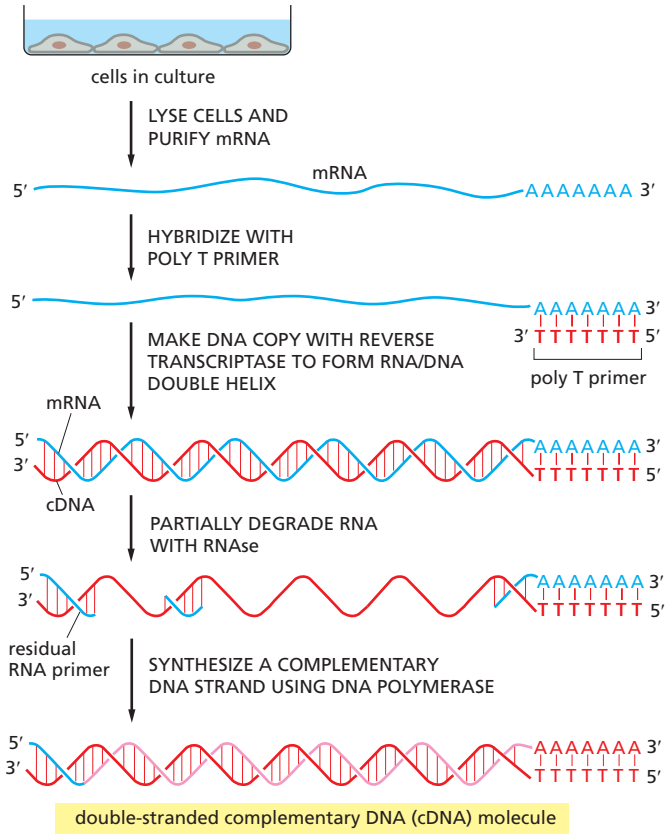


Figure 10–9 Complementary DNA (cDNA) is prepared from mRNA. Total mRNA is extracted from a selected type of cell, and double-stranded complementary DNA (cDNA) is produced using reverse transcriptase (see Figure 9–30) and DNA polymerase. For simplicity, the copying of just one of these mRNAs into cDNA is illustrated here. Following synthesis of the first cDNA strand by reverse transcriptase, treatment with RNAse leaves a few RNA fragments on the cDNA. The RNA fragment that is base-paired to the 3′ end of the first DNA strand acts as the primer for DNA polymerase to synthesize the second, complementary DNA strand. Any remaining RNA is degraded during subsequent cloning steps. As a result, the nucleotide sequences at the extreme 5′ ends of the original mRNA molecules are often absent from cDNA libraries.

There are several important differences between genomic DNA clones and cDNA clones. Genomic clones represent a random sample of all of the DNA sequences found in an organism's genome and, with very rare exceptions, will contain the same sequences regardless of the cell type from which the DNA came. Also, genomic clones from eukaryotes contain large amounts of noncoding DNA, repetitive DNA sequences, introns, regulatory DNA, and spacer DNA; sequences that code for proteins will make up only a few percent of the library (see Figure 9–33). By contrast, cDNA clones contain predominantly protein-coding sequences, and only those sequences that have been transcribed into mRNA in the cells from which the cDNA was made.

As different types of cells produce distinct sets of mRNA molecules, each yields a different cDNA library. Furthermore, patterns of gene expression change during development, so cells at different stages in their development will also yield different cDNA libraries. Thus, cDNAs can be used to assess which genes are expressed in specific cells, at particular times in development, or under a particular set of conditions.

Hybridization Provides a Sensitive Way to Detect Specific Nucleotide Sequences

Thus far, we have been talking about large collections of DNA fragments. For many studies, however, investigators wish to identify or examine an individual gene or RNA. Fortunately, an intrinsic property of nucleic acids—their ability to form complementary base pairs—provides a convenient and powerful technique for detecting a specific nucleotide sequence.

To see how, let's look at a molecule of double-stranded DNA. Under normal conditions, the two strands of a DNA double helix are held together by hydrogen bonds between the complementary base pairs (see Figure 5–4).

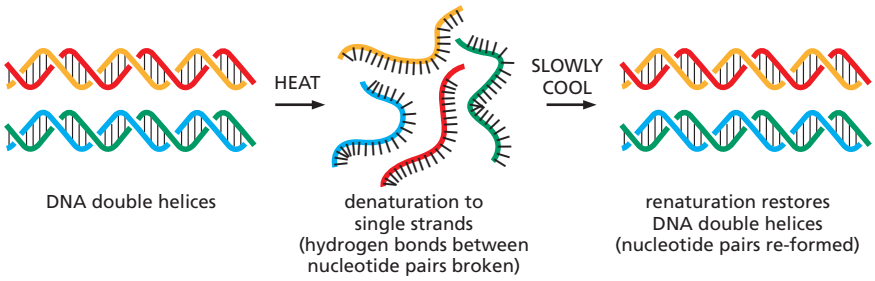


Figure 10–10 A molecule of DNA can undergo denaturation and renaturation (hybridization). For two single-stranded molecules to hybridize, they must have complementary nucleotide sequences that allow base-pairing. In this example, the *red* and *orange* strands are complementary to each other, and the *blue* and *green* strands are complementary to each other. Although denaturation by heating is shown, DNA can also be denatured by alkali treatment. The 1961 discovery that single strands of DNA could readily re-form a double helix in this way was a big surprise to scientists. Hybridization can also occur between complementary strands of DNA and RNA or between two RNAs.

But these relatively weak, noncovalent bonds can be fairly easily broken—for example, by heating the DNA to around 90°C. Such treatment will cause *DNA denaturation*, releasing the two strands from each other. When the conditions are reversed—by slowly lowering the temperature—the complementary strands will readily come back together to re-form a double helix. This *DNA renaturation*, or **hybridization**, is driven by the re-formation of the hydrogen bonds between complementary base pairs (**Figure 10–10**).

Hybridization can be employed for detecting any nucleotide sequence of interest, whether DNA or RNA. One simply designs a short, single-stranded *DNA probe* that is complementary to that sequence. Because the nucleotide sequences of so many genomes are known—and are stored in publicly accessible databases—designing such a probe is straightforward. The desired probe can then be synthesized in the laboratory—usually by a commercial organization or a centralized academic facility.

Hybridization with DNA probes has many uses in cell and molecular biology. As we will see later in this chapter, for example, DNA probes that carry a fluorescent or radioactive label can be used to detect complementary RNA molecules in tissue preparations. But one of the most powerful applications of hybridization is in the cloning of DNA by the polymerase chain reaction, as we discuss next.

QUESTION 10–3

Discuss the following statement: "From the nucleotide sequence of a cDNA clone, the complete amino acid sequence of a protein can be deduced by applying the genetic code. Thus, protein biochemistry has become superfluous because there is nothing more that can be learned by studying the protein."

DNA CLONING BY PCR

Genomic and cDNA libraries were once the only route to gene cloning, and they are still used for cloning very large genes and for sequencing whole genomes. However, a powerful and versatile method for amplifying DNA, known as the **polymerase chain reaction** (**PCR**), provides a more rapid and straightforward approach, particularly in organisms whose complete genome sequence is known. Today, most genes are cloned via PCR.

Invented in the 1980s, PCR revolutionized the way that DNA and RNA are analyzed. The technique can amplify any nucleotide sequence quickly and selectively. Unlike the traditional approach of cloning using vectors—which relies on bacteria to make copies of the desired DNA sequences—PCR is performed entirely in a test tube. Eliminating the need for bacteria makes PCR convenient and fast—billions of copies of a nucleotide sequence can be generated in a matter of hours. At the same time, PCR is remarkably sensitive: the method can be used to amplify and detect the trace amounts of DNA in a drop of blood left at a crime scene or in a few copies of a viral genome in a patient's blood sample. Because of its sensitivity, speed, and ease of use, PCR has many applications in addition to DNA cloning, including forensics and diagnostics.

In this section, we provide a brief overview of how PCR works and how it is used for a range of purposes that require the amplification of specific DNA sequences.

PCR Uses DNA Polymerase and Specific DNA Primers to Amplify DNA Sequences in a Test Tube

The success of PCR depends on the exquisite selectivity of DNA hybridization, along with the ability of DNA polymerase to copy a DNA template reliably, through repeated rounds of replication *in vitro*. The enzyme works by adding nucleotides to the 3′ end of a growing strand of DNA (see Figure 6–11). To initiate the reaction, the polymerase requires a primer—a short nucleotide sequence that provides a 3′ end from which synthesis can begin. The beauty of PCR is that the primers that are added to the reaction mixture not only serve as starting points, but they also direct the polymerase to the specific DNA sequence to be amplified. These primers are designed by the experimenter based on the DNA sequence of interest and then synthesized chemically. Thus, PCR can only be used to clone a DNA segment for which the sequence is known in advance. However, with the large and growing number of genome sequences available in public databases, this requirement is rarely a drawback.

The power of PCR comes from repetition: the cycle of amplification is carried out dozens of times over the course of a few hours. At the start of each cycle, the two strands of the double-stranded DNA template are separated and a unique primer is hybridized, or annealed, to each. DNA polymerase is then allowed to replicate each strand independently (**Figure 10–11**). In subsequent cycles, all the newly synthesized DNA molecules produced by the polymerase serve as templates for the next round of replication (**Figure 10–12**). Through this iterative process of amplification, many copies of the original sequence can be made—billions after about 20 to 30 cycles.

PCR is the method of choice for cloning relatively short DNA fragments (say, under 10,000 nucleotide pairs). Because the original template for PCR can be either DNA or RNA, the method can be used to obtain either a full genomic clone (complete with introns and exons) or a cDNA copy of an mRNA (**Figure 10–13**). A major benefit of PCR is that genes can be cloned directly from any piece of DNA or RNA without the time and effort needed to first construct a DNA library.

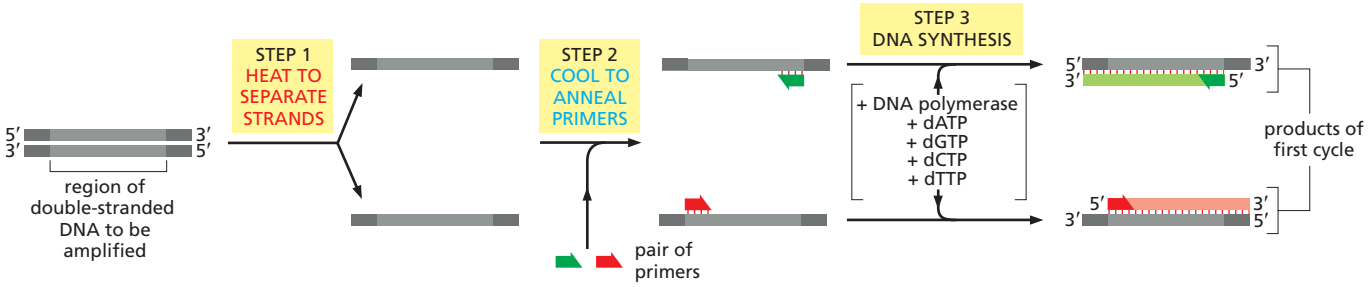


Figure 10–11 A pair of PCR primers directs the amplification of a desired segment of DNA in a test tube. Each cycle of PCR includes three steps: (1) The double-stranded DNA is heated briefly to separate the two strands. (2) The DNA is exposed to a large excess of a pair of specific primers—designed to bracket the region of DNA to be amplified—and the sample is cooled to allow the primers to hybridize to complementary sequences in the two DNA strands. (3) This mixture is incubated with DNA polymerase and the four deoxyribonucleoside triphosphates so that DNA can be synthesized, starting from the two primers. The process can then be repeated by reheating the sample to separate the double-stranded products of the previous cycle (see Figure 10–12).

The technique depends on the use of a special DNA polymerase isolated from a thermophilic bacterium; this polymerase is stable at much higher temperatures than eukaryotic DNA polymerases, so it is not denatured by the heat treatment shown in step 1. The enzyme therefore does not have to be added again after each cycle.

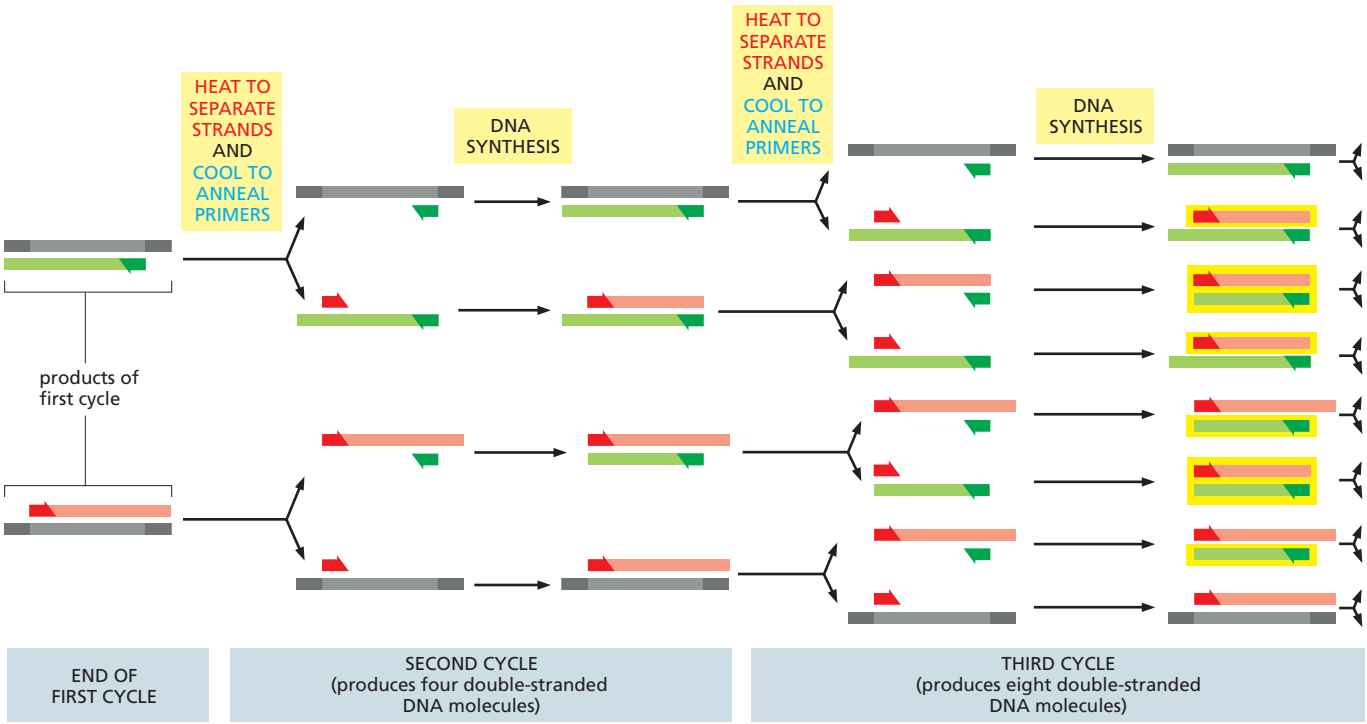


Figure 10–12 PCR uses repeated rounds of strand separation, hybridization, and synthesis to amplify DNA. As the procedure outlined in Figure 10–11 is repeated, all the newly synthesized fragments serve as templates in their turn. Because the polymerase and the primers remain in the sample after the first cycle, PCR involves simply heating and then cooling the same sample, in the same test tube, again and again. Each cycle doubles the amount of DNA synthesized in the previous cycle, so that within a few cycles, the predominant DNA is identical to the sequence bracketed by and including the two primers in the original template. In the example illustrated here, three cycles of reaction produce 16 DNA chains, 8 of which (boxed in *yellow*) correspond exactly to one or the other strand of the original bracketed sequence. After four more cycles, 240 of the 256 DNA chains will correspond exactly to the original sequence, and after several more cycles, essentially all of the DNA strands will be this length. The whole procedure is shown in **Movie 10.1**.

PCR Can Be Used for Diagnostic and Forensic Applications

In addition to its use in cloning, PCR is frequently employed to amplify DNA for other, more practical purposes. Because of its extraordinary sensitivity, PCR can be used to detect an infection at its earliest stages. In this case, short sequences complementary to the suspected pathogen's genome are used as primers, and following many cycles of amplification, even a few copies of an invading bacterial or viral genome in a patient sample can be detected (**Figure 10–14**). PCR can also be used to track epidemics, detect bioterrorist attacks, and test food products for the presence of potentially harmful microbes. It is also used to verify the authenticity of a food source—for example, whether a sample of beef actually came from a cow.

Finally, PCR is widely used in forensic medicine. The method's extreme sensitivity allows forensic investigators to isolate DNA from even the smallest traces of human blood or other tissue to obtain a *DNA fingerprint* of the person who left the sample behind. With the possible exception of identical twins, the genome of each human differs in DNA sequence from that of every other person on Earth. Using primer pairs targeted at genome sequences that are known to be highly variable in the human

QUESTION 10–4

A. If the PCR shown in Figure 10–12 is carried through an additional two rounds of amplification, how many of the DNA fragments (gray, green, red, or outlined in yellow) will be produced? If many additional cycles are carried out, which fragments will predominate?

B. Assume you start with one double-stranded DNA molecule and amplify a 500-nucleotide-pair sequence contained within it. Approximately how many cycles of PCR amplification will you need to produce 100 ng of this DNA? 100 ng is an amount that can be easily detected after staining with a fluorescent dye. (Hint: for this calculation, you need to know that each nucleotide has an average molecular mass of 330 g/mole.)

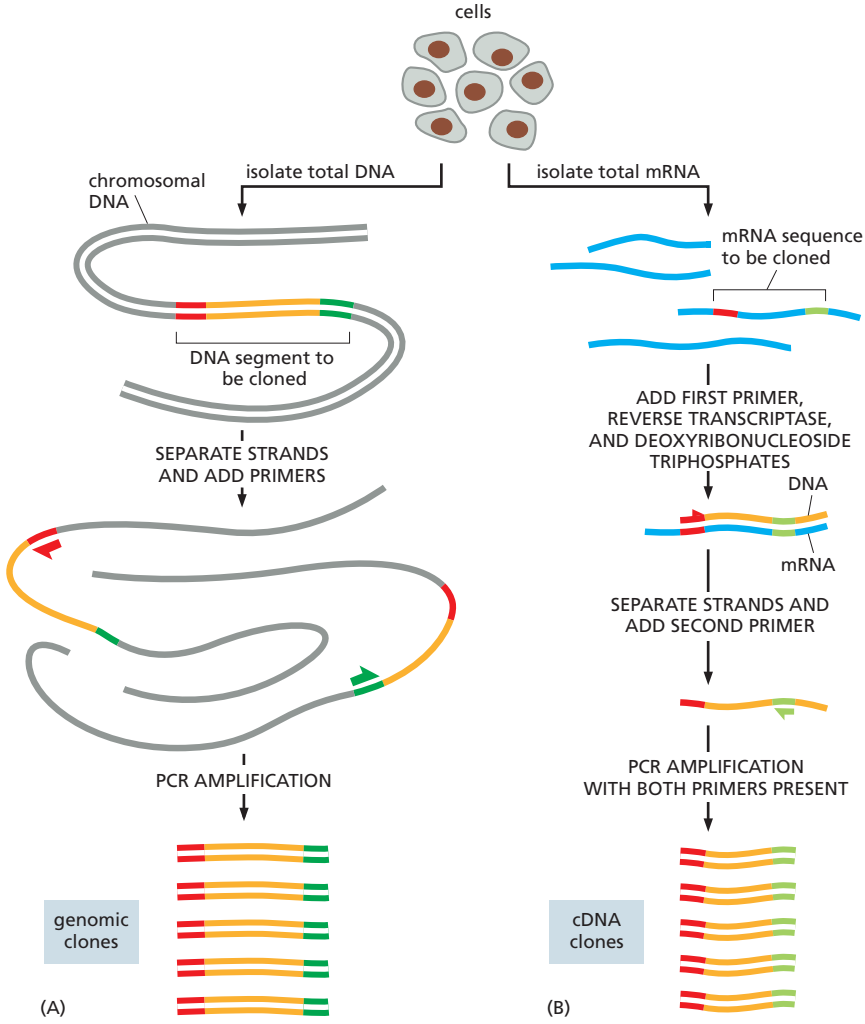


Figure 10–13 PCR can be used to obtain either genomic or cDNA clones. (A) To use PCR to clone a segment of chromosomal DNA, total DNA is first purified from cells. PCR primers that flank the stretch of DNA to be cloned are added, and many cycles of PCR are completed (see Figure 10–12). Because only the DNA between (and including) the primers is amplified, PCR provides a way to obtain selectively any short stretch of chromosomal DNA in an effectively pure form. (B) To use PCR to obtain a cDNA clone of a gene, total mRNA is first purified from cells. The first primer is added to a population of single-stranded mRNAs, and reverse transcriptase is used to make a DNA strand complementary to the specific RNA sequence of interest. A second primer is then added, and the DNA molecule is amplified through many cycles of PCR.

population, PCR makes it possible to generate a distinctive DNA fingerprint for any individual (**Figure 10–15**). Such forensic analyses can be used not only to point the finger at those who have done wrong, but—equally important—to help exonerate those who have been wrongfully convicted.

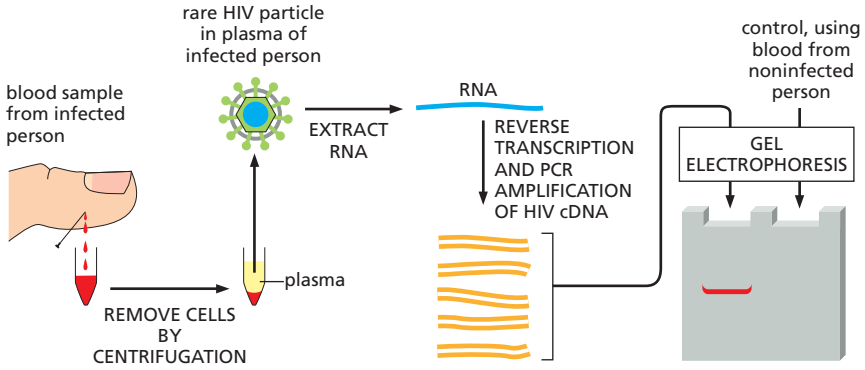


Figure 10–14 PCR can be used to detect the presence of a viral genome in a sample of blood. Because of its ability to amplify enormously the signal from every single molecule of nucleic acid, PCR is an extraordinarily sensitive method for detecting trace amounts of virus in a sample of blood or tissue without the need to purify the virus. For HIV, the virus that causes AIDS, the genome is a single-stranded molecule of RNA, as illustrated here. In addition to HIV, many other viruses that infect humans are now detected in this way.

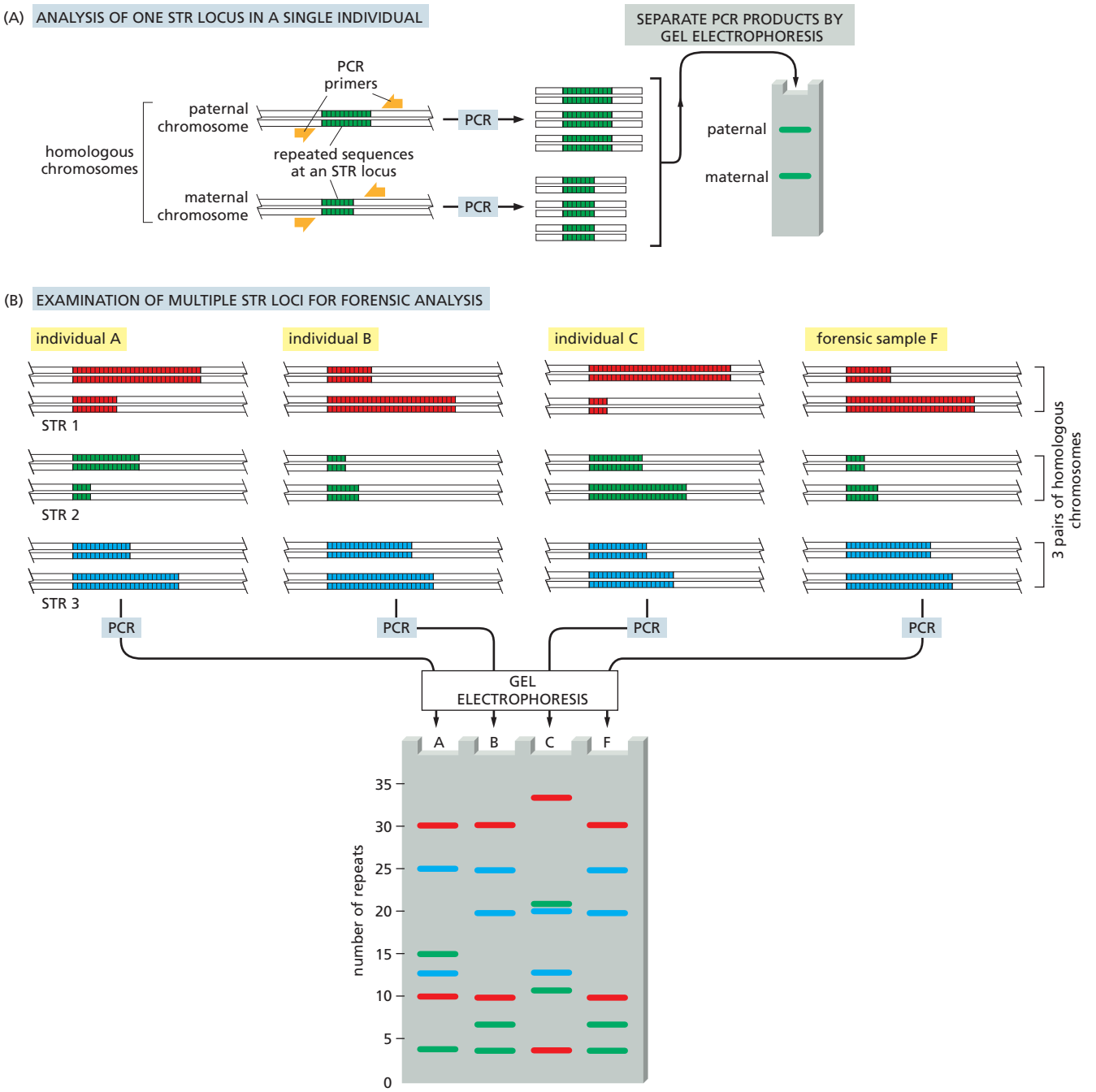


Figure 10–15 PCR is used in forensic science to distinguish one individual from another. The DNA sequences typically analyzed are short tandem repeats (STRs). These sequences, composed of stretches of CACA… or GTGT…, for example, are found in various positions (loci) in the human genome. The number of repeats in each STR locus is highly variable in the population, ranging from 4 to 40 in different individuals. Because of the variability in these sequences, individuals will usually inherit a different number of repeats at each STR locus from their mother and from their father; two unrelated individuals, therefore, rarely contain the same pair of repeat sequences at a given STR locus. (A) PCR using primers that recognize unique sequences on either side of one particular STR locus produces a pair of bands of amplified DNA from each individual, one band representing the maternal STR variant and the other representing the paternal STR variant. The length of the amplified DNA, and thus its position after gel electrophoresis, will depend on the exact number of repeats at the locus. (B) In the schematic example shown here, the same three STR loci are analyzed in samples from three suspects (individuals A, B, and C), producing six bands for each individual. Although different people can have several bands in common, the overall pattern is quite distinctive for each person. The band pattern can therefore serve as a *DNA fingerprint* to identify an individual nearly uniquely. The fourth lane in the gel (lane F) contains the products of the same PCR amplifications carried out on a hypothetical forensic DNA sample, which could have been obtained from a single hair or a tiny spot of blood left at a crime scene.

The more loci that are examined, the more confidence we can have about the results. When examining the variability at 5–10 different STR loci, the odds that two random individuals would share the same fingerprint by chance are approximately one in 10 billion. In the case shown here, individuals A and C can be eliminated from inquiries, while B is a clear suspect. A similar approach is now used routinely in paternity testing.

SEQUENCING DNA

Because information is encoded in the linear sequence of nucleotides in an organism's genome, the key to understanding the function and regulation of genes and genomes lies in the sequence of the DNA. Nucleotide sequences can reveal clues to the evolutionary relationships among different organisms, and provide insights into the causes of human disease. Knowing the sequence of a gene is a prerequisite for cloning that gene by PCR, and it allows large-scale production of any protein a gene might encode.

Because sequence information is so valuable, a great deal of effort has been dedicated over the past few decades to the development of DNA sequencing technologies with greater speed and sensitivity. As a result, we now have a variety of sophisticated and powerful methods that make it possible to obtain the complete nucleotide sequence of a genome in a fraction of the time, and at a fraction of the cost, required even 10 years ago.

In this section, we briefly describe the principles underlying the major DNA sequencing methods used today, and we provide a glimpse of some new sequencing technologies that are just around the corner.

Dideoxy Sequencing Depends on the Analysis of DNA Chains Terminated at Every Position

In the late 1970s, researchers developed several schemes for determining, simply and quickly, the nucleotide sequence of any purified DNA fragment. The method that became the most widely used—and continues to be employed in some applications today—is called **dideoxy sequencing** or **Sanger sequencing** (after the scientist who invented it). This technique uses DNA polymerase, along with special chain-terminating nucleotides called dideoxyribonucleoside triphosphates (**Figure 10–16**), to make partial copies of the DNA fragment to be sequenced. Dideoxy sequencing reactions ultimately produce a collection of different DNA copies that terminate at every position in the original DNA sequence.

Although the original method could be quite laborious—particularly reading the nucleotide sequences from the bands on a sequencing gel—the procedure is now fully automated: robotic devices mix the reagents—including the four different chain-terminating dideoxynucleotides, each tagged with a different-colored fluorescent dye—and load the reaction samples onto long, thin capillary gels, which separate the reaction products into a series of distinct bands. A detector then records the color of each band, and a computer translates the information into a nucleotide sequence (**Figure 10–17**).

The automated dideoxy method made it possible to sequence the first genomes of humans and of many other organisms, including most of those discussed in this book. How such sequence information was analyzed to assemble a complete genome sequence—for example, the initial draft of the human genome—is described in **How We Know**, pp. 348–349.

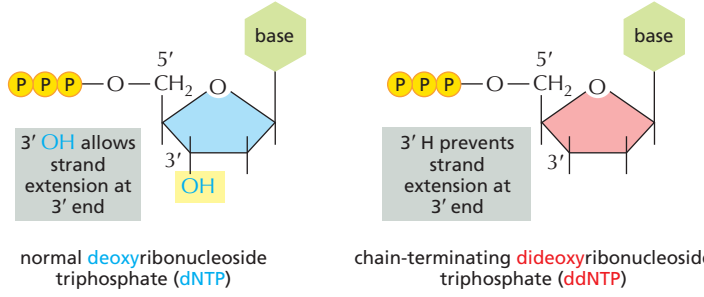


Figure 10–16 The dideoxy method of sequencing DNA relies on chain-terminating dideoxynucleoside triphosphates (ddNTPs). These ddNTPs are derivatives of the normal deoxyribonucleoside triphosphates that lack the 3′ hydroxyl group. When incorporated into a growing DNA strand, they block further elongation of that strand.

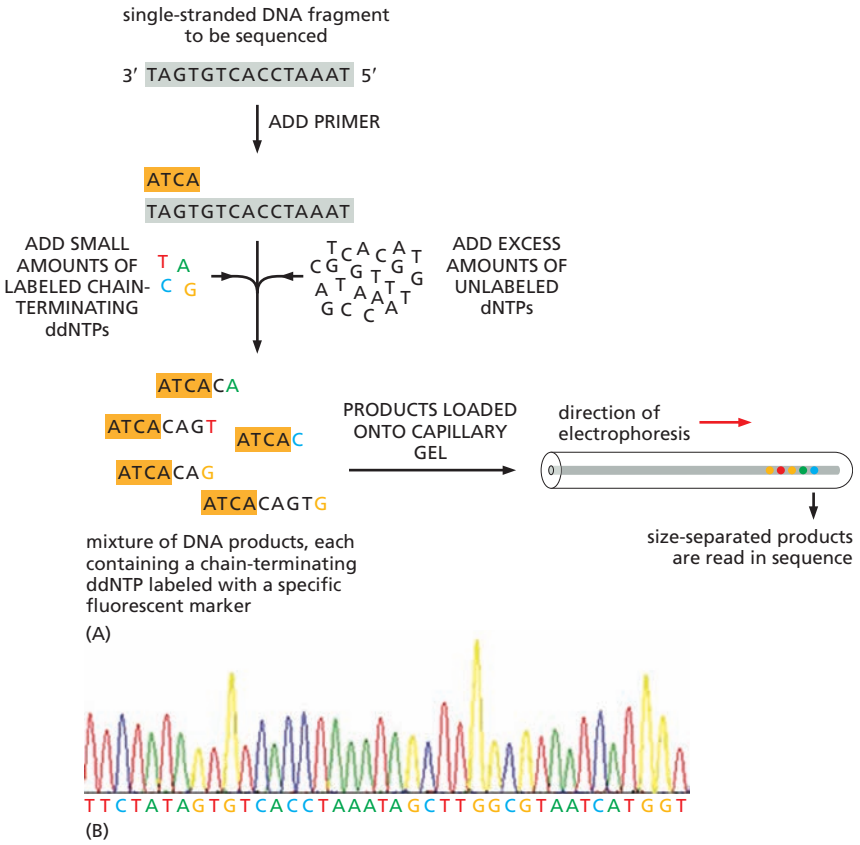


Figure 10–17 Automated dideoxy sequencing relies on a set of four ddNTPs, each bearing a uniquely colored fluorescent tag. (A) To determine the complete sequence of a single-stranded fragment of DNA (*gray*), the DNA is first hybridized with a short DNA primer (*orange*). The DNA is then mixed with DNA polymerase (not shown), an excess amount of normal dNTPs, and a mixture containing small amounts of all four chain-terminating ddNTPs, each of which is labeled with a fluorescent tag of a different color. Because the chain-terminating ddNTPs will be incorporated only occasionally, each reaction produces a diverse set of DNA copies that terminate at different points in the sequence. The reaction products are loaded onto a long, thin capillary gel and separated by electrophoresis. A camera reads the color of each band on the gel and feeds the data to a computer that assembles the sequence (not shown). The sequence read from the gel will be complementary to the sequence of the original DNA molecule. (B) A tiny part of the data from such an automated sequencing run. Each colored peak represents a nucleotide in the DNA sequence.

Next-Generation Sequencing Techniques Make Genome Sequencing Faster and Cheaper

Newer methods for the determination of nucleotide sequence, developed over the past decade or so, have made genome sequencing much more rapid—and much cheaper. As the cost of sequencing DNA has plummeted, the number of genomes that have been sequenced has skyrocketed. These rapid methods allow multiple genomes to be sequenced in parallel in a matter of weeks. With these techniques—collectively referred to as *second-generation sequencing methods*—investigators have been able to examine thousands of human genomes, catalog the variation in nucleotide sequences from people around the world, and uncover the mutations that increase the risk of various diseases—from cancer to autism—as we discuss in Chapter 19.

Although each method differs in detail, many rely on the sequencing of libraries of DNA fragments that, taken together, represent the DNA of the entire genome. Instead of using bacterial cells to generate these libraries (as seen in Figure 10–8), however, the libraries are synthesized by PCR amplification of a collection of DNA fragments, each attached to a solid support such as a glass slide or bead. The resulting PCR-generated copies, instead of drifting away in solution, remain bound in proximity to their original "parent" DNA fragment. The process thus generates DNA clusters, each containing about 1000 identical copies of a single DNA fragment. All of these clusters are then sequenced at the same time. One of the most common methods for doing so is called *Illumina sequencing*. Like automated dideoxy sequencing, Illumina sequencing is based on the use of chain-terminating nucleotides with uniquely colored fluorescent tags. In the Illumina method, however, the fluorescent tags and the chemical group that blocks elongation are removable. Once DNA

SEQUENCING THE HUMAN GENOME

When DNA sequencing techniques became fully automated, determining the order of the nucleotides in a piece of DNA went from being an elaborate Ph.D. thesis project to a routine laboratory chore. Feed DNA into the sequencing machine, add the necessary reagents, and out comes the sought-after result: the order of As, Ts, Gs, and Cs. Nothing could be simpler.

So why was sequencing the human genome such a formidable task? Largely because of its size. The DNA sequencing methods employed at the time were limited by the physical size of the gel used to separate the labeled fragments (see, for example, Figure Q10–9). At most, only a few hundred nucleotides could be read from a single gel. How, then, do you handle a genome that contains billions of nucleotide pairs?

The solution is to break the genome into fragments and sequence these smaller pieces. The main challenge then comes in piecing the short fragments together in the correct order to yield a comprehensive sequence of a whole chromosome, and ultimately a whole genome. There are two main strategies for accomplishing this genomic breakage and reassembly: the shotgun method and the clone-by-clone approach.

Shotgun sequencing

The most straightforward approach to sequencing a genome is to break it into random fragments, separate and sequence each of the single-stranded fragments, and then use a powerful computer to order these pieces using sequence overlaps to guide the assembly (**Figure 10–18**). This approach is called the shotgun sequencing strategy. As an analogy, imagine shredding several copies of *Essential Cell Biology* (*ECB*), mixing up the pieces, and then trying to put one whole copy of the book back together again by matching up the words or phrases or sentences that appear on each piece. (Several copies would be needed to generate enough overlap for reassembly.) It could be done, but it would be much easier if the book were, say, only two pages long.

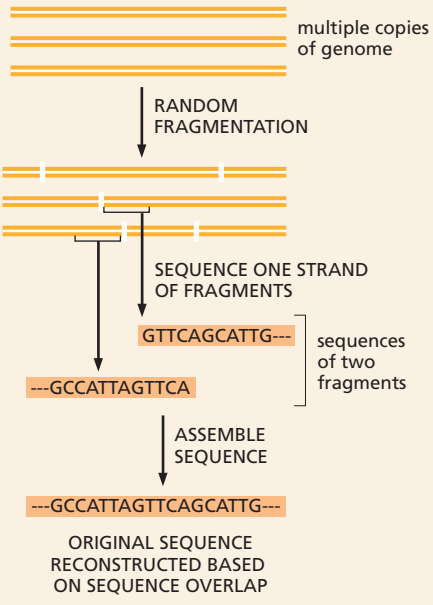


Figure 10–18 Shotgun sequencing is the method of choice for small genomes. The genome is first broken into much smaller, overlapping fragments. Each fragment is then sequenced, and the genome is assembled based on overlapping sequences.

For this reason, a straight-out shotgun approach is the strategy of choice only for sequencing small genomes. The method proved its worth in 1995, when it was used to sequence the genome of the infectious bacterium *Haemophilus influenzae*, the first organism to have its complete genome sequence determined. The trouble with shotgun sequencing is that the reassembly process can be derailed by repetitive nucleotide sequences. Although rare in bacteria, these sequences make up a large fraction of vertebrate genomes (see Figure 9–33). Highly repetitive DNA segments make it difficult to piece DNA sequences back together accurately (**Figure 10–19**). Returning to the *ECB* analogy, this chapter alone contains more than a few instances of the phrase "the human genome." Imagine that one slip of paper from the shredded *ECB*s contains the information: "So why was sequencing the human genome" (which appears at the start of this section); another contains the information: "the human genome sequence consortium combined shotgun sequencing with a clone-by-clone approach" (which appears below). You might be tempted to join these two segments together based on the overlapping phrase "the human genome." But you would wind up with the nonsensical statement: "So why was sequencing the human genome sequence consortium combined shotgun sequencing with a clone-by-clone approach." You would also lose the several paragraphs of important text that originally appeared between these two instances of "the human genome."

And that's just in this section. The phrase "the human genome" appears in many chapters of this book. Such repetition compounds the problem of placing each fragment in its correct context. To circumvent these assembly problems, researchers in the human genome sequence consortium combined shotgun sequencing with a clone-by-clone approach.

Clone-by-clone

In this approach, researchers started by preparing a genomic DNA library. They broke the human genome into overlapping fragments, 100–200 kilobase pairs in size. They then plugged these segments into bacterial artificial chromosomes (BACs) and inserted them into *E. coli*. (BACs are similar to the bacterial plasmids discussed earlier, except they can carry much larger pieces of DNA.) As the bacteria divided, they copied the BACs, thus producing a collection of overlapping cloned fragments (see Figure 10–8).

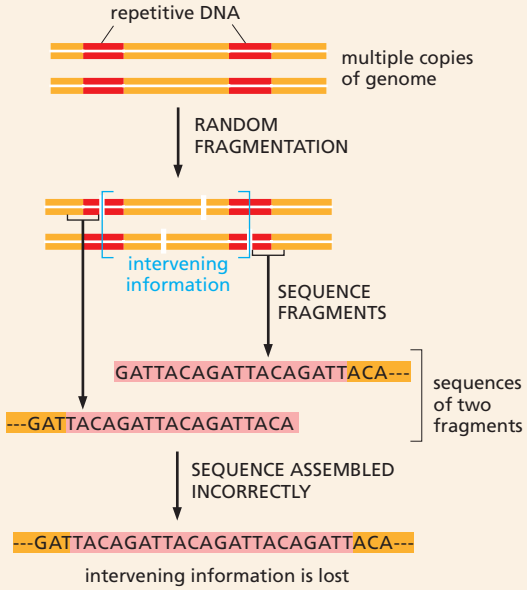


Figure 10–19 Repetitive DNA sequences in a genome make it difficult to accurately assemble its fragments. In this example, the DNA contains two segments of repetitive DNA, each made of many copies of the sequence GATTACA. When the resulting sequences are examined, two fragments from different parts of the DNA appear to overlap. Assembling these sequences incorrectly would result in a loss of the information (in brackets) that lies between the original repeats.

The researchers then determined where each of these DNA fragments fit into the existing map of the human genome. To do this, different restriction enzymes were used to cut each clone to generate a unique restriction-site "signature." The locations of the restriction sites in each fragment allowed researchers to map each BAC clone onto a restriction map of a whole human genome that had been generated previously using the same set of restriction enzymes (**Figure 10–20**).

Knowing the relative positions of the cloned fragments, the researchers then selected some 30,000 BACs, sheared each into smaller fragments, and determined the nucleotide sequence of each BAC separately using the shotgun method. They could then assemble the whole genome sequence by stitching together the sequences of thousands of individual BACs that span the length of the genome.

The beauty of this approach was that it was relatively easy to accurately determine where the BAC fragments belong in the genome. This mapping step reduced the likelihood that regions containing repetitive sequences were assembled incorrectly, and it virtually eliminated the possibility that sequences from different chromosomes were mistakenly joined together. Returning to the textbook analogy, the BAC-based approach is akin to first separating your copies of *ECB* into individual pages and then shredding each page into its own separate pile. It should be much easier to put the book back together when one pile of fragments contains words from page 1, a second pile from page 2, and so on. And there's virtually no chance of mistakenly sticking a sentence from page 40 into the middle of a paragraph on page 412.

All together now

The clone-by-clone approach produced the first draft of the human genome sequence in 2000 and the completed sequence in 2004. As the set of instructions that specify all of the RNA and protein molecules needed to build a human being, this string of genetic bits holds the secrets to human development and physiology. But the sequence was also of great value to researchers interested in comparative genomics or in the physiology of other organisms: it eased the assembly of nucleotide sequences from other mammalian genomes—mice, rats, dogs, and other primates. It also made it much easier to determine the nucleotide sequences of the genomes of individual humans by providing a framework on which the new sequences could be simply superimposed.

The first human sequence was the only mammalian genome completed in this methodical way. But the Human Genome Project was an unqualified success in that it provided the techniques, confidence, and momentum that drove the development of the next generation of DNA sequencing methods, which are now rapidly transforming all areas of biology.

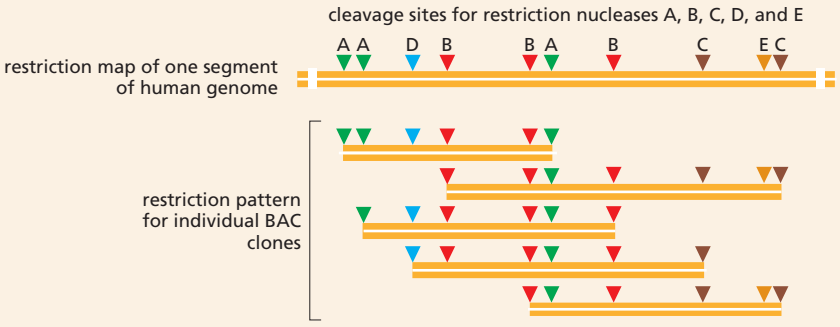


Figure 10–20 Individual BAC clones are positioned on the physical map of the human genome sequence on the basis of their restriction-site "signatures." Clones are digested with five different restriction enzymes, and the sites at which the different enzymes cut each clone are recorded. The distinctive pattern of restriction sites allows investigators to order the fragments and place them on a restriction map of a human genome that had been previously generated using the same nucleases.

polymerase has added the labeled, chain-terminating nucleotide, a photo of the slide is taken and the identity of the nucleotide added at each cluster is recorded; the label and the chain-terminator are then stripped away, allowing DNA polymerase to add the next nucleotide (**Figure 10–21**).

More recent technological advances have led to the development of *third-generation sequencing methods* that permit the sequencing of just a single molecule of DNA. One of these techniques, called Single Molecule Real Time sequencing, employs a special apparatus in which a single DNA polymerase and a DNA template with an attached primer are anchored together in a tiny compartment with differently colored fluorescent dNTPs. As DNA synthesis proceeds, the attachment of each nucleotide to the growing DNA strand is determined one base at a time, revealing the sequence of the template; as in other sequencing methods, large numbers of reactions are measured in parallel in separate compartments. In another method, still under development, a single DNA molecule is pulled slowly through a tiny channel, like thread through the eye of a needle. Because each of the four nucleotides has different, characteristic chemical properties, the way a nucleotide obstructs the pore as it passes through reveals its identity—information that is then used to compile the sequence of the DNA molecule. Further refinement of these and other technologies will continue to drive down the amount of time and money required to sequence a human genome.

Comparative Genome Analyses Can Identify Genes and Predict Their Function

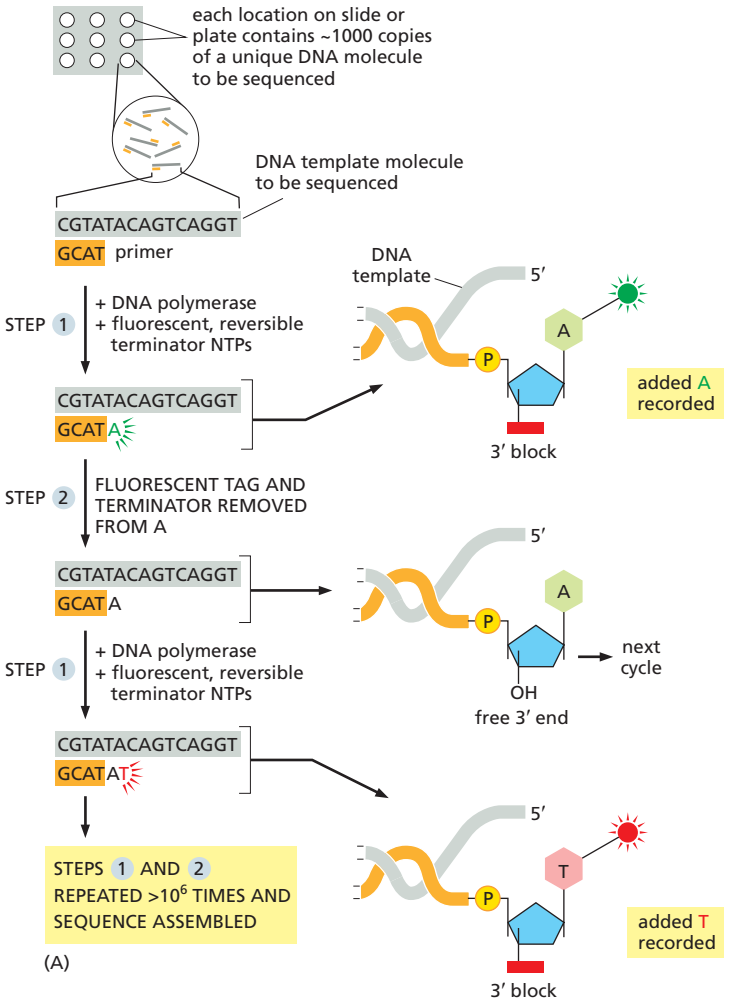
Strings of nucleotides, at first glance, reveal nothing about how that genetic information directs the development of a living organism—or even what type of organism it might encode. One way to learn something about the function of a particular nucleotide sequence is to compare it with the multitude of sequences available in public databases. Using a computer program to search for sequence similarity, one can determine whether a nucleotide sequence contains a gene and what that gene is likely to do—based on the gene's known activity in other organisms.

Comparative analyses have revealed that the coding regions of genes from a wide variety of organisms show a large degree of sequence conservation (see Figure 9–20). The sequences of noncoding regions, however, tend to diverge rapidly over evolutionary time (see Figure 9–19). Thus, a search for sequence similarity can often indicate from which organism a particular piece of DNA was derived, and which species are most closely related. Such information is particularly useful when the origin of a DNA sample is unknown—because it was extracted, for example, from a sample of soil or seawater or the blood of a patient with an undiagnosed infection.

EXPLORING GENE FUNCTION

Knowing where a nucleotide sequence comes from—or even what activity it might have—is only the first step toward determining what role it has in the development or physiology of an organism. The knowledge that a particular DNA sequence encodes a transcription regulator, for example, does not reveal when and where that protein is produced, or which genes it might regulate. To learn that, investigators must head back to the laboratory.

This is where creativity comes in. There are as many ways to study how genes function as there are scientists with an interest in studying the question. The techniques an investigator chooses often depend on his or



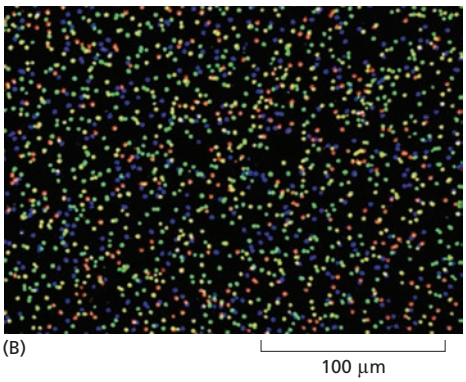


Figure 10–21 Illumina sequencing is based on the basic principles of automated dideoxy sequencing. (A) A genome or other large DNA sample of interest is broken into millions of short fragments. These fragments are attached to a glass surface and amplified by PCR to generate DNA clusters, each containing about a thousand copies of a single DNA fragment. The large number of clusters provides complete coverage of the genome. In the first step, the anchored DNA clusters are incubated with DNA polymerase and a special set of four nucleoside triphosphates (NTPs) with two reversible chemical modifications: a uniquely colored fluorescent marker and a 3′ chemical group that terminates DNA synthesis. No normal dNTPs are present in the reaction. After a nucleotide is added by DNA polymerase, a high-resolution digital camera records the color of the fluorescence at each DNA cluster. In the second step, the DNA is chemically treated to remove the fluorescent markers and chemical blockers. A new batch of fluorescent, reversible terminator NTPs is then added to initiate another round of DNA synthesis. These steps are repeated until the sequence is complete. The snapshots of each round of synthesis are compiled by computer to yield the sequence of each DNA fragment. The sequence of the millions of overlapping DNA fragments can then be used to reconstruct the complete genome sequence. (B) An image of a glass slide showing individual DNA clusters after a round of DNA synthesis with colored NTPs. (B, courtesy of Illumina, Inc.)

her background and training: a geneticist might, for example, engineer mutant organisms in which the activity of the gene has been disrupted, whereas a biochemist might take the same gene and produce large amounts of its protein to determine its three-dimensional structure.

In this section, we present a few of the approaches that investigators currently use to study gene function. We explore a variety of techniques for investigating when and where a gene is expressed. We then describe how disrupting the activity of a gene in a cell, tissue, or whole plant or animal can provide insights into what that gene normally does. Finally, we explain how proteins can be produced in large amounts for biochemical and structural studies.

Analysis of mRNAs Provides a Snapshot of Gene Expression

As we discuss in Chapter 8, a cell expresses only a subset of the thousands of genes available in its genome. This subset of genes differs from one cell type to another, and under different conditions in the same cell type. One way to determine which genes are being expressed in a population of cells or in a tissue is to analyze which mRNAs are being produced.

To sequence all the RNAs produced by a cell, investigators make use of the next-generation sequencing technologies described earlier. In most cases, a collection of RNAs is converted into complementary DNA (cDNA) by reverse transcriptase, and these cDNAs are then sequenced. This

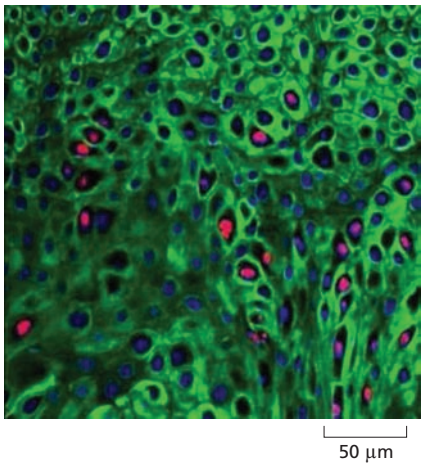


Figure 10–22 *In situ* hybridization can be used to detect the presence of a virus in cells. In this micrograph, the nuclei of cultured epithelial cells infected with the human papillomavirus (HPV) are stained *pink* by a fluorescent probe that recognizes a viral DNA sequence. The cytoplasm of all cells is stained *green*. (Courtesy of Hogne Røed Nilsen.)

method, called **RNA-Seq** or deep RNA sequencing, provides a quantitative analysis of the *transcriptome*—the complete collection of RNAs produced by a cell under a certain set of conditions. It also reveals the number of times a particular sequence appears in a sample and can detect rare mRNAs, RNA transcripts that are alternatively spliced, mRNAs that harbor sequence variations, and noncoding RNAs. This remarkably powerful technology has led to dramatic new insights into the genes expressed in a variety of cells and tissues at different times in development, during different stages of the cell-division cycle, in response to treatment with different drugs, or as a result of different mutations.

In Situ Hybridization Can Reveal When and Where a Gene Is Expressed

Although RNA-Seq can provide a list of genes that are being expressed by a particular tissue at a particular time, it does not reveal exactly where in the tissue those RNAs are produced. To do that, investigators use a technique called ***in situ* hybridization** (from the Latin *in situ*, "in place"), which allows a specific nucleic acid sequence—either DNA or RNA—to be visualized in its normal location.

In situ hybridization uses single-stranded DNA or RNA probes, labeled with either fluorescent dyes or radioactive isotopes, to detect complementary nucleic acid sequences within a tissue (**Figure 10–22**) or even on an isolated chromosome (**Figure 10–23**). The latter application is used in the clinic to determine, for example, whether fetuses carry abnormal chromosomes. *In situ* hybridization is also used to study the expression patterns of a particular gene or collection of genes in an adult or developing tissue, providing important clues about when and where these genes carry out their functions.

Reporter Genes Allow Specific Proteins to Be Tracked in Living Cells

For a gene that encodes a protein, the location of the protein within the cell, tissue, or organism yields clues to the gene's function. Traditionally, the most effective way to visualize a protein within a cell or tissue involved using a labeled antibody. That approach requires the generation of an antibody that specifically recognizes the protein of interest—a process that can be time-consuming and offers no guarantee of success.

An alternative approach is to use the regulatory DNA sequences of the protein-coding gene to drive the expression of some type of **reporter gene**, which encodes a protein that can be easily monitored by its fluorescence or enzymatic activity. A recombinant gene of this type usually mimics the expression of the gene of interest, producing the reporter

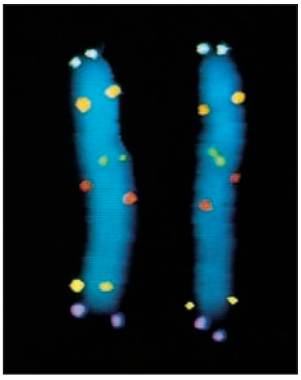


Figure 10–23 *In situ* hybridization can be used to locate genes on isolated chromosomes. Here, six different DNA probes have been used to mark the locations of their respective nucleotide sequences on human Chromosome 5 isolated from a mitotic cell in metaphase (see Figure 5–15 and Panel 18–1, pp. 628–629). The DNA probes have been labeled with different chemical groups and are detected using fluorescent antibodies specific for those groups. Both the maternal and paternal copies of Chromosome 5 are shown, aligned side-by-side. Each probe produces two dots on each chromosome because chromosomes undergoing mitosis have already replicated their DNA; therefore, each chromosome contains two identical DNA helices. The technique employed here is nicknamed FISH, for fluorescence *in situ* hybridization. (Courtesy of David C. Ward.)

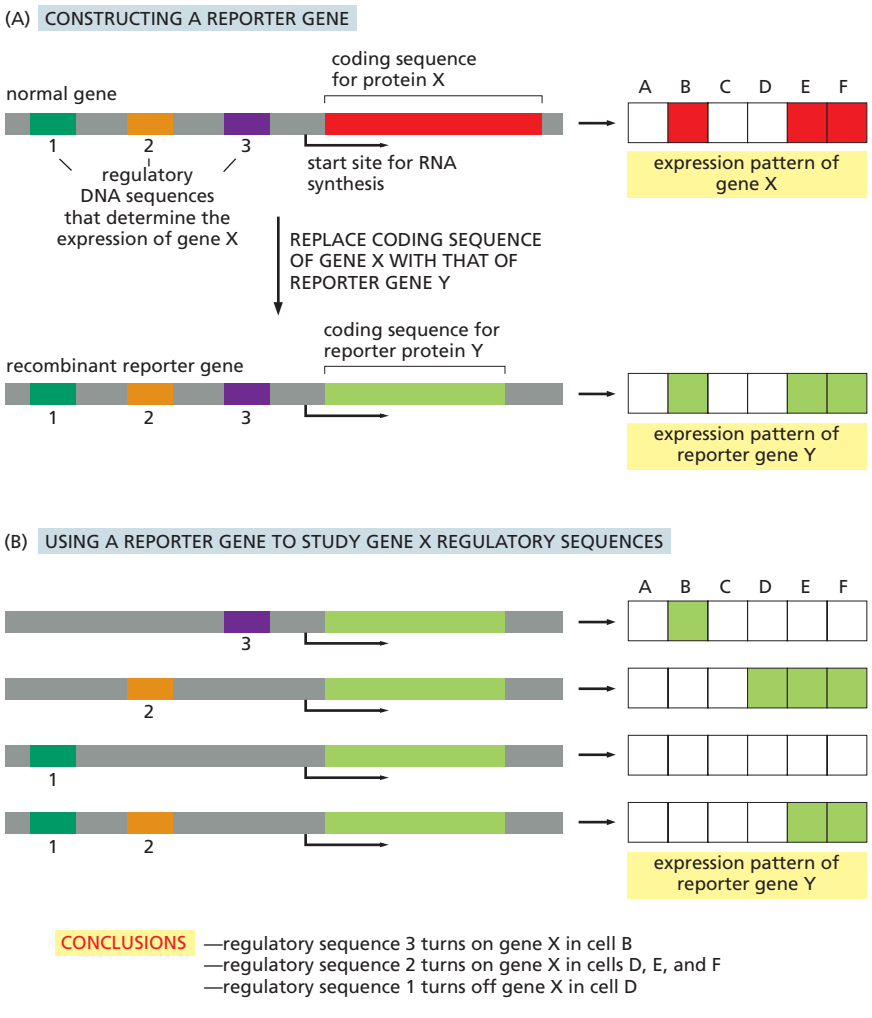


Figure 10–24 Reporter genes can be used to determine the pattern of a gene's expression. (A) Suppose the goal is to find out which cell types (A–F) express protein X, but it is difficult to detect the protein directly—with antibodies, for example. Using recombinant DNA techniques, the coding sequence for protein X can be replaced with the coding sequence for reporter protein Y, which can be easily monitored visually; two commonly used reporter proteins are the enzyme β-galactosidase (see Figure 8–14C) and green fluorescent protein (GFP, see Figure 10–25). The expression of the reporter protein Y will now be controlled by the regulatory sequences (here labeled 1, 2, and 3) that control the expression of the normal protein X. (B) To determine which regulatory sequences normally control expression of gene X in particular cell types, reporters with various combinations of the regulatory regions associated with gene X can be constructed. These recombinant DNA molecules are then tested for expression after their introduction into the different cell types.

protein when, where, and in the same amounts as the normal protein would be made (**Figure 10–24A**). This approach can also be used to study the regulatory DNA sequences that control the gene's expression (**Figure 10–24B**).

One of the most popular reporter proteins is **green fluorescent protein** (**GFP**), the molecule that gives luminescent jellyfish their greenish glow. If the gene that encodes GFP is fused to the regulatory sequences of a gene of interest, the expression of the resulting reporter gene can be monitored by fluorescence microscopy (**Figure 10–25**). The use of multiple GFP variants that fluoresce at different wavelengths can provide insights into how different cells interact in a living tissue (**Figure 10–26**).

In some cases, the DNA encoding GFP is attached directly to the protein-coding region of the gene of interest, resulting in a GFP fusion protein

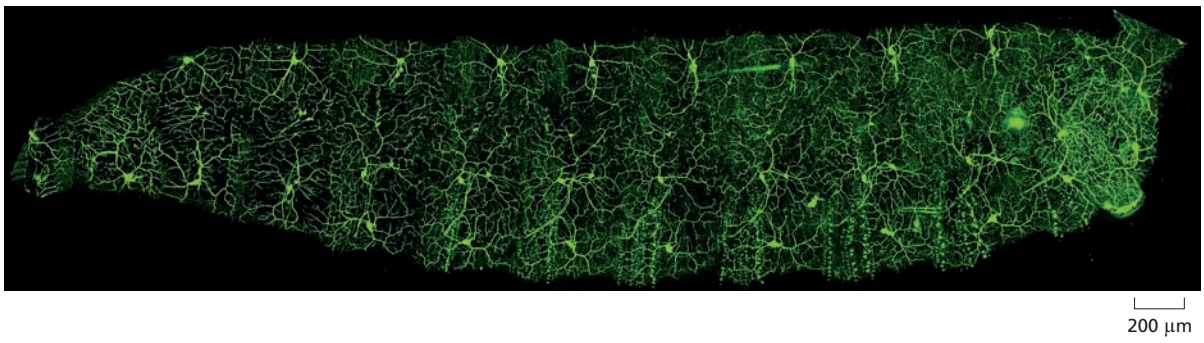


Figure 10–25 Green fluorescent protein (GFP) can be used to identify specific cells in a living animal. For this experiment, carried out in the fruit fly, recombinant DNA techniques were used to join the gene encoding GFP to the regulatory DNA sequences that direct the production of a particular *Drosophila* protein. Both the GFP and the normal fly protein are made only in a specialized set of neurons. This image of a live fly larva was captured by a fluorescence microscope and shows approximately 20 neurons, each with long extensions (axons and dendrites) that communicate with other (nonfluorescent) cells. These neurons, located just under the body surface, allow the organism to sense its immediate environment. (Courtesy of Samantha Galindo/Grueber Lab/Columbia University's Zuckerman Institute.)

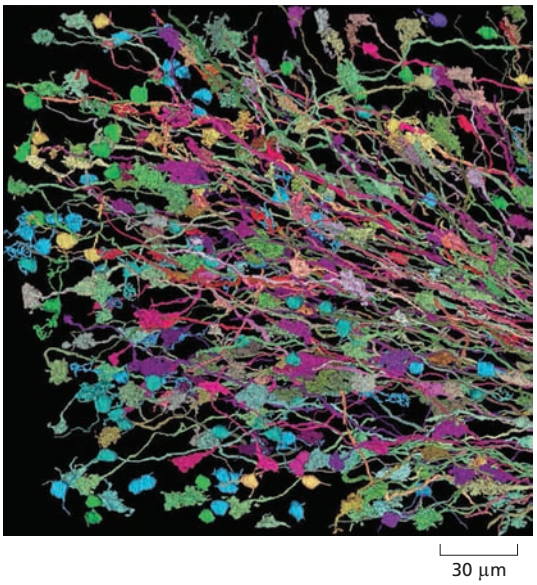


Figure 10–26 GFPs that fluoresce at different wavelengths help reveal the connections that individual neurons make within the brain. This image shows differently colored neurons in one region of a mouse brain. The neurons express different combinations of differently colored GFPs, making it possible to distinguish and trace many individual neurons within a population. The stunning appearance of these labeled neurons earned the animals that bear them the colorful nickname "brainbow mice." (From J. Livet et al., *Nature* 450:56–62, 2007. With permission from Macmillan Publishers Ltd.)

that often behaves in the same way as the normal protein produced by the gene. GFP fusion has become a standard strategy for tracking not only the location but also the movement of specific proteins in living cells (see How We Know, pp. 520–521).

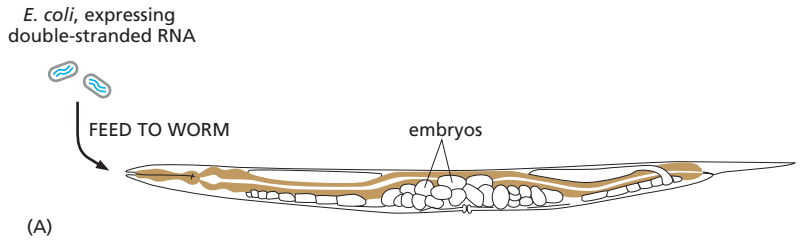
The Study of Mutants Can Help Reveal the Function of a Gene

Although it may seem counterintuitive, one of the best ways to determine a gene's function is to see what happens to an organism when the gene is inactivated by a mutation. Before the advent of gene cloning, geneticists would often study the mutant organisms that arise at random in a population. The mutants of most interest were often selected because of their unusual *phenotype*—fruit flies with white eyes or curly wings, for example. The gene responsible for the mutant phenotype could then be studied by breeding experiments, as Gregor Mendel did with peas in the nineteenth century (discussed in Chapter 19).

Although mutant organisms can arise spontaneously, they do so infrequently. The process can be accelerated by treating organisms with radiation or chemical mutagens, which randomly disrupt gene activity. Such random mutagenesis generates large numbers of mutant organisms, each of which can then be studied individually. This "classical genetic approach," which we discuss in detail in Chapter 19, is most applicable to organisms that reproduce rapidly and can be analyzed genetically in the laboratory—such as bacteria, yeasts, nematode worms, and fruit flies—although it has also been used to study zebrafish and mice, which require more time to reproduce and develop.

RNA Interference (RNAi) Inhibits the Activity of Specific Genes

DNA technology has made possible more targeted genetic approaches to studying gene function. Instead of beginning with a randomly generated mutant and then identifying the responsible gene, a gene of known sequence can be inactivated deliberately, and the effects on the cell or organism's phenotype can be observed. Because this strategy is essentially the reverse of that used in classical genetics—which goes from mutants to genes—it is often referred to as *reverse genetics*.



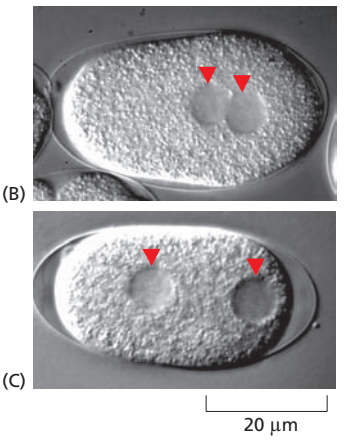


Figure 10–27 Gene function can be tested by RNA interference. (A) Double-stranded RNA (dsRNA) can be introduced into *C. elegans* by feeding the worms *E. coli* that express the dsRNA. Gene function is reduced in all tissues, including the reproductive tissues where embryos are produced by self-fertilization. (B) In a wild-type worm embryo, the egg and sperm pronuclei (*red* arrowheads) come together in the posterior half of the embryo shortly after fertilization. (C) In an embryo in which a particular gene has been silenced by RNAi, the pronuclei fail to migrate. This experiment revealed an important but previously unknown function of this gene in embryonic development. (B and C, from P. Gönczy et al., *Nature* 408:331–336, 2000. With permission from Macmillan Publishers Ltd.)

One of the fastest and easiest ways to silence genes in cells and organisms is via **RNA interference** (**RNAi**). Discovered in 1998, RNAi exploits a natural mechanism used in a wide variety of plants and animals to protect themselves against infection with certain viruses and the proliferation of mobile genetic elements (discussed in Chapter 9). The technique involves introducing into a cell or organism double-stranded RNA molecules with a nucleotide sequence that matches the gene to be inactivated. The double-stranded RNA is cleaved and processed by special RNAi machinery to produce shorter, double-stranded fragments called small interfering RNAs (siRNAs). These siRNAs are separated to form single-stranded RNA fragments that hybridize with the target gene's mRNAs and direct their degradation (see Figure 8–28). In some organisms, the same fragments can direct the production of more siRNAs, allowing continued inactivation of the target mRNAs.

RNAi is frequently used to inactivate genes in cultured mammalian cell lines, *Drosophila*, and the nematode *C. elegans*. Introducing double-stranded RNAs into *C. elegans* is particularly easy: the worm can be fed with *E. coli* that have been genetically engineered to produce the double-stranded RNAs that trigger RNAi (**Figure 10–27**). These RNAs are converted into siRNAs, which are then distributed throughout the animal's body to inhibit expression of the target gene in various tissues. For the many organisms whose genomes have been completely sequenced, RNAi can, in principle, be used to explore the function of any gene, and large collections of DNA vectors that produce these double-stranded RNAs are available for several species.

A Known Gene Can Be Deleted or Replaced with an Altered Version

Despite its usefulness, RNAi has some limitations. Non-target genes are sometimes inhibited along with the gene of interest, and certain cell types are resistant to RNAi entirely. Even for cell types in which the mechanism functions effectively, gene inactivation by RNAi is often temporary, earning the description "gene knockdown."

Fortunately, there are other, more specific and effective means of eliminating gene activity in cells and organisms. The coding sequence of a cloned gene can be mutated *in vitro* to change the functional properties of its protein product. Alternatively, the coding region can be left intact and the regulatory region of the gene changed, so that the amount of protein made will be altered or the gene will be expressed in a different type of cell or at a different time during development. By re-introducing this altered gene back into the organism from which it originally came, one can produce a mutant organism that can be studied to determine the gene's function. Often the altered gene is inserted into the genome of reproductive cells so that it can be stably inherited by subsequent generations. Organisms whose genomes have been altered in this way are known as **transgenic organisms**, or *genetically modified organisms* (*GMOs*); the introduced gene is called a *transgene*.

To study the function of a gene that has been altered *in vitro,* ideally one would like to generate an organism in which the normal gene has been

replaced by the altered one. In this way, the function of the mutant protein can be analyzed in the absence of the normal protein. A common way of doing this in mice makes use of cultured mouse embryonic stem (ES) cells (discussed in Chapter 20). These cells are first subjected to targeted gene replacement before being transplanted into a developing embryo to produce a mutant mouse, as illustrated in Figure 10–28.

Using a similar strategy, the activities of both copies of a gene can be eliminated entirely, creating a "**gene knockout**." To do this, one can either introduce an inactive, mutant version of the gene into cultured ES cells or delete the gene altogether. The ability to use ES cells to produce such "knockout mice" revolutionized the study of gene function, and the

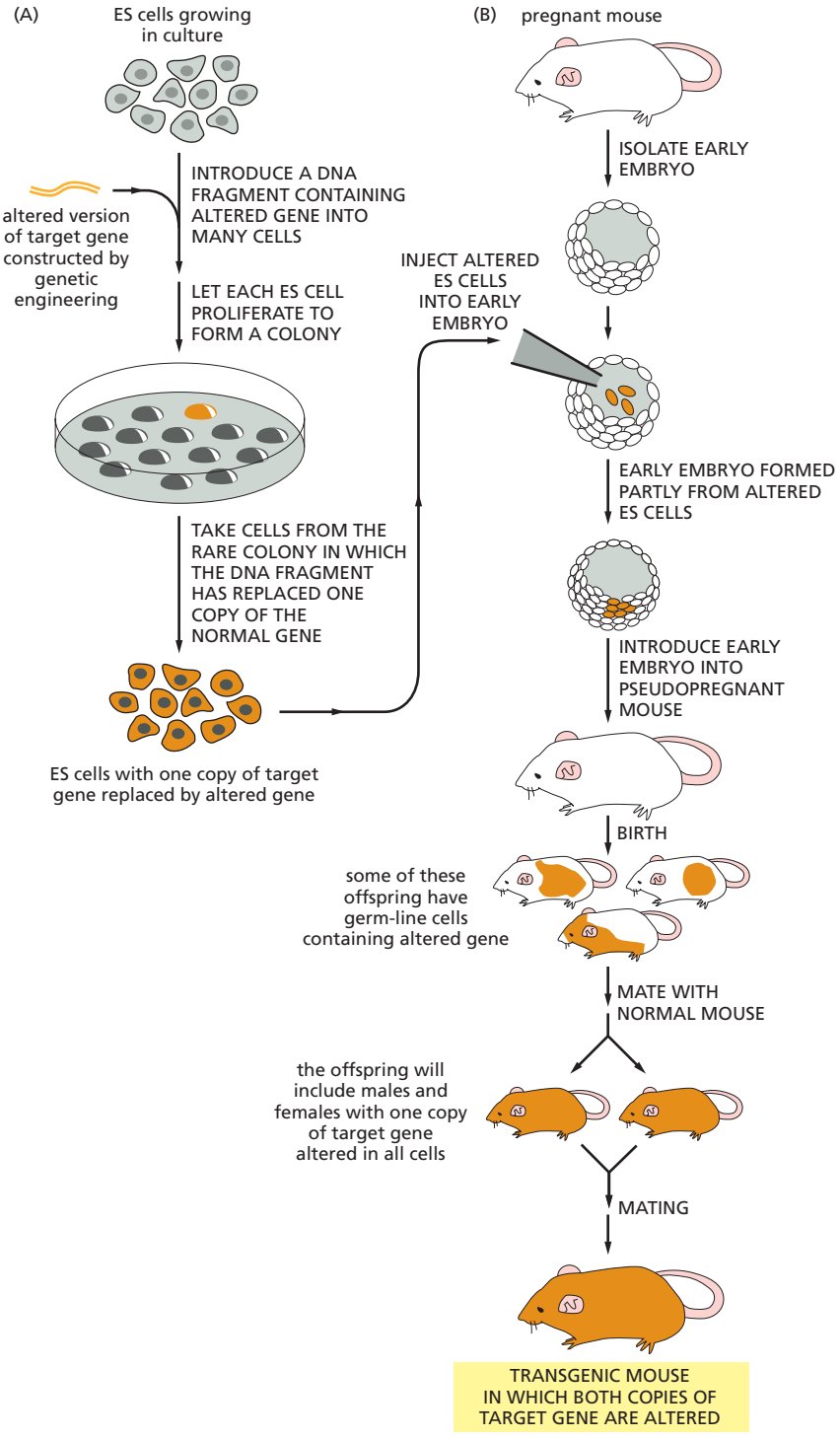
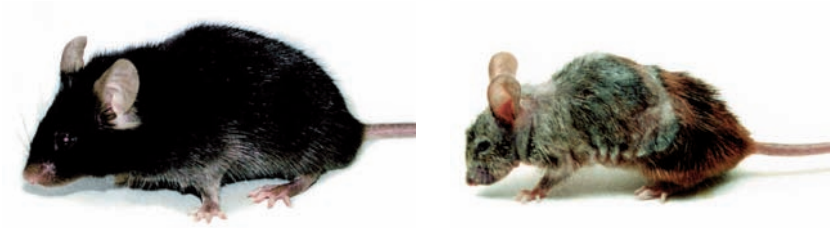


Figure 10–28 Targeted gene replacement in mice utilizes embryonic stem (ES) cells. (A) First, an altered version of the gene is introduced into cultured ES cells. In a few rare ES cells, the altered gene will replace the corresponding normal gene through homologous recombination (as described in Chapter 6, pp. 220–222 and Figure 6–31). Although the procedure is often laborious, these rare cells can be identified and cultured to produce many descendants, each of which carries an altered gene in place of one of its two normal corresponding genes. (B) Next, the altered ES cells are injected into a very early mouse embryo; the cells are incorporated into the growing embryo, which then develops into a mouse that contains some somatic cells (colored *orange*) that carry the altered gene. Some of these mice may also have germ-line cells that contain the altered gene; when bred with a normal mouse, some of the progeny of these mice will contain the altered gene in all of their cells. Such a mouse is called a "knock-in" mouse. If two such mice are bred, one can obtain progeny that contain two copies of the altered gene—one on each chromosome—in all of their cells.



(A) (B)

Figure 10–29 Transgenic mice with a mutant DNA helicase show premature aging. The helicase, encoded by the *Xpd* gene, is involved in both transcription and DNA repair. Compared with a wild-type mouse (A), a transgenic mouse that expresses a defective version of *Xpd* (B) exhibits many of the symptoms of premature aging, including osteoporosis, emaciation, early graying, infertility, and reduced life-span. The mutation in *Xpd* used here impairs the activity of the helicase and mimics a human mutation that causes trichothiodystrophy, a disorder characterized by brittle hair, skeletal abnormalities, and a greatly reduced life expectancy. These results support the hypothesis that an accumulation of DNA damage contributes to the aging process in both humans and mice. (From J. de Boer et al., *Science* 296:1276–1279, 2002. With permission from AAAS.)

technique is now being used to systematically determine the function of every mouse gene (**Figure 10–29**).

A variation of this technique can be used to produce *conditional knockout mice*, in which a known gene can be disrupted more selectively—only in a particular cell type or at a certain time in development. The strategy involves the introduction of an enzyme, called a recombinase, that can be directed to selectively excise—and thus disable—a gene of interest (**Figure 10–30**). Such conditional knockouts are useful for studying genes with a critical function during development, because mice missing these crucial genes often die before birth.

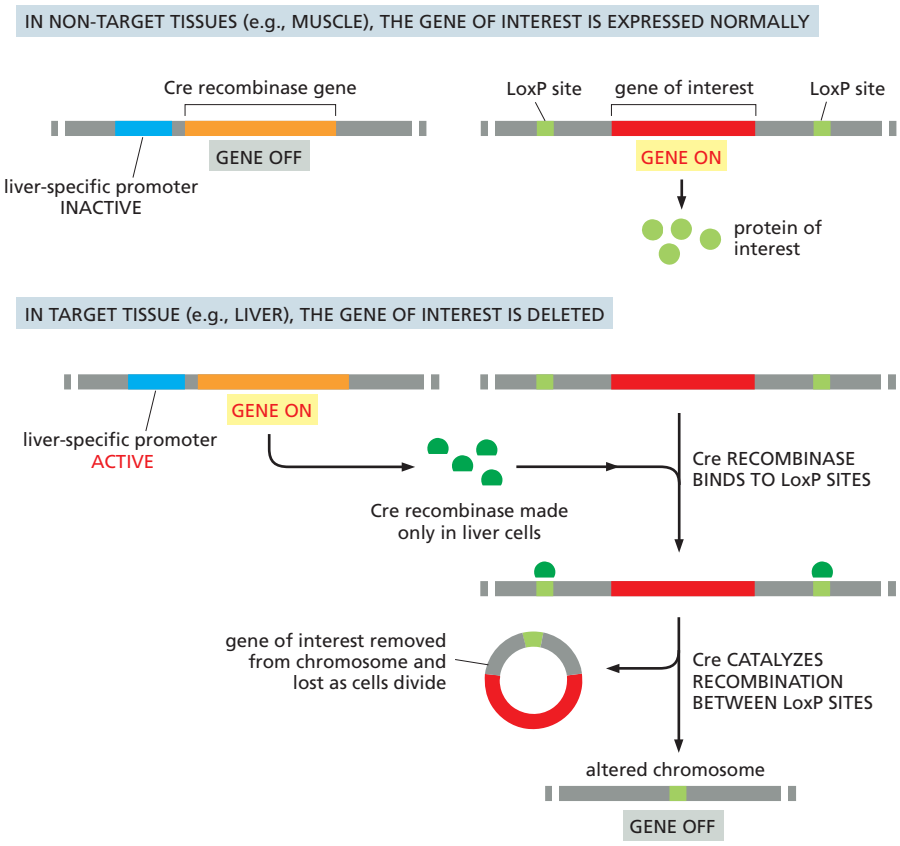


Figure 10–30 In conditional knockouts, a gene can be selectively disabled in a particular target tissue. The approach requires the insertion of two engineered segments of DNA into an animal's germ-line cells. The first contains the gene encoding a recombinase (in this case, Cre recombinase) that is under the control of a tissue-specific promoter. This promoter ensures that recombinase will be produced only in the target tissue. The second DNA molecule contains the gene of interest flanked by nucleotide sequences (in this case, LoxP recombination sites) that are recognized by the recombinase. The mouse is engineered so that this version of the gene of interest is the only copy the animal has.

In non-target tissues, no recombinase will be produced and the gene of interest will be expressed normally. In the target tissue, however, the tissue-specific promoter will be activated, allowing the recombinase to be produced. The enzyme will then bind to the LoxP sites and catalyze a recombination reaction that will excise the gene of interest—thus disabling it specifically in the target tissue.

Genes Can Be Edited with Great Precision Using the Bacterial CRISPR System

Bacteria employ several mechanisms to protect themselves from foreign DNA. One line of defense is provided by the restriction enzymes, as previously discussed. Recently, the discovery of another bacterial defense system led to the development of a powerful new method for editing genes in a variety of cells, tissues, and organisms. This system, called **CRISPR**, relies on a bacterial enzyme called Cas9, which produces a double-strand break in a molecule of DNA. Unlike restriction enzymes, Cas9 is not sequence-specific; to direct Cas9 to its target sequence, investigators provide the enzyme with a guide RNA molecule. This guide RNA, carried by Cas9, allows the enzyme to search the genome and bind to a segment of DNA with a complementary sequence (**Figure 10–31A**). The gene coding for Cas9 has been genetically engineered into a variety of organisms; thus, to use the CRISPR system to target a gene—or multiple genes—researchers need only introduce the appropriate guide RNAs (**Movie 10.2**).

As we saw in Chapter 6, double-strand breaks, like the one induced by Cas9, are often repaired by homologous recombination—a process that uses the information on an undamaged segment of DNA to repair the break. Thus, to replace a target gene using CRISPR, investigators simply provide an altered version of the gene to serve as a template for the homologous repair. In this way, a target gene can be selectively cut by the CRISPR system and replaced at high efficiency by an experimentally altered version of the gene (**Figure 10–31B**).The CRISPR system therefore provides another means of generating transgenic organisms.

Researchers are also adapting the CRISPR system for turning selected genes on or off. In this case, a catalytically inactive Cas9 protein can be

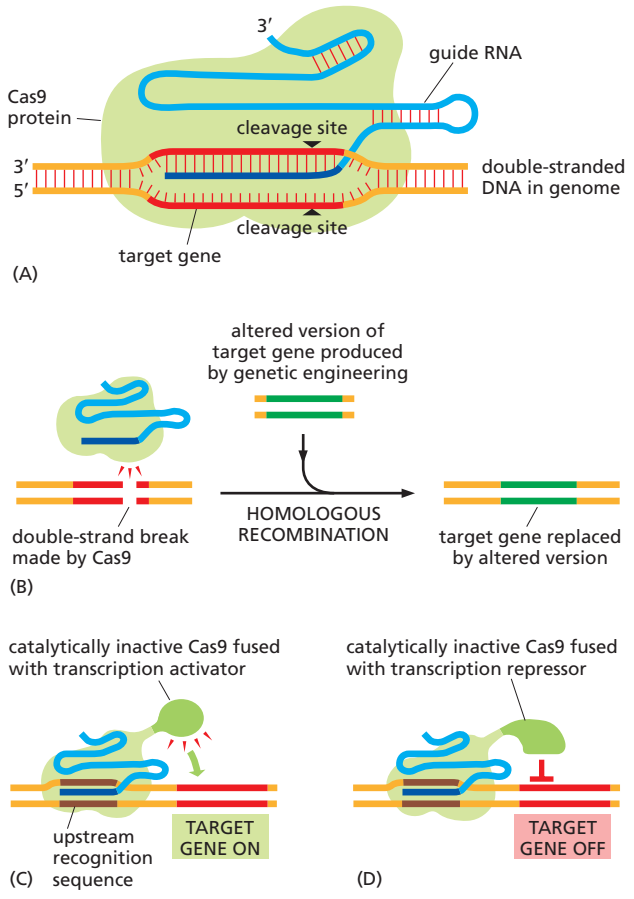


Figure 10–31 The CRISPR system can be used to study gene function in a variety of species. (A) The Cas9 protein, along with a guide RNA designed by the experimenter, are both artificially expressed in the cell or species of interest. One portion of the guide RNA (*light blue*) associates with Cas9, and another segment (*dark blue*) is designed to match a particular target sequence in the genome. (B) Once Cas9 has made a double-strand break in the target gene, that gene can be replaced with an experimentally altered gene by the enzymes that repair double-strand breaks through homologous recombination (see Figure 6–31). In this way, the CRISPR system promotes the precise and rapid replacement of a target gene. (C and D) By using a mutant form of Cas9 that can no longer cleave DNA, Cas9 can be used to activate a normally dormant gene (C) or turn off an actively expressed gene (D). (Adapted from P. Mali et al., *Nat. Methods* 10:957–963, 2013.)

fused to a transcription activator or repressor; this hybrid transcription regulator can then be directed to a target gene by the appropriate guide RNA (**Figure 10–31C and D**).

The transfer of the CRISPR system from bacteria to virtually all other experimental organisms—including mice, zebrafish, worms, flies, rice, and wheat—has revolutionized the study of gene function. Like the earlier discoveries of restriction enzymes and RNAi, this incredible breakthrough came from the work of scientists who were studying a fascinating biological phenomenon without—at first—realizing the enormous impact these discoveries would have on all aspects of biology, including human health. Such unintentional application highlights the fundamental importance of basic research.

Mutant Organisms Provide Useful Models of Human Disease

Technically speaking, transgenic approaches—including CRISPR—could be used to alter genes in the human germ line. Such manipulations would be unethical. However, transgenic technologies are currently being used to generate animal models of human diseases in which mutant genes play a major part.

With the explosion of DNA sequencing technologies, investigators can rapidly search the genomes of patients for mutations that cause or greatly increase the risk of their disease (discussed in Chapter 19). These mutations can then be introduced into animals, such as mice, that can be studied in the laboratory. The resulting transgenic animals, which often mimic some of the phenotypic abnormalities associated with the condition in patients, can be used to explore the cellular and molecular basis of the disease and to screen for drugs that could potentially be used therapeutically in humans.

An encouraging example is provided by *fragile X syndrome*, a neuropsychiatric disorder associated with intellectual impairment, neurological abnormalities, and often autism. The disease is caused by a mutation in the *fragile X mental retardation gene* (*FMR1*), which encodes a protein that inhibits the translation of mRNAs into proteins at synapses—the junctions where nerve cells communicate with one another (see Figure 12–39). Transgenic mice in which the *FMR1* gene has been disabled show many of the same neurological and behavioral abnormalities seen in patients with the disorder, and drugs that return synaptic protein synthesis to near-normal levels also reverse many of the problems seen in these mutant mice. Preliminary studies suggest that at least one of these drugs may benefit patients with the disease.

Transgenic Plants Are Important for both Cell Biology and Agriculture

Although we tend to think of DNA technology in terms of animal biology, these techniques have also had a profound impact on the study of plants. In fact, certain features of plants make them especially amenable to these methods.

When a piece of plant tissue is cultured in a sterile medium containing nutrients and appropriate growth regulators, some of the cells are stimulated to proliferate indefinitely in a disorganized manner, producing a mass of relatively undifferentiated cells called a *callus*. If the nutrients and growth regulators are carefully manipulated, one can induce the formation of a shoot within the callus, and in many species a whole new plant can be regenerated from such shoots. In a number of plants—including

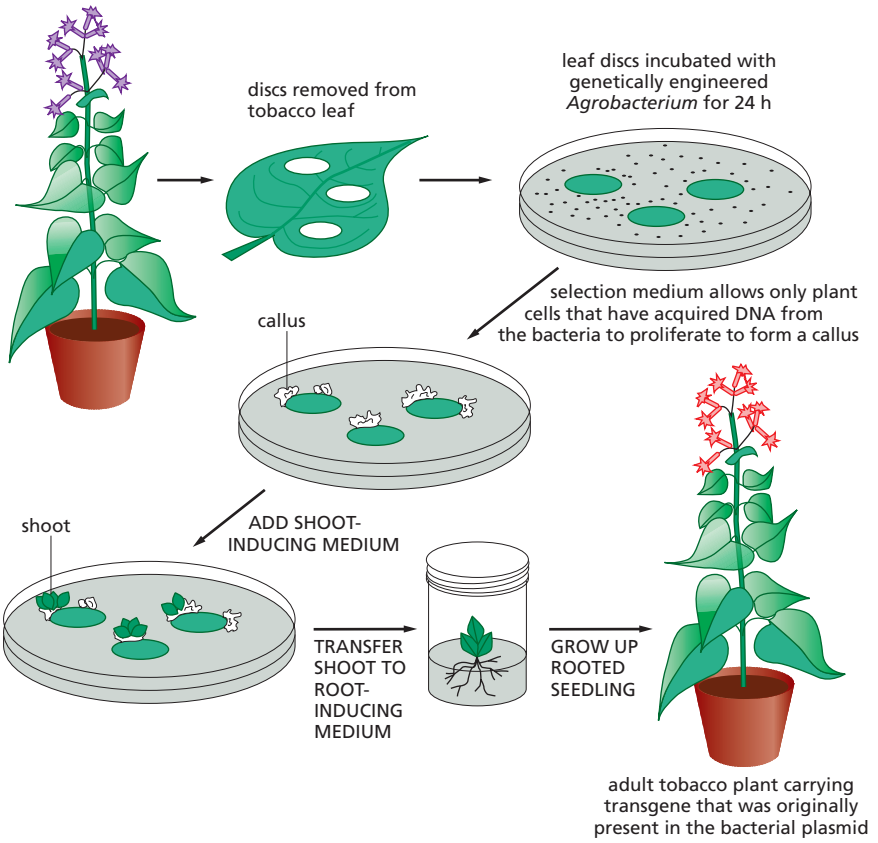
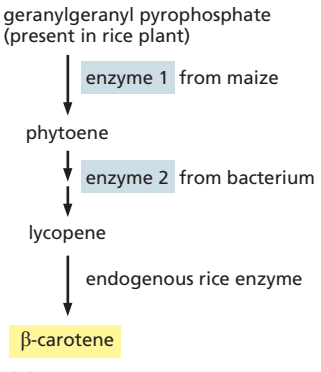


Figure 10–32 Transgenic plants can be made using recombinant DNA techniques optimized for plants. A disc is cut out of a leaf and incubated in a culture of *Agrobacterium* that carries a recombinant plasmid with both a selectable marker and a desired genetically engineered gene. The wounded plant cells at the edge of the disc release substances that attract the bacteria, which inject their DNA into the plant cells. Only those plant cells that take up the appropriate DNA and express the selectable marker gene survive and proliferate and form a callus. The manipulation of growth factors supplied to the callus induces it to form shoots, which subsequently root and grow into adult plants carrying the engineered gene.

tobacco, petunia, carrot, potato, and *Arabidopsis*—a single cell from such a callus can be grown into a small clump of cells from which a whole plant can be regenerated (see Figure 8–2B). Just as mutant mice can be derived by the genetic manipulation of embryonic stem cells in culture, transgenic plants can be created from plant cells transfected with DNA in culture (**Figure 10–32**).

The ability to produce transgenic plants has greatly accelerated progress in many areas of plant cell biology. It has played an important part, for example, in isolating receptors for growth regulators and in analyzing the mechanisms of morphogenesis and of gene expression in plants. These techniques have also opened up many new possibilities in agriculture that could benefit both the farmer and the consumer. They have made it possible, for example, to modify the ratio of lipid, starch, and protein in seeds, to impart pest and virus resistance to plants, and to create modified plants that tolerate extreme habitats such as salt marshes or water-stressed soil. One variety of rice has been genetically engineered to produce β-carotene, the precursor of vitamin A (**Figure 10–33**). If it replaced conventional rice, this "golden rice"—so called because of its



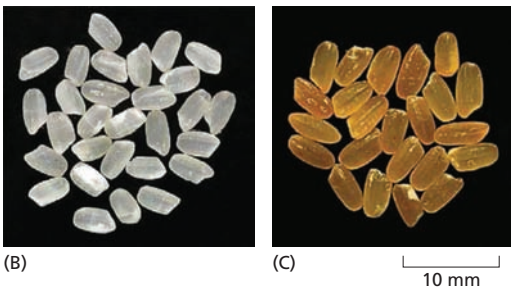


Figure 10–33 DNA technology allows the production of rice grains with high levels of β-carotene. To help reduce vitamin A deficiency in the developing world, a strain of rice, called "golden rice," was developed in which the edible part of the grain (called the endosperm) contains large amounts of β-carotene, which is converted in the human gut to vitamin A. (A) Rice plants, like most other plants, can synthesize β-carotene in their leaves from an abundant precursor (geranylgeranyl pyrophosphate) found in all plant tissues. However, the genes that code for two of the enzymes that act early in this biosynthetic pathway are turned off in the endosperm, preventing the production of β-carotene in rice grains. To produce golden rice, the genes for these two enzymes in the pathway were obtained from organisms that produce large amounts of β-carotene: one from maize and the other from a bacterium. Using DNA technology, these genes were connected to a promoter that drives gene expression in rice endosperm. Using the method outlined in Figure 10–32, this engineered DNA was then used to generate a transgenic rice plant that expresses these enzymes in endosperm, resulting in rice grains that contain high levels of β-carotene. Compared to the milled grains of wild-type rice (B), the grains of the transgenic rice are a deep yellow/orange due to the presence of β-carotene (C). (B and C, from J.A. Paine et al., *Nature Biotechnology*, Letters 23: 482–487, 2005. With permission from Macmillan Publishers Ltd.)

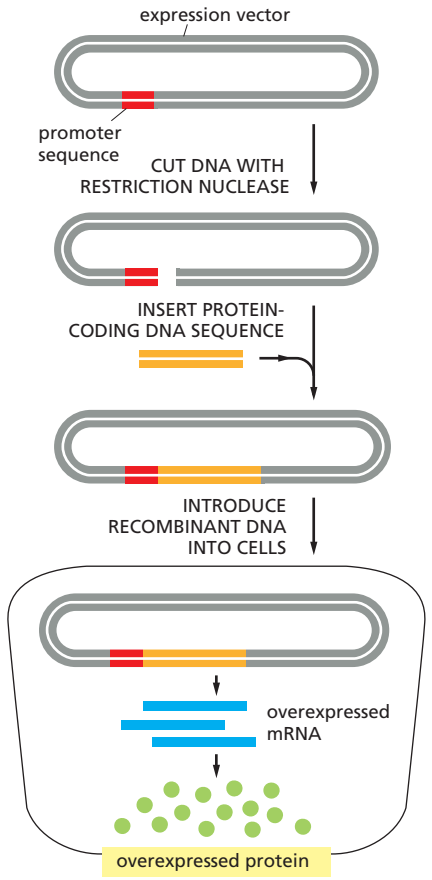


Figure 10–34 Large amounts of a protein can be produced from a protein-coding DNA sequence inserted into an expression vector and introduced into cells. Here, a plasmid vector has been engineered to contain a highly active promoter, which causes unusually large amounts of mRNA to be produced from the inserted protein-coding gene. Depending on the characteristics of the cloning vector, the plasmid is introduced into bacterial, yeast, insect, or mammalian cells, where the inserted gene is efficiently transcribed and translated into protein.

yellow/orange color—could help to alleviate severe vitamin A deficiency, which causes blindness in hundreds of thousands of children in the developing world each year.

Even Rare Proteins Can Be Made in Large Amounts Using Cloned DNA

One of the most important contributions of DNA cloning and genetic engineering to cell biology is that they make it possible to produce any protein, including the rare ones, in large amounts. Such high-level production is usually accomplished by using specially designed vectors known as *expression vectors*. These vectors include transcription and translation signals that direct an inserted gene to be expressed at high levels. Different expression vectors are designed for use in bacterial, yeast, insect, or mammalian cells, each containing the appropriate regulatory sequences for transcription and translation in these cells (**Figure 10–34**). The expression vector is replicated at each round of cell division, so that the transfected cells in the culture are able to synthesize large amounts of the protein of interest—sometimes comprising 1–10% of the total cell protein. It is usually a simple matter to purify this protein away from the other proteins made by the host cell.

This technology is now used to make large amounts of many medically useful proteins, including hormones (such as insulin), growth factors, therapeutic antibodies, and viral coat proteins for use in vaccines. Expression vectors also allow scientists to produce many proteins of biological interest in large enough amounts for detailed structural and functional studies that were once impossible—especially for proteins that are normally present in very small amounts, such as some receptors and transcription regulators. Recombinant DNA techniques thus allow scientists to move with ease from protein to gene, and vice versa, so that the functions of both can be explored from multiple directions (**Figure 10–35**).

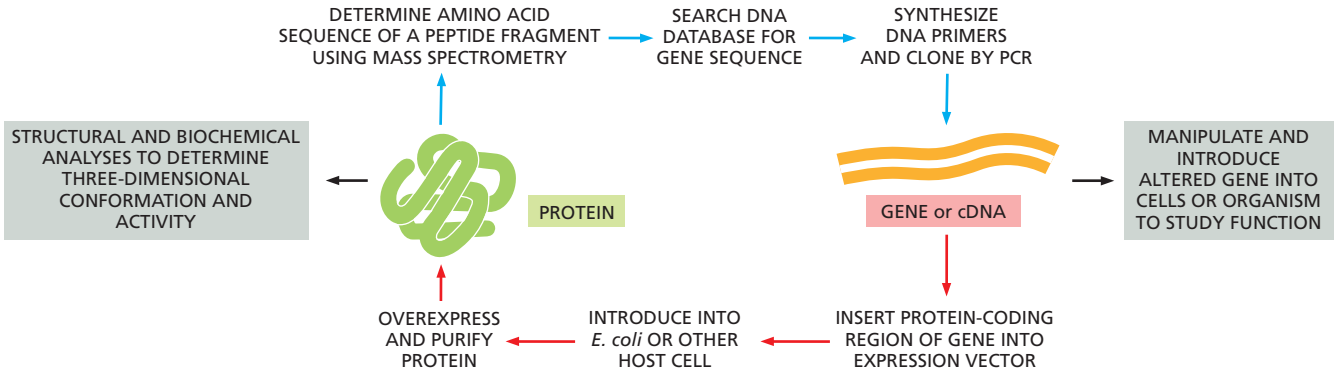


Figure 10–35 Recombinant DNA techniques make it possible to move experimentally from gene to protein or from protein to gene. A small quantity of a purified protein or peptide fragment is used to obtain a partial amino acid sequence, which is used to search a DNA database for the corresponding nucleotide sequence. This sequence is used to synthesize DNA primers, which can be used to clone the gene by PCR from a sequenced genome (see Figure 10–13). Once the gene has been isolated and sequenced, its protein-coding sequence can be inserted into an expression vector to produce large quantities of the protein (see Figure 10–34), which can then be studied biochemically or structurally. In addition to producing protein, the gene or DNA can also be manipulated and introduced into cells or organisms to study its function.

KEY TERMS

complementary DNA (cDNA)
cDNA library
CRISPR
dideoxy (Sanger) sequencing
DNA cloning
DNA library
DNA ligase
gene knockout
genomic library
green fluorescent protein (GFP)
hybridization
in situ hybridization
plasmid
polymerase chain reaction (PCR)
recombinant DNA
reporter gene
restriction enzyme
RNA interference (RNAi)
RNA-Seq
transformation
transgenic organism

ESSENTIAL CONCEPTS

- DNA technology has revolutionized the study of cells, making it possible to pick out any gene at will from the thousands of genes in a cell and to determine its nucleotide sequence.
- A crucial element in this technology is the ability to cut a large DNA molecule into a specific and reproducible set of DNA fragments using restriction enzymes, each of which cuts the DNA double helix only at a particular nucleotide sequence.
- DNA fragments can be separated from one another on the basis of size by gel electrophoresis.
- DNA cloning techniques enable any DNA sequence to be selected from millions of other sequences and produced in unlimited amounts in pure form.
- DNA fragments can be joined together *in vitro* by using DNA ligase to form recombinant DNA molecules that are not found in nature.
- DNA fragments can be maintained and amplified by inserting them into a larger DNA molecule capable of replication, such as a plasmid. This recombinant DNA molecule is then introduced into a rapidly dividing host cell, usually a bacterium, so that the DNA is replicated at each cell division.
- A collection of cloned fragments of chromosomal DNA representing the complete genome of an organism is known as a genomic library. The library is often maintained as millions of clones of bacteria, each different clone carrying a different fragment of the organism's genome.
- cDNA libraries contain cloned DNA copies of the total mRNA of a particular type of cell or tissue. Unlike genomic DNA clones, cDNA clones contain predominantly protein-coding sequences; they lack introns, regulatory DNA sequences, and promoters. Thus they are useful when the cloned gene is needed to make a protein.
- Nucleic acid hybridization can detect any given DNA or RNA sequence in a mixture of nucleic acid fragments. This technique depends on highly specific base-pairing between a labeled, single-stranded DNA or RNA probe and another nucleic acid with a complementary sequence.
- The polymerase chain reaction (PCR) is a powerful form of DNA amplification that is carried out *in vitro* using a purified DNA polymerase. Cloning via PCR requires prior knowledge of the sequence to be amplified, because two synthetic oligonucleotide primers must be synthesized that bracket the portion of DNA to be replicated.

- DNA sequencing techniques have become increasingly fast and cheap, so that the entire genome sequences of thousands of different organisms are now known, including thousands of individual humans.
- Using DNA technology, a protein can be joined to a molecular tag, such as green fluorescent protein (GFP), which allows its movement to be tracked inside a cell and, in some cases, inside a living organism.
- *In situ* nucleic acid hybridization can be used to detect the precise location of genes on chromosomes and of RNAs in cells and tissues.
- RNA-Seq can be used to monitor the expression of all of the genes in a cell or tissue.
- Cloned genes can be altered *in vitro* and stably inserted into the genome of a cell or an organism to study their function. Such mutants are called transgenic organisms.
- The expression of particular genes can be inhibited in cells or organisms by the technique of RNA interference (RNAi), which prevents an mRNA from being translated into protein.
- Genes can be deleted or modified with high specificity by the CRISPR system, which uses guide mRNAs to promote DNA cleavage at a specific nucleotide sequence in the genome.
- Bacteria, yeasts, and mammalian cells can be engineered to synthesize large quantities of any protein whose gene has been cloned, making it possible to study proteins that are otherwise rare or difficult to isolate.

QUESTIONS

QUESTION 10–5

What are the consequences for a dideoxy DNA sequencing reaction if the ratio of dideoxyribonucleoside triphosphates to deoxyribonucleoside triphosphates is increased? What happens if this ratio is decreased?

QUESTION 10–6

Almost all the cells in an individual animal contain identical genomes. In an experiment, a tissue composed of several different cell types is fixed and subjected to *in situ* hybridization with a DNA probe to a particular gene. To your surprise, the hybridization signal is much stronger in some cells than in others. How might you explain this result?

QUESTION 10–7

After decades of work, Dr. Ricky M. isolated a small amount of attractase—an enzyme that produces a powerful human pheromone—from hair samples of Hollywood celebrities. To take advantage of attractase for his personal use, he obtained a complete genomic clone of the attractase gene, connected it to a strong bacterial promoter on an expression plasmid, and introduced the plasmid into *E. coli* cells. He was devastated to find that no attractase was produced in the cells. What is a likely explanation for his failure?

QUESTION 10–8

Which of the following statements are correct? Explain your answers.

A. Restriction enzymes cut DNA at specific sites that are always located between genes.

B. DNA migrates toward the positive electrode during electrophoresis.

C. Clones isolated from cDNA libraries contain promoter sequences.

D. PCR utilizes a heat-stable DNA polymerase because for each amplification step, double-stranded DNA must be heat-denatured.

E. Digestion of genomic DNA with AluI, a restriction enzyme that recognizes a four-nucleotide sequence, produces fragments that are all exactly 256 nucleotides in length.

F. To make a cDNA library, both a DNA polymerase and a reverse transcriptase must be used.

G. DNA fingerprinting by PCR relies on the fact that different individuals have different numbers of repeats in STR regions in their genome.

H. It is possible for a coding region of a gene to be present in a genomic library prepared from a particular tissue but to be absent from a cDNA library prepared from the same tissue.

QUESTION 10–9

A. What is the sequence of the DNA that was used in the sequencing reaction shown in Figure Q10–9? The four lanes show the products of sequencing reactions that contained ddG (lane 1), ddA (lane 2), ddT (lane 3), and ddC (lane 4). The numbers to the right of the autoradiograph represent the positions of marker DNA fragments of 50 and 116 nucleotides.

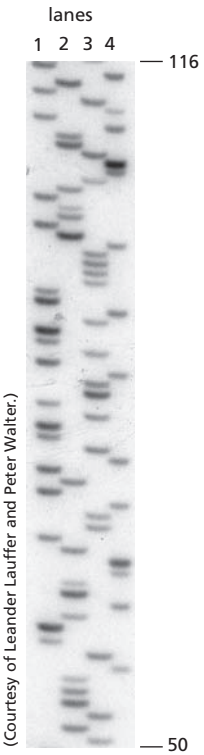


Figure Q10–9

B. This DNA was derived from the middle of a cDNA clone of a mammalian protein. Using the genetic code table (see Figure 7–27), can you determine the amino acid sequence of this portion of the protein?

QUESTION 10–10

A. How many different DNA fragments would you expect to obtain if you cleaved human genomic DNA with HaeIII? (Recall that there are 3×10^9 nucleotide pairs per haploid genome.) How many fragments would you expect with EcoRI?

B. Human genomic libraries used for DNA sequencing are often made from fragments obtained by cleaving human DNA with HaeIII in such a way that the DNA is only partially digested; that is, not all the possible HaeIII sites have been cleaved. What is a possible reason for doing this?

QUESTION 10–11

A molecule of double-stranded DNA was cleaved with restriction enzymes, and the resulting products were separated by gel electrophoresis (Figure Q10–11). You do not know if the molecule is linear DNA or a DNA circle. DNA fragments of known sizes were electrophoresed on the same gel for use as size markers (*left* lane). The size of the DNA markers is given in kilobase pairs (kb), where 1 kb = 1000 nucleotide pairs. Using the size markers as a guide, estimate the length of each restriction fragment obtained. From this information, construct a map of the original DNA molecule indicating the relative positions of all the restriction enzyme cleavage sites.

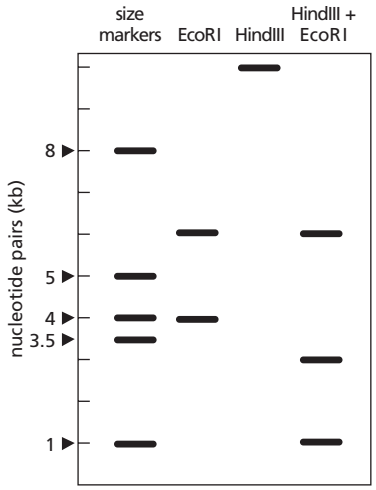


Figure Q10–11

QUESTION 10–12

There has been a colossal snafu in the maternity ward of your local hospital. Four sets of male twins, born within an hour of each other, were inadvertently shuffled in the excitement occasioned by that unlikely event. You have been called in to set things straight. As a first step, you would like to match each baby with his twin. (Many newborns look alike so you don't want to rely on appearance alone.) To that end you analyze a small blood sample from each infant using a hybridization probe that detects short tandem repeats (STRs) located in widely scattered regions of the genome. The results are shown in Figure Q10–12.

Figure Q10–12

A. Which infants are twins? Which are identical twins?

B. How could you match a pair of twins to the correct parents?

QUESTION 10–13

One of the first organisms that was genetically modified using recombinant DNA technology was a bacterium that normally lives on the surface of strawberry plants. This bacterium makes a protein, called ice-protein, that causes the efficient formation of ice crystals around it when the temperature drops to just below freezing. Thus, strawberries harboring this bacterium are particularly susceptible to frost damage because their cells are destroyed by the ice crystals. Consequently, strawberry farmers have a considerable interest in preventing ice crystallization.

A genetically engineered version of this bacterium was constructed in which the ice-protein gene was knocked out. The mutant bacteria were then introduced in large numbers into strawberry fields, where they displaced the normal bacteria by competition for their ecological niche. This approach has been successful: strawberries bearing the mutant bacteria show a much reduced susceptibility to frost damage.

At the time they were first carried out, the initial open-field trials triggered an intense debate because they represented the first release into the environment of an organism that had been genetically engineered using recombinant DNA technology. Indeed, all preliminary experiments were carried out with extreme caution and in strict containment.

Do you think that bacteria lacking the ice-protein could be isolated without the use of modern DNA technology? Is it likely that such mutations have already occurred in nature? Would the use of a mutant bacterial strain isolated from nature be of lesser concern? Should we be concerned about the risks posed by the application of recombinant DNA techniques in agriculture and medicine? Do the potential benefits outweigh the risks? Explain your answers.