

20

DNA Tools and Biotechnology

KEY CONCEPTS

- 20.1 DNA sequencing and DNA cloning are valuable tools for genetic engineering and biological inquiry
- 20.2 Biologists use DNA technology to study gene expression and function
- 20.3 Cloned organisms and stem cells are useful for basic research and other applications
- 20.4 The practical applications of DNA-based biotechnology affect our lives in many ways



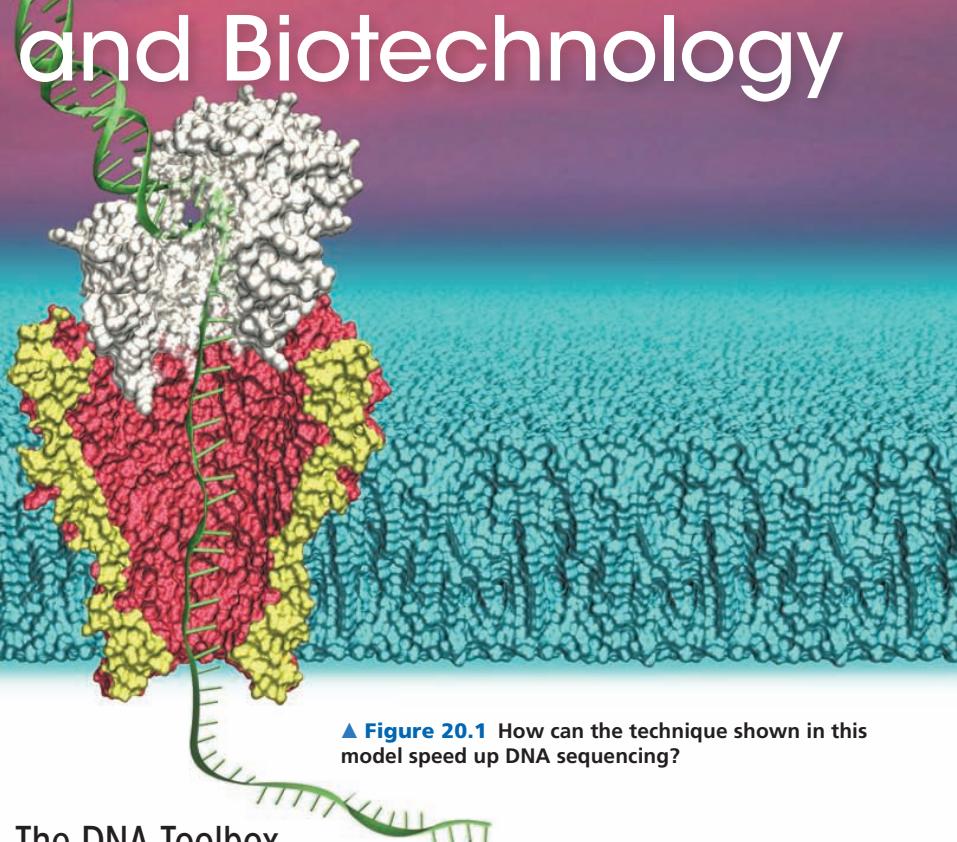
The DNA Toolbox

The last five to ten years have seen some extraordinary feats in biology, among them determination of the complete DNA sequences of several extinct species, including woolly mammoths (see below, left), Neanderthals, and a 700,000-year-old horse. Pivotal to those discoveries was the sequencing of the human genome, essentially completed in 2003. This endeavor marked a turning point in biology because it sparked remarkable technological advances in DNA sequencing.

The first human genome sequence took several years at a cost of 1 billion dollars; the time and cost of sequencing a genome have been in free fall since then. **Figure 20.1** shows a model of a sequencing technique in which the nucleotides of a single strand of DNA are passed one by one through a tiny pore in a membrane, and the resulting tiny changes in an electrical current are used to determine the nucleotide sequence. Developers of this technique, which you will learn more about later in the chapter, claim that ultimately we will be able to sequence a human genome in about 6 hours on a \$900 device the size of a pack of gum.

In this chapter, we'll first describe the main techniques for sequencing and manipulating DNA—**DNA technology**—and for using these DNA tools to analyze gene expression. Next, we'll explore advances in cloning organisms and producing stem cells, techniques that have both expanded our basic understanding of biology and enhanced our ability to apply that understanding to global problems. In the last

▲ **Figure 20.1** How can the technique shown in this model speed up DNA sequencing?



section, we'll survey the practical applications of DNA-based **biotechnology**, the manipulation of organisms or their components to make useful products. Today, the applications of DNA technology affect everything from agriculture to criminal law to medical research. We will end by considering some of the important social and ethical issues that arise as biotechnology becomes more pervasive in our lives.

CONCEPT 20.1

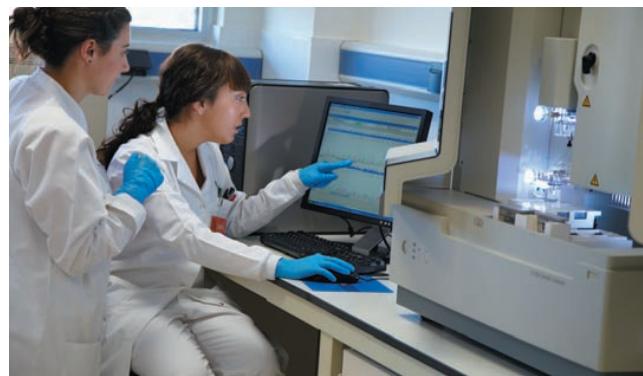
DNA sequencing and DNA cloning are valuable tools for genetic engineering and biological inquiry

The discovery of the structure of the DNA molecule, with its two complementary strands, opened the door for the development of DNA sequencing and many other techniques used in biological research today. Key to many of these techniques is **nucleic acid hybridization**, the base pairing of one strand of a nucleic acid to the complementary sequence on a strand from *another* nucleic acid molecule. In this section, we'll first describe DNA sequencing techniques. Then we'll explore other important methods used in **genetic engineering**, the direct manipulation of genes for practical purposes.

DNA Sequencing

Researchers can exploit the principle of complementary base pairing to determine the complete nucleotide sequence of a DNA molecule, a process called **DNA sequencing**. The DNA is first cut into fragments, and then each fragment is sequenced. Today, sequencing is carried out by machines (**Figure 20.2**). The first automated procedure used a technique called *dideoxyribonucleotide* (or *dideoxy*) *chain termination sequencing*. In this technique, one strand of a DNA fragment is used as a template for synthesis of a nested set of complementary fragments; these are further analyzed to yield the sequence, as shown in detail in **Figure 20.3**. Biochemist Frederick Sanger received the Nobel Prize in 1980 for developing this method. Dideoxy sequencing is still widely used today for routine small-scale sequencing jobs, in machines like that shown in Figure 20.2a.

In the last ten years, “next-generation sequencing” techniques have been developed that do not rely on chain termination. Instead, DNA fragments are amplified (copied) to yield an enormous number of identical fragments (**Figure 20.4**). A specific strand of each fragment is immobilized, and the complementary strand is synthesized, one nucleotide at a time. A chemical technique enables electronic monitors to identify in real time which of the four nucleotides is added; this method is thus called *sequencing by synthesis*. Thousands or hundreds of thousands of fragments, each about 400–1,000 nucleotides long, are sequenced in parallel in machines like those shown in Figure 20.2b, accounting for the high rate of nucleotides



(a) Standard sequencing machine



(b) Next-generation sequencing machines

▲ **Figure 20.2 DNA sequencing machines.** (a) This standard sequencing machine uses the dideoxy chain termination sequencing method (see Figure 20.3). It can sequence up to about 120,000 nucleotides in 10 hours and is used for sequencing small numbers of samples with shorter sequences. (b) Next-generation sequencing machines use “sequencing by synthesis” (see Figure 20.4). Today's machines can sequence 700–900 million nucleotides in 10 hours and are used for larger sequencing jobs.

sequenced per hour. This is an example of “high-throughput” DNA technology, and is currently the method of choice for studies where massive numbers of DNA samples—even representing an entire genome—are being sequenced.

Further technical developments have given rise to “third-generation sequencing,” with each new technique being faster and less expensive than the previous. In these new methods, the DNA is neither cut into fragments nor amplified. Instead, a single, very long DNA molecule is sequenced on its own. Several groups have been working on the idea of moving a single strand of a DNA molecule through a very small pore (a *nanopore*) in a membrane, detecting the bases one by one by their interruption of an electrical current. One model of this concept is shown in Figure 20.1, in which the pore is a protein channel embedded in a lipid membrane. (Other researchers are using artificial membranes and nanopores.) The idea is that each type of base would interrupt the current for a slightly different length of time. This example is only one of many different approaches to further increase the rate and cut the cost of sequencing.

▼ Figure 20.3

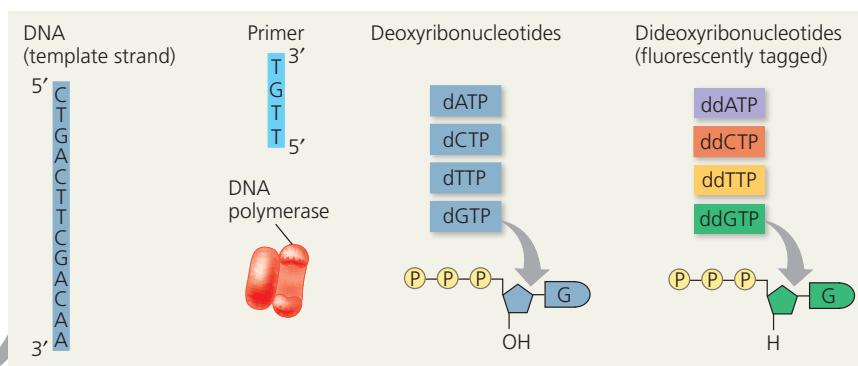
Research Method

Dideoxy Chain Termination Method for Sequencing DNA

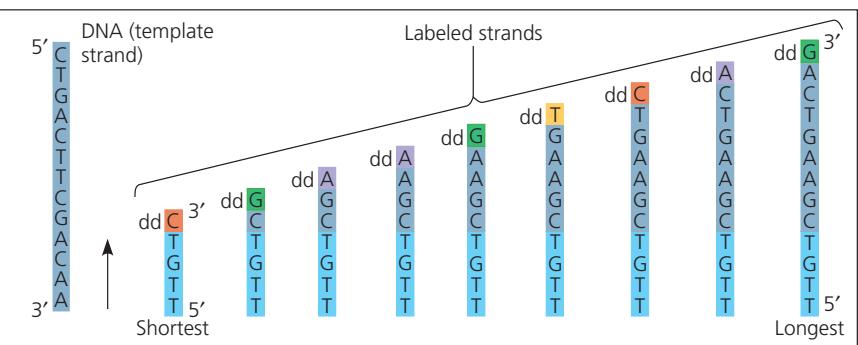
Application The sequence of nucleotides in any DNA fragment of up to 800–1,000 base pairs in length can be determined rapidly with machines that carry out sequencing reactions and separate the labeled reaction products by length.

Technique This method is based on synthesis of a nested set of DNA strands complementary to one strand of a DNA fragment. Each new strand starts with the same primer and ends with a dideoxyribonucleotide (ddNTP), a modified deoxyribonucleotide (dNTP). Incorporation of a ddNTP terminates a growing DNA strand because it lacks a 3' —OH group, the site for attachment of the next nucleotide (see Figure 16.14). In the set of new strands, each nucleotide position along the original sequence is represented by strands ending at that point with the complementary ddNTP. Because each type of ddNTP is tagged with a distinct fluorescent label, the identity of the ending nucleotides of the new strands, and ultimately the entire original sequence, can be determined.

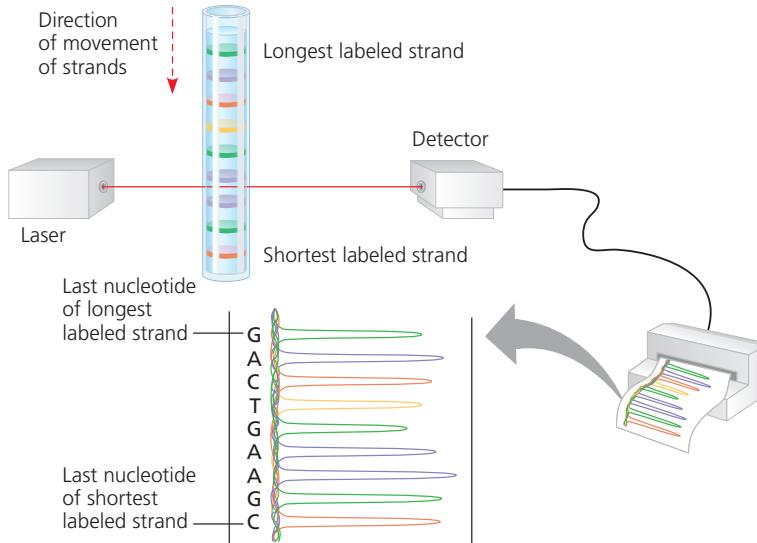
- 1 The fragment of DNA to be sequenced is denatured into single strands and incubated in a test tube with the necessary ingredients for DNA synthesis: a primer designed to base-pair with the known 3' end of the template strand, DNA polymerase, the four dNTPs, and the four ddNTPs, each tagged with a specific fluorescent molecule.



- 2 Synthesis of each new strand starts at the 3' end of the primer and continues until a ddNTP happens to be inserted instead of the equivalent dNTP. The incorporated ddNTP prevents further elongation of the strand. Eventually, a set of labeled strands of every possible length is generated, with the color of the tag representing the last nucleotide in the sequence.



- 3 The labeled strands in the mixture are separated by passage through a gel that allows shorter strands to move through more quickly than longer ones. For DNA sequencing, the gel is in a capillary tube, and its small diameter allows a fluorescence detector to sense the color of each fluorescent tag as the strands come through. Strands differing in length by as little as one nucleotide can be distinguished from each other.

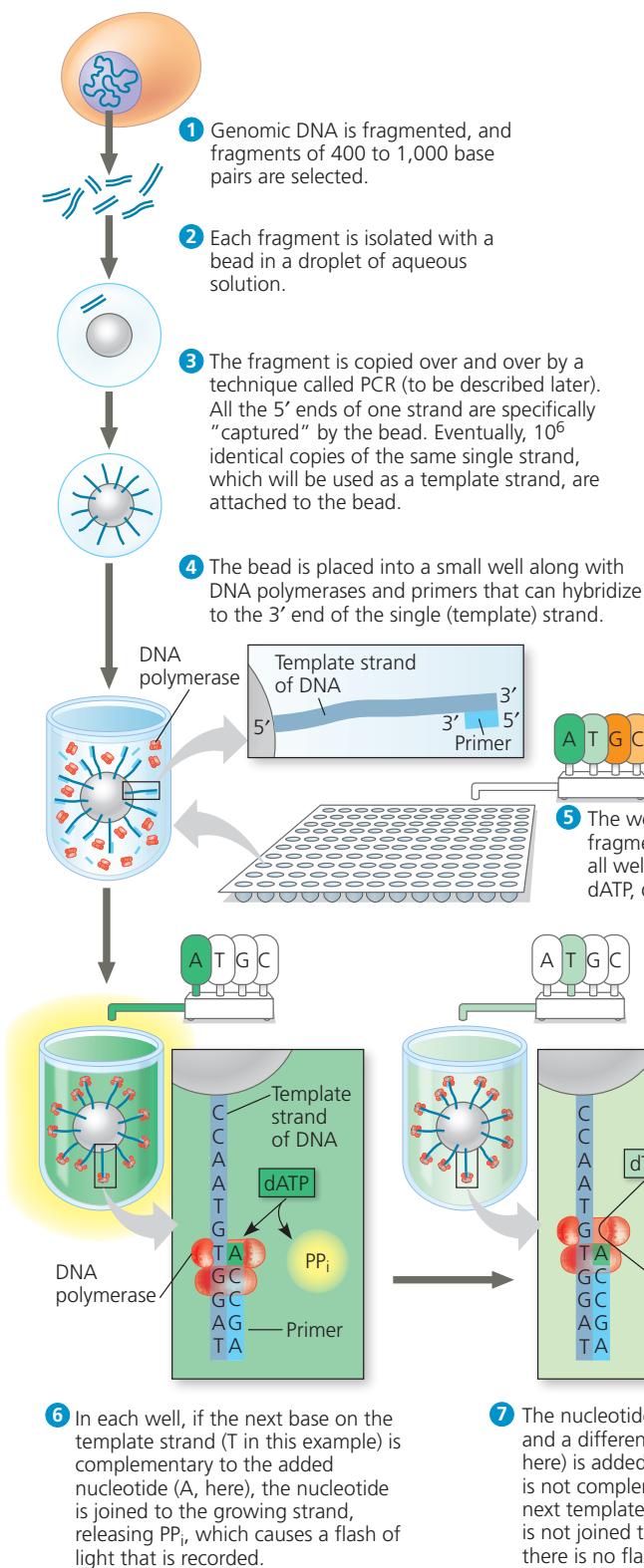


Results The color of the fluorescent tag on each strand indicates the identity of the nucleotide at its 3'-end. The results can be printed out as a spectrogram, and the sequence, which is complementary to the template strand, can then be read from bottom (shortest strand) to top (longest strand). (Notice that the sequence here begins after the primer.)

▼ Figure 20.4

Research Method

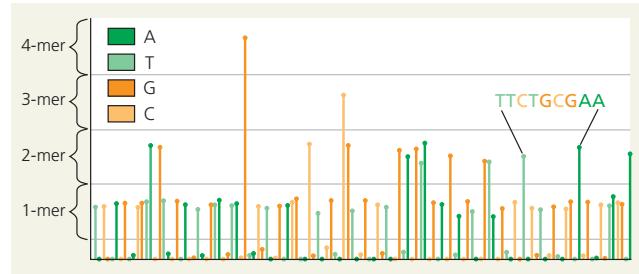
Next-Generation Sequencing



Application In current next-generation sequencing techniques, each fragment is 400–1,000 nucleotides long; by sequencing the fragments in parallel, 700–900 million nucleotides can be sequenced in 10 hours.

Technique See numbered steps and diagrams.

Results Each of the 2,000,000 wells in the multiwell plate, which holds a different fragment, yields a different sequence. The results for one fragment are shown below as a "flow-gram." The sequences of the entire set of fragments are analyzed using computer software, which "stitches" them together into a whole sequence—here, an entire genome.



INTERPRET THE DATA If the template strand has two or more identical nucleotides in a row, their complementary nucleotides will be added one after the other in the same flow step. How are two or more of the same nucleotide (in a row) detected in the flow-gram? (See sample on the right.) Write out the sequence of the first 25 nucleotides in the flow-gram above, starting from the left. (Ignore the very short lines.)

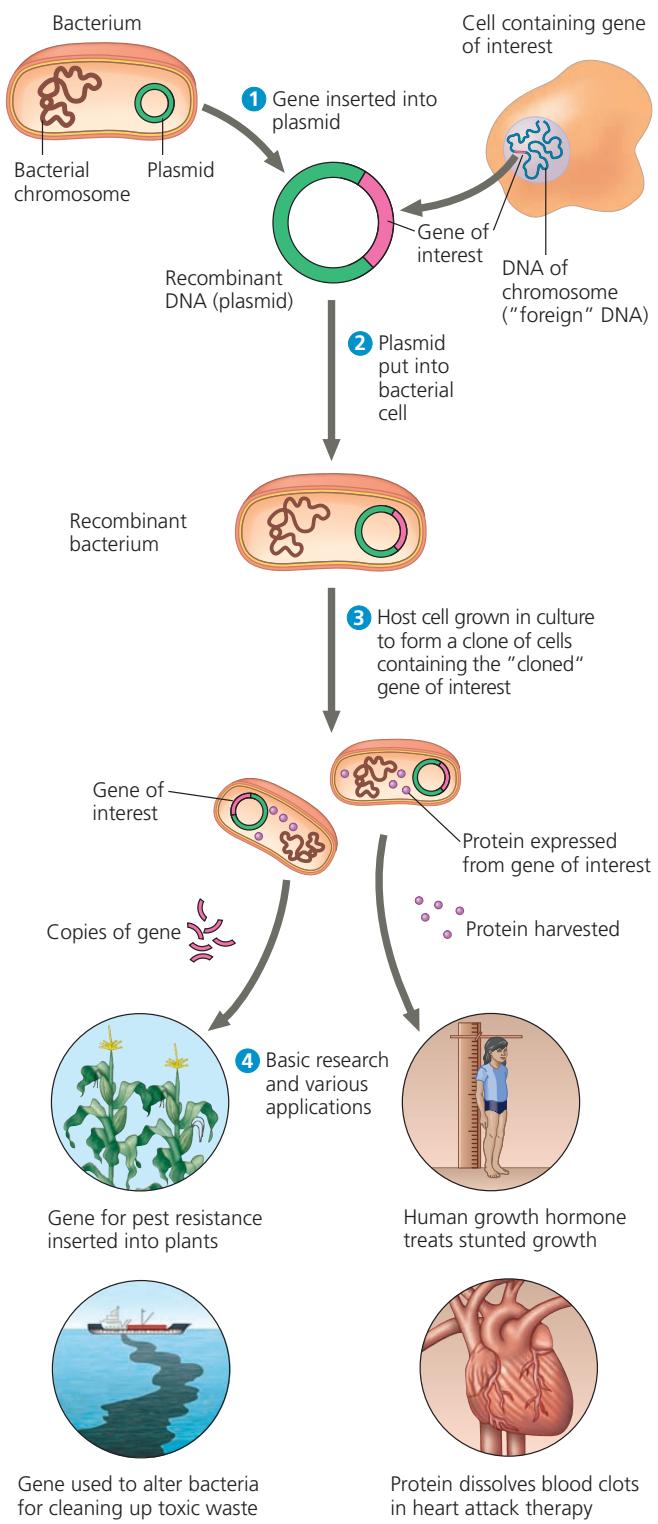
Improved DNA sequencing techniques have transformed the way in which we can explore fundamental biological questions about evolution and how life works. Little more than a decade after the human genome sequence was announced, researchers had completed sequencing roughly 4,000 bacterial, 190 archaeal, and 180 eukaryotic genomes, with more than 17,000 additional species under way. Complete genome sequences have been determined for cells from several cancers, for ancient humans, and for the many bacteria that live in the human intestine. In Chapter 21, you'll learn more about how this new sequencing technology has informed us about the evolution of species and the evolution of the genome itself. Now, let's consider how individual genes are studied.

Making Multiple Copies of a Gene or Other DNA Segment

A molecular biologist studying a particular gene or group of genes faces a challenge. Naturally occurring DNA molecules are very long, and a single molecule usually carries many hundreds or even thousands of genes. Moreover, in many eukaryotic genomes, protein-coding genes occupy only a small proportion of the chromosomal DNA, the rest being noncoding nucleotide sequences. A single human gene, for example, might constitute only 1/100,000 of a chromosomal DNA molecule. As a further complication, the distinctions between a gene and the surrounding DNA are subtle, consisting only of differences in nucleotide sequence. To work directly with specific genes, scientists have developed methods for preparing well-defined segments of DNA in multiple identical copies, a process called **DNA cloning**.

Most methods for cloning pieces of DNA in the laboratory share certain general features. One common approach uses bacteria, most often *Escherichia coli*. Recall from Figure 16.12 that the *E. coli* chromosome is a large circular molecule of DNA. In addition to their bacterial chromosome, *E. coli* and many other bacteria also have **plasmids**, small, circular DNA molecules that are replicated separately. A plasmid has only a small number of genes; these genes may be useful when the bacterium is in a particular environment but may not be required for survival or reproduction under most conditions.

To clone pieces of DNA using bacteria, researchers first obtain a plasmid (originally isolated from a bacterial cell and genetically engineered for efficient cloning) and insert DNA from another source ("foreign" DNA) into it (**Figure 20.5**). The resulting plasmid is now a **recombinant** DNA molecule, a molecule containing DNA from two different sources, very often different species. The plasmid is then returned to a bacterial cell, producing a *recombinant bacterium*. This single cell reproduces through repeated cell divisions to form a clone of cells, a population of genetically identical cells. Because the dividing bacteria replicate the recombinant plasmid and pass it on to their descendants, the



▲ Figure 20.5 Gene cloning and some uses of cloned genes. In this simplified diagram of gene cloning, we start with a plasmid (originally isolated from a bacterial cell) and a gene of interest from another organism. Only one plasmid and one copy of the gene of interest are shown at the top of the figure, but the starting materials would include many of each.

foreign DNA and any genes it carries are cloned at the same time. The production of multiple copies of a single gene is a type of DNA cloning called **gene cloning**.

In our example in Figure 20.5, the plasmid acts as a **cloning vector**, a DNA molecule that can carry foreign DNA into a host cell and replicate there. Bacterial plasmids are widely used as cloning vectors for several reasons: They can be readily obtained from commercial suppliers, manipulated to form recombinant plasmids by insertion of foreign DNA in a test tube (*in vitro*, from the Latin meaning “in glass”), and then easily introduced into bacterial cells. Moreover, recombinant bacterial plasmids (and the foreign DNA they carry) multiply rapidly owing to the high reproductive rate of their host (bacterial) cells. The foreign DNA in Figure 20.5 is a gene from a eukaryotic cell; we will describe in more detail how the foreign DNA segment was obtained later in this section.

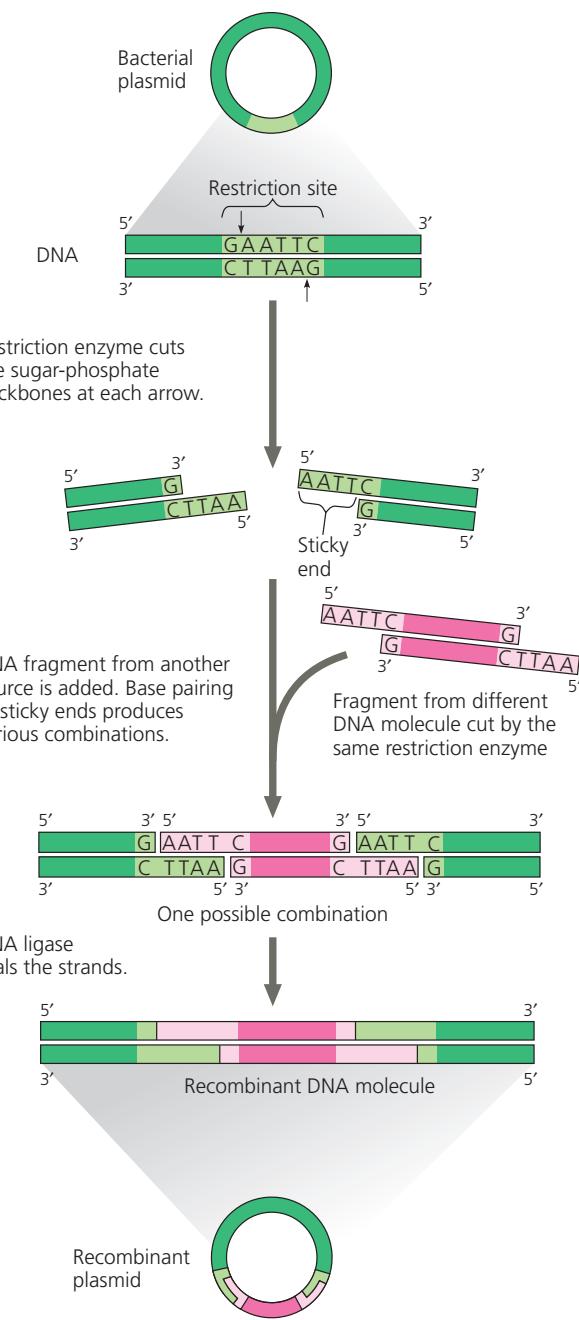
Gene cloning is useful for two basic purposes: to make many copies of, or *amplify*, a particular gene and to produce a protein product (see Figure 20.5). Researchers can isolate copies of a cloned gene from bacteria for use in basic research or to endow another organism with a new metabolic capability, such as pest resistance. For example, a resistance gene present in one crop species might be cloned and transferred into plants of another species. Alternatively, a protein with medical uses, such as human growth hormone, can be harvested in large quantities from cultures of bacteria carrying a cloned gene for the protein. (We’ll describe the techniques for expressing cloned genes later.) Since one gene is only a very small part of the total DNA in a cell, the ability to amplify such rare DNA fragments, by cloning or other means, is crucial for any application involving a single gene.

Using Restriction Enzymes to Make a Recombinant DNA Plasmid

Gene cloning and genetic engineering generally rely on the use of enzymes that cut DNA molecules at a limited number of specific locations. These enzymes, called restriction endonucleases, or **restriction enzymes**, were discovered in the late 1960s by biologists doing basic research on bacteria. Restriction enzymes protect the bacterial cell by cutting up foreign DNA from other organisms or phages (see Concept 19.2).

Hundreds of different restriction enzymes have been identified and isolated. Each restriction enzyme is very specific, recognizing a particular short DNA sequence, or **restriction site**, and cutting both DNA strands at precise points within this restriction site. The DNA of a bacterial cell is protected from the cell’s own restriction enzymes by the addition of methyl groups ($-\text{CH}_3$) to adenines or cytosines within the sequences recognized by the enzymes.

Figure 20.6 shows how restriction enzymes are used during DNA cloning to join DNA fragments together. At the top is a bacterial plasmid (like the one in Figure 20.5) with



▲ Figure 20.6 Using a restriction enzyme and DNA ligase to make a recombinant DNA plasmid. The restriction enzyme in this example (called EcoRI) recognizes a specific six-base-pair sequence, the restriction site, present at one place in this plasmid. The enzyme makes staggered cuts in the sugar-phosphate backbones within this sequence, producing fragments with sticky ends. Foreign DNA fragments with complementary sticky ends can base-pair with the plasmid ends; the ligated product is a recombinant plasmid. (The two original sticky ends can also base-pair, forming a circle; ligation would result in the original non-recombinant plasmid.)

DRAW IT The restriction enzyme HindIII recognizes the sequence 5'-AAGCTT-3', cutting between the two As. Draw the double-stranded sequence before and after the enzyme cuts it.

a single restriction site recognized by a particular restriction enzyme from *E. coli*. As shown in Figure 20.6, most restriction sites are symmetrical. That is, the sequence of nucleotides is the same on both strands when read in the 5' → 3' direction. The most commonly used restriction enzymes recognize sequences containing 4–8 nucleotide pairs. Because any sequence this short usually occurs (by chance) many times in a long DNA molecule, a restriction enzyme will make many cuts in such a DNA molecule, yielding a set of **restriction fragments**. All copies of a given DNA molecule always yield the same set of restriction fragments when exposed to the same restriction enzyme.

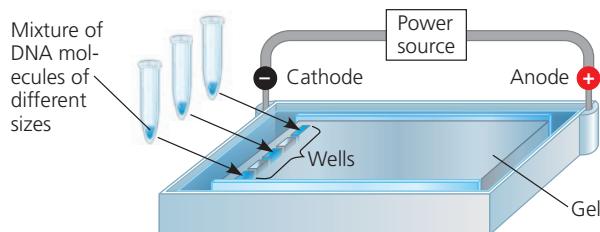
The most useful restriction enzymes cleave the sugar-phosphate backbones in the two DNA strands in a staggered manner, as indicated in Figure 20.6. The resulting double-stranded restriction fragments have at least one single-stranded end, called a **sticky end**. These short extensions can form hydrogen-bonded base pairs with complementary sticky ends on any other DNA molecules cut with the same enzyme. The associations formed in this way are only temporary but can be made permanent by the enzyme **DNA ligase**. This enzyme catalyzes the formation of covalent bonds that close up the sugar-phosphate backbones of DNA strands (see Figure 16.16). You can see at the bottom of Figure 20.6 that the ligase-catalyzed joining of DNA from two different sources produces a stable recombinant DNA molecule, in this example a recombinant plasmid.

In order to check the recombinant plasmid product after it has been copied many times in host cells (see Figure 20.5), a researcher might cut the products again using the same restriction enzyme. This would be expected to yield two kinds of DNA fragments, one the size of the original plasmid and one the size of the inserted DNA fragment, the two starting materials at the top of Figure 20.6. To separate and visualize DNA fragments of different lengths, researchers carry out a technique called **gel electrophoresis**. This technique uses a gel made of a polymer as a molecular sieve to separate out a mixture of nucleic acids (or proteins) on the basis of size, electrical charge, and other physical properties (**Figure 20.7**). Gel electrophoresis is used in conjunction with many different techniques in molecular biology, including DNA sequencing (see Figure 20.3).

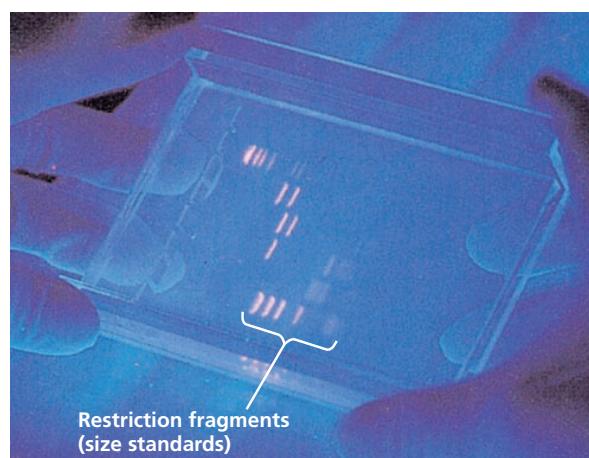
Now that we have discussed the cloning vector in some detail, let's consider the foreign DNA to be inserted. The most common way to obtain many copies of the gene to be cloned is by PCR, described next.

Amplifying DNA: The Polymerase Chain Reaction (PCR) and Its Use in DNA Cloning

Today, most researchers have some information about the sequence of the gene or other DNA segment they want to clone. Using this information, they can start with the entire



(a) Each sample, a mixture of different DNA molecules, is placed in a separate well near one end of a thin slab of agarose gel. The gel is set into a small plastic support and immersed in an aqueous, buffered solution in a tray with electrodes at each end. The current is then turned on, causing the negatively charged DNA molecules to move toward the positive electrode.



(b) Shorter molecules are slowed down less than longer ones, so they move faster through the gel. After the current is turned off, a DNA-binding dye is added that fluoresces pink in UV light. Each pink band corresponds to many thousands of DNA molecules of the same length. The horizontal ladder of bands at the bottom of the gel is a set of restriction fragments of known sizes for comparison with samples of unknown length.

▲ Figure 20.7 Gel electrophoresis. A gel made of a polymer acts as a molecular sieve to separate nucleic acids or proteins differing in size, electrical charge, or other physical properties as they move in an electric field. In the example shown here, DNA molecules are separated by length in a gel made of a polysaccharide called agarose.

collection of genomic DNA from the particular species of interest and obtain many copies of the desired gene by using a technique called the **polymerase chain reaction**, or **PCR**.

Figure 20.8 illustrates the steps in PCR. Within a few hours, this technique can make billions of copies of a specific target DNA segment in a sample, even if that segment makes up less than 0.001% of the total DNA in the sample.

In the PCR procedure, a three-step cycle brings about a chain reaction that produces an exponentially growing population of identical DNA molecules. During each cycle, the reaction mixture is heated to denature (separate) the strands of the double-stranded DNA and then cooled to allow annealing (hydrogen bonding) of short, single-stranded DNA primers complementary to sequences on opposite strands

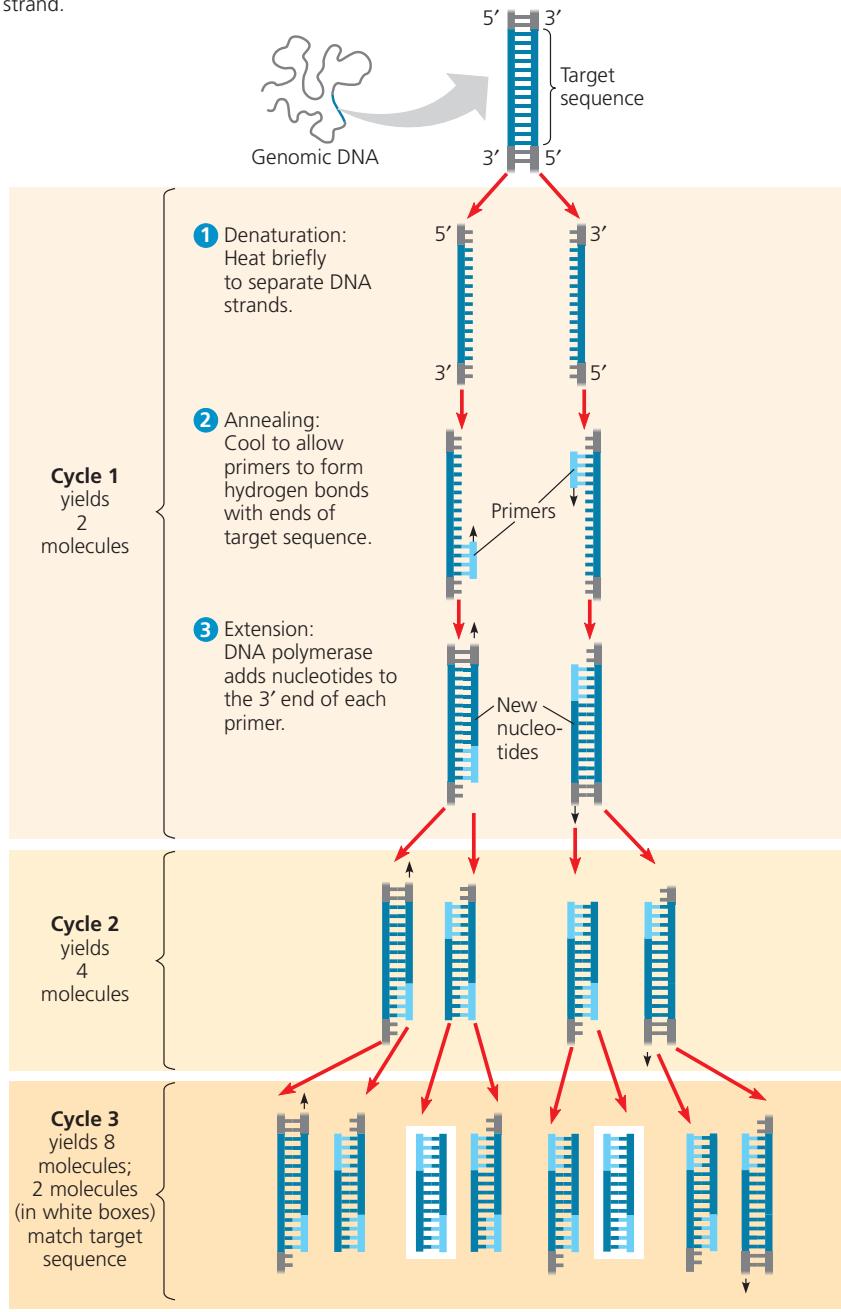
▼ Figure 20.8

Research Method

The Polymerase Chain Reaction (PCR)

Application With PCR, any specific segment (the so-called target sequence) within a DNA sample can be copied many times (amplified), completely *in vitro*.

Technique PCR requires double-stranded DNA containing the target sequence, a heat-resistant DNA polymerase, all four nucleotides, and two 15- to 20-nucleotide DNA strands that serve as primers. One primer is complementary to one end of the target sequence on one strand; the second primer is complementary to the other end of the sequence on the other strand.



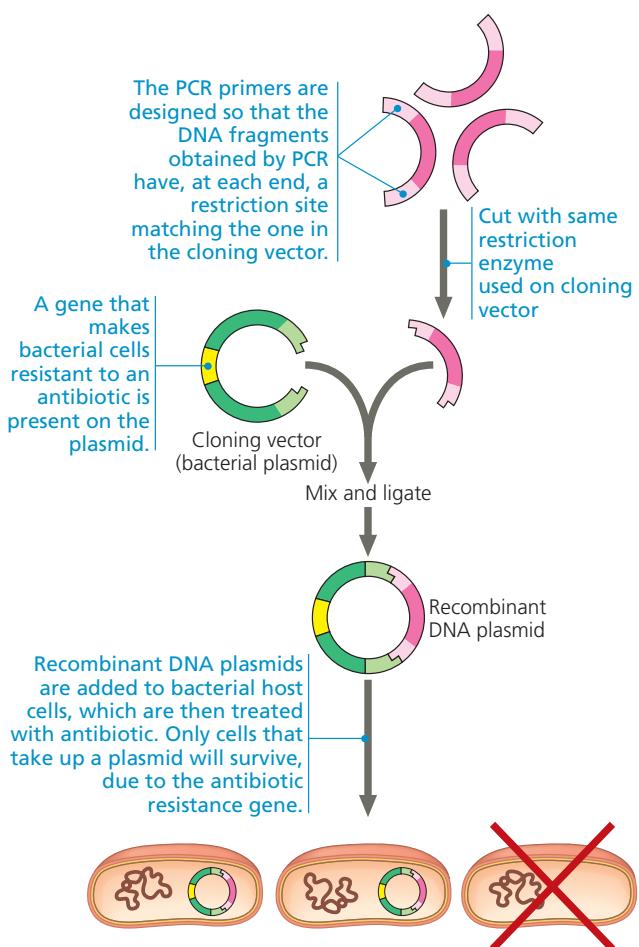
© Pearson Education, Inc.

Results After 3 cycles, two molecules match the target sequence exactly. After 30 more cycles, over 1 billion (10^9) molecules match the target sequence.

at each end of the target sequence; finally, a heat-stable DNA polymerase extends the primers in the 5' → 3' direction. If a standard DNA polymerase were used, the protein would be denatured along with the DNA during the first heating step and would have to be replaced after each cycle. The key to automating PCR was the discovery of an unusual heat-stable DNA polymerase called Taq polymerase, named after the bacterial species from which it was first isolated. This bacterial species, *Thermus aquaticus*, lives in hot springs, and the stability of its DNA polymerase at high temperatures is an evolutionary adaptation that enables the bacterium to survive and reproduce at temperatures up to 95°C.

PCR is speedy and very specific. Only minuscule amounts of DNA need be present in the starting material, and this DNA can be partially degraded, as long as a few molecules contain the complete target sequence. The key to the high specificity is the pair of primers used for each PCR amplification. The primer sequences are chosen so they hybridize *only* to sequences at opposite ends of the target segment, one on the 3' end of each strand. (For high specificity, the primers must be at least 15 or so nucleotides long.) By the end of the third cycle, one-fourth of the molecules are identical to the target segment, with both strands the appropriate length. With each successive cycle, the number of target segment molecules of the correct length doubles, so the number of molecules equals 2^n , where n is the number of cycles. After 30 more cycles, about a billion copies of the target sequence are present!

Despite its speed and specificity, PCR amplification cannot substitute for gene cloning in cells when large amounts of a gene are required. This is because occasional errors during PCR replication limit the number of good copies and the length of DNA fragments that can be copied. Instead, PCR is used to provide a supply of the specific DNA fragment for cloning. PCR primers are synthesized to include a restriction site at each end of the DNA fragment that matches the site in the cloning vector. Then the fragment and vector are cut, allowed to hybridize, and ligated



▲ Figure 20.9 Use of a restriction enzyme and PCR in gene cloning. This figure takes a closer look at the process shown at the top of Figure 20.5. PCR is used to produce multiple copies of the DNA fragment or gene of interest. The ends of the fragments have the same restriction site as the cloning vector. The plasmid and the DNA fragments are cut with the same restriction enzyme, combined so the sticky ends can hybridize, ligated together, and introduced into bacterial host cells. The plasmid also contains an antibiotic resistance gene that allows only cells with a plasmid to survive when the antibiotic is present. Other genetic engineering techniques are used to ensure that cells with nonrecombinant plasmids can be eliminated.

together (**Figure 20.9**). The plasmids from the resulting bacterial clones are sequenced so that clones carrying plasmids with error-free inserts can be selected.

Devised in 1985, PCR has had a major impact on biological research and genetic engineering. PCR has been used to amplify DNA from a wide variety of sources: a 40,000-year-old frozen woolly mammoth (see the photo on the first page of this chapter); fingerprints or tiny amounts of blood, tissue, or semen found at crime scenes; single embryonic cells for rapid prenatal diagnosis of genetic disorders (see Figure 14.19); and cells infected with viruses that are difficult to detect, such as HIV. (To test for HIV, viral genes are amplified.) We'll return to applications of PCR later.

Expressing Cloned Eukaryotic Genes

Once a gene has been cloned in host cells, its protein product can be expressed in large amounts for research or for practical applications, which we'll explore in Concept 20.4. Cloned genes can be expressed in either bacterial or eukaryotic cells; each option has advantages and disadvantages.

Bacterial Expression Systems

Getting a cloned eukaryotic gene to function in bacterial host cells can be difficult because certain aspects of gene expression are different in eukaryotes and bacteria. To overcome differences in promoters and other DNA control sequences, scientists usually employ an **expression vector**, a cloning vector that contains a highly active bacterial promoter just upstream of a restriction site where the eukaryotic gene can be inserted in the correct reading frame. The bacterial host cell will recognize the promoter and proceed to express the foreign gene now linked to that promoter. Such expression vectors allow the synthesis of many eukaryotic proteins in bacterial cells.

Another problem with expressing cloned eukaryotic genes in bacteria is the presence of noncoding regions (introns; see Concept 17.3) in most eukaryotic genes. Introns can make a eukaryotic gene very long and unwieldy, and they prevent correct expression of the gene by bacterial cells, which do not have RNA-splicing machinery. This problem can be surmounted by using a form of the gene that includes only the exons. (This is called *complementary DNA*, or *cDNA*; see Figure 20.11.)

Eukaryotic DNA Cloning and Expression Systems

Molecular biologists can avoid eukaryotic-bacterial incompatibility by using eukaryotic cells such as yeasts, rather than bacteria, as hosts for cloning and expressing eukaryotic genes of interest. Yeasts, single-celled fungi, offer two advantages: They are as easy to grow as bacteria, and they have plasmids, a rarity among eukaryotes. Scientists have even constructed recombinant plasmids that combine yeast and bacterial DNA and can replicate in either type of cell.

Another reason to use eukaryotic host cells for expressing a cloned eukaryotic gene is that many eukaryotic proteins will not function unless they are modified after translation—for example, by the addition of carbohydrate groups (glycosylation) or lipid groups. Bacterial cells cannot carry out these modifications, and if the gene product requiring such processing is from a mammal, even yeast cells may not be able to modify the protein correctly. Several cultured cell types have proved successful as host cells for this purpose, including some mammalian cell lines and an insect cell line that can be infected by a virus (called baculovirus) carrying recombinant DNA.

Besides using vectors, scientists have developed a variety of other methods for introducing recombinant DNA into

eukaryotic cells. In **electroporation**, a brief electrical pulse applied to a solution containing cells creates temporary holes in their plasma membranes, through which DNA can enter. (This technique is now commonly used for bacteria as well.) Alternatively, scientists can inject DNA directly into single eukaryotic cells using microscopically thin needles. Another way to get DNA into plant cells is by using the soil bacterium *Agrobacterium tumefaciens*, as we'll discuss later. Whatever the method, if the introduced DNA is incorporated into a cell's genome by genetic recombination, then it can be expressed by the cell.

To study the function of a particular protein, researchers can introduce different mutant forms of the gene for that protein into eukaryotic cells. The cells express different versions of the protein, and the resulting phenotypes provide information about the normal protein's function. For this purpose, researchers may use a cloning vector with viral DNA sequences that allow the introduced DNA to integrate into a chromosome and be stably expressed. The same approach can be used to express noncoding RNAs in order to study their role as agents of gene regulation in the cell. (Noncoding RNAs are discussed in Concept 18.3.)

Cross-Species Gene Expression and Evolutionary Ancestry

EVOLUTION The ability to express eukaryotic proteins in bacteria (even if the proteins aren't modified properly) is quite remarkable when we consider how different eukaryotic and bacterial cells are. In fact, examples abound of genes that are taken from one species and function perfectly well when transferred into another very different species. These observations underscore the shared evolutionary ancestry of species living today.

One example involves a gene called *Pax-6*, which has been found in animals as diverse as vertebrates and fruit flies. The vertebrate *Pax-6* gene product (the PAX-6 protein) triggers a complex program of gene expression resulting in formation of the vertebrate eye, which has a single lens. Expression of the fly *Pax-6* gene leads to formation of the compound fly eye, which is quite different from the vertebrate eye. When the mouse *Pax-6* gene was cloned and introduced into a fly embryo so that it replaced the fly's own *Pax-6* gene, researchers were surprised to see that the mouse version of the gene led to formation of a compound fly eye (see Figure 50.16). Conversely, when the fly *Pax-6* gene was transferred into a vertebrate embryo—a frog, in this case—a frog eye formed. Although the genetic programs triggered in vertebrates and flies generate very different eyes, the two versions of the *Pax-6* gene can substitute for each other to trigger lens development, evidence of their evolution from a gene in a very ancient common ancestor.

Simpler examples are seen in Figure 17.6, where a firefly gene is expressed in a tobacco plant and a jellyfish gene in

a pig. Because of their ancient evolutionary roots, all living organisms share the same basic mechanisms of gene expression. This commonality is the basis of many recombinant DNA techniques described in this chapter.

CONCEPT CHECK 20.1

1. The restriction site for an enzyme called *Pvu*I is the following sequence:
$$\begin{array}{c} 5'-C\ G\ A\ T\ C\ G-3' \\ | \\ 3'-G\ C\ T\ A\ G\ C-5' \end{array}$$
Staggered cuts are made between the T and C on each strand. What type of bonds are being cleaved?
2. **DRAW IT** One strand of a DNA molecule has the following sequence: 5'-CCTTGACGATCGTTACCG-3'. Draw the other strand. Will *Pvu*I cut this molecule? If so, draw the products.
3. What are some potential difficulties in using plasmid vectors and bacterial host cells to produce large quantities of proteins from cloned eukaryotic genes?
4. **MAKE CONNECTIONS** Compare Figure 20.8 with Figure 16.20. How does replication of DNA ends during PCR proceed without shortening the fragments each time?

For suggested answers, see Appendix A.

CONCEPT 20.2

Biologists use DNA technology to study gene expression and function

To see how a biological system works, scientists seek to understand the functioning of the system's component parts. Analysis of when and where a gene or group of genes is expressed can provide important clues about their function.

Analyzing Gene Expression

Biologists driven to understand the assorted cell types of a multicellular organism, cancer cells, or the developing tissues of an embryo first try to discover which genes are expressed by the cells of interest. The most straightforward way to do this is usually to identify the mRNAs being made. We'll first examine techniques that look for patterns of expression of specific individual genes. Next, we'll explore ways to characterize groups of genes being expressed by cells or tissues of interest. As you will see, all of these procedures depend in some way on base pairing between complementary nucleotide sequences.

Studying the Expression of Single Genes

Suppose we have cloned a gene that we suspect plays an important role in the embryonic development of *Drosophila melanogaster* (the fruit fly). The first thing we might want to know is which embryonic cells express the gene—in other

words, where in the embryo is the corresponding mRNA found? We can detect the mRNA by nucleic acid hybridization with molecules of complementary sequence that we can follow in some way. The complementary molecule, a short, single-stranded nucleic acid that can be either RNA or DNA, is called a **nucleic acid probe**. Using our cloned gene as a template, we can synthesize a probe complementary to the mRNA. For example, if part of the sequence on the mRNA were



then we would synthesize this single-stranded DNA probe:



Each probe molecule is labeled during synthesis with a fluorescent tag so we can follow it. A solution containing probe molecules is applied to *Drosophila* embryos, allowing the probe to hybridize specifically with any complementary sequences on the many mRNAs in embryonic cells that are transcribing the gene. Because this technique allows us to see the mRNA in place (or *in situ*, in Latin) in the intact organism, this technique is called ***in situ* hybridization**. Different probes can be labeled with different fluorescent dyes, sometimes with strikingly beautiful results (**Figure 20.10**).

Other mRNA detection techniques may be preferable for comparing the amounts of a specific mRNA in several samples at the same time—for example, in different cell types or in embryos at different stages of development. One method that is widely used is called the **reverse transcriptase-polymerase chain reaction**, or RT-PCR.

RT-PCR begins by turning sample sets of mRNAs into double-stranded DNAs with the corresponding sequences. First, the enzyme reverse transcriptase (from a retrovirus; see Figure 19.8) is used *in vitro* to make a single-stranded DNA *reverse transcript* of each mRNA molecule (**Figure 20.11**). Recall that the 3' end of the mRNA has a stretch of adenine (A) nucleotides called a poly-A tail. This allows use of a short complementary strand of thymine deoxyribonucleotides (poly-dT) as a primer for synthesis of this DNA strand. Following enzymatic degradation of the mRNA, a second DNA strand, complementary to the first, is synthesized by DNA polymerase. The resulting double-stranded DNA is called **complementary DNA (cDNA)**. (Made from mRNA, cDNA lacks introns and can be used for protein expression in bacteria, as mentioned earlier.) To analyze the timing of expression of the *Drosophila* gene of interest, for example, we would first isolate all the mRNAs from different stages of *Drosophila* embryos and make cDNA from each stage (**Figure 20.12**). Then we'd use PCR to find any cDNA derived from the gene of interest.

As you will recall from Figure 20.8, PCR is a way of rapidly making many copies of one specific stretch of double-stranded DNA, using primers that hybridize to the opposite ends of the segment of interest. In our case, we would add primers corresponding to a segment of our *Drosophila* gene,

The yellow probe hybridizes with mRNAs in cells that are expressing the *wingless (wg)* gene, which encodes a secreted signaling protein.



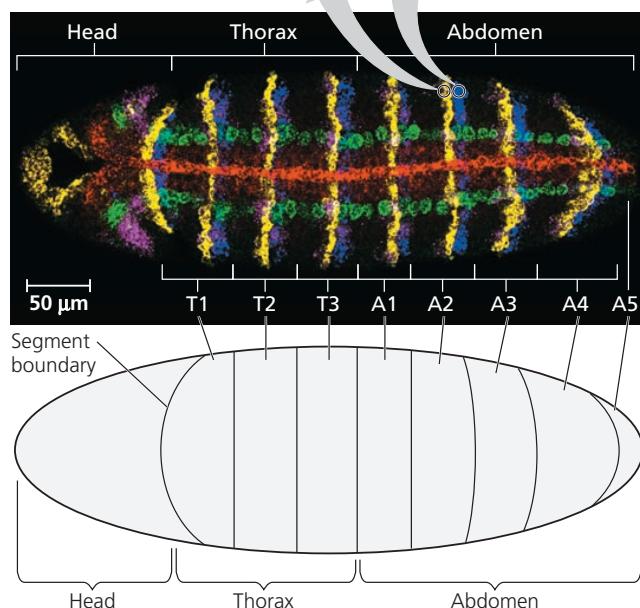
wg mRNA

The blue probe hybridizes with mRNAs in cells that are expressing the *engrailed (en)* gene, which encodes a transcription factor.



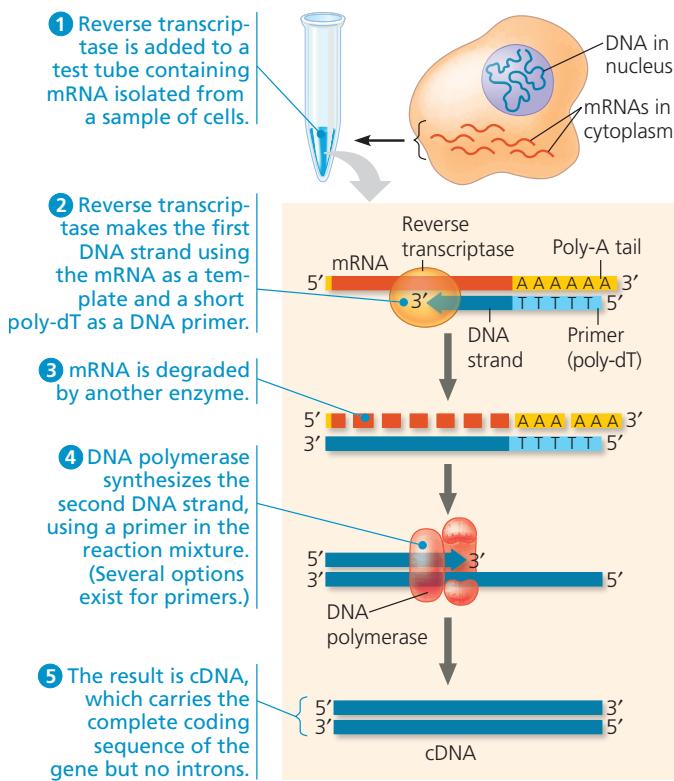
en mRNA

Cells expressing the *wg* gene Cells expressing the *en* gene



▲ Figure 20.10 Determining where genes are expressed by *in situ* hybridization analysis. A *Drosophila* embryo was incubated in a solution containing probes for five different mRNAs, each probe labeled with a different fluorescently colored tag. The embryo was then viewed from the belly (ventral) side using fluorescence microscopy; the resulting fluorescent micrograph is shown in the middle, above. Each color marks where a specific gene is expressed as mRNA. The arrows from the groups of yellow and blue cells above the micrograph show a magnified view of nucleic acid hybridization of the appropriately colored probe to the mRNA. Yellow cells (expressing the *wg* gene) interact with blue cells (expressing the *en* gene); their interaction helps establish the pattern in a body segment. The diagram at the bottom clarifies the eight segments visible in this view.

using the cDNA from each embryonic stage as a template for PCR amplification in separate samples. When the products are analyzed on a gel, only samples that originally contained mRNA from the gene of interest will show bands containing copies of the amplified region. A recent enhancement involves using a fluorescent dye that fluoresces only when bound to a double-stranded PCR product. The newer PCR machines can detect the light and measure the PCR product, thus avoiding the need for electrophoresis while also



▲ Figure 20.11 Making complementary DNA (cDNA) from eukaryotic genes. Complementary DNA is made *in vitro* using mRNA as a template for the first strand. The mRNA contains only exons, so the resulting double-stranded cDNA carries the continuous coding sequence of the gene. Only one mRNA is shown here, but the final collection of cDNAs would reflect all the mRNAs present in the cell sample. Figure 20.12 shows how the cDNA of interest is identified.

providing quantitative data, a distinct advantage. RT-PCR can also be carried out with mRNAs collected from different tissues at one time to discover which tissue is producing a specific mRNA.

In the **Scientific Skills Exercise**, you can work with data from an experiment that analyzed expression of a gene involved in paw formation in the mouse. The study investigated mRNA expression using two techniques. One of these methods was qualitative (*in situ* hybridization), whereas the other approach was quantitative (PCR).

Studying the Expression of Interacting Groups of Genes

A major goal of biologists is to learn how genes act together to produce and maintain a functioning organism. Now that the genomes of a number of species have been sequenced, it is possible to study the expression of large groups of genes—the so-called *systems approach*. Researchers use what is known about the whole genome to investigate which genes are transcribed in different tissues or at different stages of development. One aim is to identify networks of gene expression across an entire genome.

Research Method

RT-PCR Analysis of the Expression of Single Genes

Application RT-PCR uses the enzyme reverse transcriptase (RT) in combination with PCR and gel electrophoresis. RT-PCR can be used to compare gene expression between samples—for instance, in different embryonic stages, in different tissues, or in the same type of cell under different conditions.

Technique In this example, samples containing mRNAs from six embryonic stages of *Drosophila* were analyzed for a specific mRNA as shown below. (In steps 1 and 2, the mRNA from only one stage is shown.)

1 **cDNA synthesis** is carried out by incubating the mRNAs with reverse transcriptase and other necessary components.

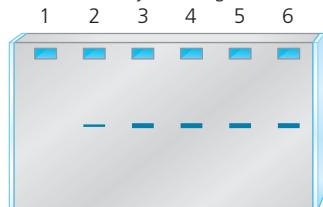
2 **PCR amplification** of the sample is performed using primers specific to the *Drosophila* gene of interest.

3 **Gel electrophoresis** will reveal amplified DNA products only in samples that contained mRNA transcribed from the specific *Drosophila* gene.

mRNAs

cDNAs

Embryonic stages



Results The mRNA for this gene first is expressed at stage 2 and continues to be expressed through stage 6. The size of the amplified fragment (shown by its position on the gel) depends on the distance between the primers that were used (not on the size of the mRNA).

Genome-wide expression studies can be carried out using **DNA microarray assays**. A DNA microarray consists of tiny amounts of a large number of single-stranded DNA fragments representing different genes fixed to a glass slide in a tightly spaced array, or grid. (The microarray is also called a *DNA chip* by analogy to a computer chip.) Ideally, these fragments represent all the genes of an organism.

The basic strategy in such studies is to isolate the mRNAs made in a cell of interest and use these molecules as templates for making the corresponding cDNAs by reverse transcription. In microarray assays, these cDNAs are labeled with fluorescent molecules and then allowed to hybridize to a DNA microarray. Most often, the cDNAs from two

SCIENTIFIC SKILLS EXERCISE

Analyzing Quantitative and Spatial Gene Expression Data

How Is a Particular *Hox* Gene Regulated During Paw Development? *Hox* genes code for transcription factor proteins, which in turn control sets of genes important for animal development (see Concept 21.6 for more information on *Hox* genes). One group of *Hox* genes, the *Hoxd* genes, plays a role in establishing the pattern of the different digits (fingers and toes) at the end of a limb. Unlike the *mPGES-1* gene mentioned in the Chapter 18 Scientific Skills Exercise, *Hox* genes have very large, complicated regulatory regions, including control elements that may be hundreds of kilobases (kb; thousands of nucleotides) away from the gene.

In cases like this, how do biologists locate the DNA segments that contain important elements? They begin by removing (deleting) large segments of DNA and studying the effect on gene expression. In this exercise, you'll compare data from two different but complementary approaches that look at the expression of a specific *Hoxd* gene (*Hoxd13*). One approach quantifies overall expression; the other approach is less quantitative but gives important spatial localization information.

How the Experiment Was Done Researchers interested in the regulation of *Hoxd13* gene expression genetically engineered a set of mice (*transgenic* mice) that had different segments of

DNA deleted upstream of the gene. They then compared levels and patterns of *Hoxd13* gene expression in the developing paws of 12.5-day-old transgenic mouse embryos (with the DNA deletions) with those seen in wild-type mouse embryos of the same age.

They used two different approaches: In some mice, they extracted the mRNA from the embryonic paws and quantified the overall level of *Hoxd13* mRNA in the whole paw using quantitative RT-PCR. In another set of the same transgenic mice, they used *in situ* hybridization to pinpoint exactly where in the paws the *Hoxd13* gene was expressed as mRNA. The particular technique that was used causes the *Hoxd13* mRNA to appear blue, or black for the highest mRNA levels.

Data from the Experiment The top-most diagram (upper right) depicts the very large regulatory region upstream of the *Hoxd13* gene. The area between the slashes represents the long stretch of DNA located between the promoter and the regulatory region.

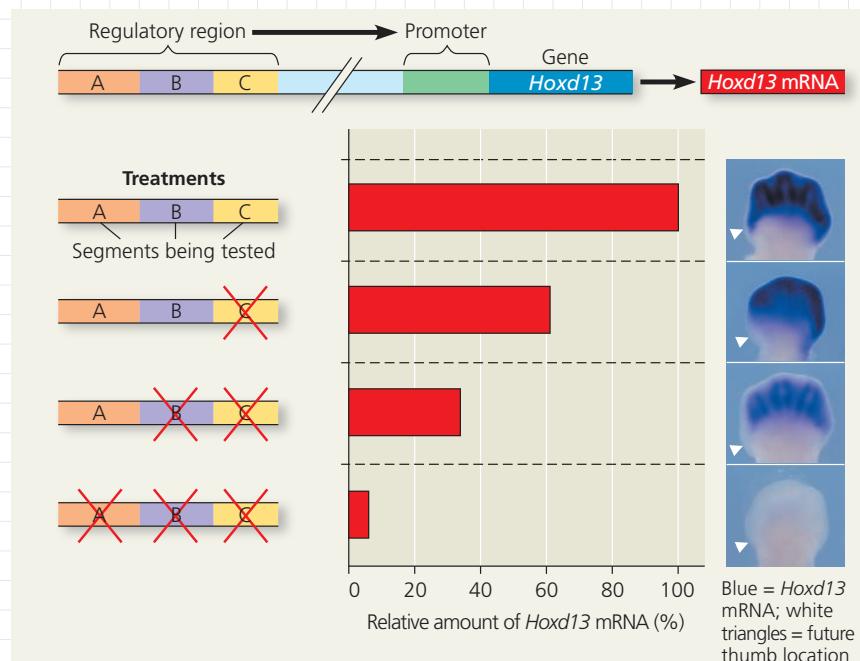
The diagrams to the left of the bar graph show, first, the intact DNA (830 kb) and, next, the three altered DNA sequences. (Each is called a “deletion” because a particular section of DNA has been deleted from it.) A red X indicates the segment (A, B, and/or C) that was deleted in each experimental treatment.

The horizontal bar graph shows the amount of *Hoxd13* mRNA that was present in the digit-formation zone of each mutant 12.5-day-old embryo paw relative to the amount that was in the digit-formation zone of the mouse that had the intact regulatory region (top bar = 100%).

The images on the right are fluorescent micrographs of the embryo paws showing the location of the *Hoxd13* mRNA (stain appears blue or black). The white triangles show the location where the thumb will form.

Interpret the Data

1. The researchers hypothesized that all three regulatory segments (A, B, and C) were required for full expression of the *Hoxd13* gene. By measuring the amount of *Hoxd13* mRNA in the embryo paw zones where digits develop, they could measure the effect of the

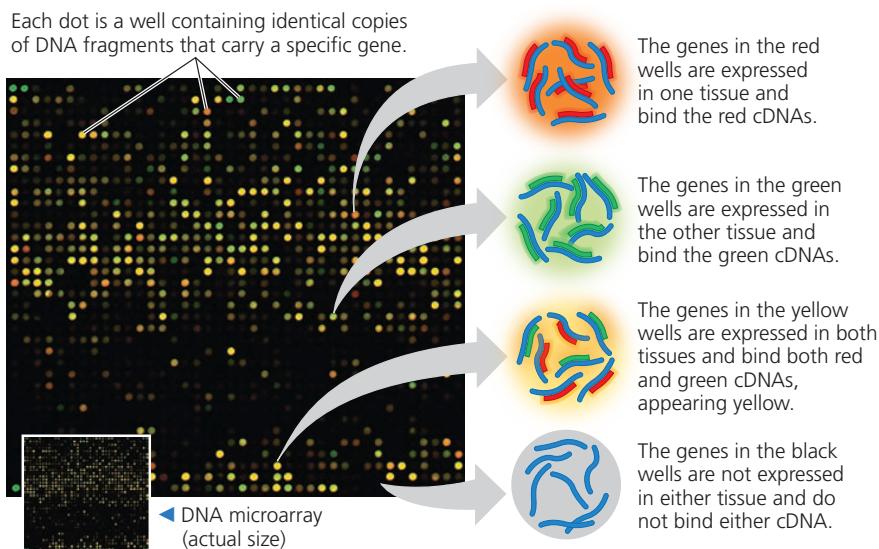


regulatory segments singly and in combination. Refer to the graph to answer these questions, noting that the segments being tested are shown on the vertical axis and the relative amount of *Hoxd13* mRNA is shown on the horizontal axis. (a) Which of the four treatments was used as a control for the experiment? (b) The hypothesis is that all three segments together are required for highest expression of the *Hoxd13* gene. Is this supported by the results? Explain your answer.

2. (a) What is the effect on the amount of *Hoxd13* mRNA when segments B and C are both deleted, compared with the control? (b) Is this effect visible in the blue-stained regions of the *in situ* hybridizations? How would you describe the spatial pattern of gene expression in the embryo paws that lack segments B and C? (You'll need to look carefully at different regions of each paw and how they differ.)
3. (a) What is the effect on the amount of *Hoxd13* mRNA when just segment C is deleted, compared with the control? (b) Is this effect visible in the *in situ* hybridizations? How would you describe the spatial pattern of gene expression in embryo paws that lack just segment C, compared with the control and with the paws that lack segments B and C?
4. If the researchers had only measured the amount of *Hoxd13* mRNA and not done the *in situ* hybridizations, what important information about the role of the regulatory segments in *Hoxd13* gene expression during paw development would have been missed? Conversely, if the researchers had only done the *in situ* hybridizations, what information would have been inaccessible?

A version of this Scientific Skills Exercise can be assigned in MasteringBiology.

Data from T. Montavon et al., A regulatory archipelago controls *Hox* genes transcription in digits, *Cell* 147:1132–1145 (2011).



▲ Figure 20.13 DNA microarray assay of gene expression levels. Researchers synthesized two sets of cDNAs, fluorescently labeled red or green, from mRNAs from two different human tissues. These cDNAs were hybridized with a microarray containing 5,760 human genes (about 25% of human genes), resulting in the pattern shown here. The intensity of fluorescence at each spot is a measure of the relative expression in the two samples of the gene represented by that spot: Red indicates expression in one sample, green in the other, yellow in both, and black in neither.

samples (for example, different tissues) are labeled with molecules that emit different colors and tested on the same microarray. **Figure 20.13** shows the result of such an experiment, identifying the subsets of genes in the genome that are being expressed in one tissue compared with another. DNA technology makes such studies possible; with automation, they are easily performed on a large scale. Scientists can now measure the expression of thousands of genes at one time.

Alternatively, with the advent of rapid, inexpensive DNA sequencing methods, researchers can now afford to simply sequence the cDNA samples from different tissues or different embryonic stages in order to discover which genes are expressed. This straightforward method is called *RNA sequencing* or *RNA-seq* (pronounced “RNA-seek”), even though it is the cDNA that is actually sequenced. As the price of DNA sequencing plummets, this method is becoming more widely used for many applications. In most cases, however, expression of individual genes would still need to be confirmed by RT-PCR.

By uncovering gene interactions and providing clues to gene function, DNA microarray assays and RNA-seq may contribute to a better understanding of diseases and suggest new diagnostic techniques or therapies. For instance, comparing patterns of gene expression in breast cancer tumors and noncancerous breast tissue has already resulted in more informed and effective treatment protocols (see Figure 18.27). Ultimately, information from these methods should provide a grander view of how ensembles of genes interact to form an organism and maintain its vital systems.

Determining Gene Function

Once they identify a gene of interest, how do scientists determine its function? A gene's sequence can be compared with sequences in other species. If the function of a similar gene in another species is known, one might suspect that the gene product in question performs a comparable task. Data about the location and timing of gene expression may reinforce the suggested function. To obtain stronger evidence, one approach is to disable the gene and then observe the consequences in the cell or organism. In one such technique, called *in vitro mutagenesis*, specific mutations are introduced into a cloned gene, and the mutated gene is returned to a cell in such a way that it disables (“knocks out”) the normal cellular copies of the same gene. If the introduced mutations alter or destroy the function of the gene product, the phenotype of the mutant cell may help reveal the function

of the missing normal protein. Using molecular and genetic techniques worked out in the 1980s, researchers can generate mice with any given gene disabled, in order to study the role of that gene in development and in the adult. Mario Capecchi, Martin Evans, and Oliver Smithies received the Nobel Prize in 2007 for developing this technique.

A newer method for silencing expression of selected genes exploits the phenomenon of **RNA interference (RNAi)**, described in Chapter 18. This experimental approach uses synthetic double-stranded RNA molecules matching the sequence of a particular gene to trigger breakdown of the gene's messenger RNA or to block its translation. In organisms such as the nematode and the fruit fly, RNAi has already proved valuable for analyzing the functions of genes on a large scale.

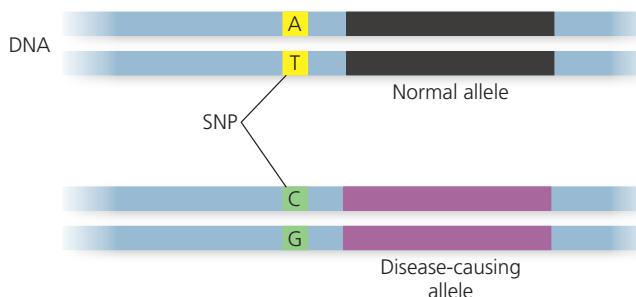
In humans, ethical considerations prohibit knocking out genes to determine their functions. An alternative approach is to analyze the genomes of large numbers of people with a certain phenotypic condition or disease, such as heart disease or diabetes, to try to find differences they all share compared with people without that condition. The assumption is that these differences may be associated with one or more malfunctioning genes, thus in a sense being naturally occurring gene knockouts. Such large-scale analyses, called **genome-wide association studies**, do not require complete sequencing of all the genomes in the two groups. Instead, researchers test for *genetic markers*, DNA sequences that vary in the population. In a gene, such sequence variation is the basis of different alleles, as we have seen for

sickle-cell disease (see Figure 17.25). And just like the coding sequences of genes, noncoding DNA at a specific locus on a chromosome may exhibit small nucleotide differences among individuals. Variations in coding or noncoding DNA sequence among a population are called polymorphisms (from the Greek for “many forms”).

Among the most useful genetic markers in tracking down genes that contribute to diseases and disorders are single base-pair variations in the genomes of the human population. A single base-pair site where variation is found in at least 1% of the population is called a **single nucleotide polymorphism (SNP)**, pronounced “snip”). A few million SNPs occur in the human genome, about once in 100–300 base pairs of both coding and noncoding DNA sequences. It isn’t necessary to sequence the DNA of multiple individuals to find SNPs; today they can be detected by very sensitive microarray analysis or by PCR.

Once a SNP is identified that is found in all affected people, researchers focus on that region and sequence it. In nearly all cases, the SNP itself does not contribute directly to the disease in question by altering the encoded protein; in fact, most SNPs are in noncoding regions. Instead, if the SNP and a disease-causing allele are close enough, scientists can take advantage of the fact that crossing over between the marker and the gene is very unlikely during gamete formation. Therefore, the marker and gene will almost always be inherited together, even though the marker is not part of the gene (Figure 20.14). SNPs have been found that correlate with diabetes, heart disease, and several types of cancer, and the search is on for genes that might be involved.

The experimental approaches you have learned about thus far focused on working with molecules, mainly DNA and proteins. In a parallel line of inquiry, biologists have



▲ Figure 20.14 Single nucleotide polymorphisms (SNPs) as genetic markers for disease-causing alleles. This diagram depicts homologous segments of DNA from two groups of individuals, those in one group having a particular disease or condition with a genetic basis. Unaffected people have an A/T pair at a particular SNP locus, while affected people have a C/G pair at that locus. A SNP that varies in this way is likely to be closely linked to one or more alleles of genes that contribute to the disease in question.

MAKE CONNECTIONS What does it mean for a SNP to be “closely linked” to a disease-causing allele, and how does this allow the SNP to be used as a genetic marker? (See Concept 15.3.)

been developing powerful techniques for cloning whole multicellular organisms. One aim of this work is to obtain special types of cells, called stem cells, that can give rise to all types of tissues. Being able to manipulate stem cells would allow scientists to use the DNA-based methods previously discussed to alter stem cells for the treatment of diseases. Methods involving the cloning of organisms and production of stem cells are the subject of the next section.

CONCEPT CHECK 20.2

1. Describe the role of complementary base pairing during RT-PCR and DNA microarray analysis.
2. **WHAT IF?** Consider the microarray in Figure 20.13. If a sample from normal tissue is labeled with a green fluorescent dye, and a sample from cancerous tissue is labeled red, what color spots would represent genes you would be interested in if you were studying cancer? Explain.

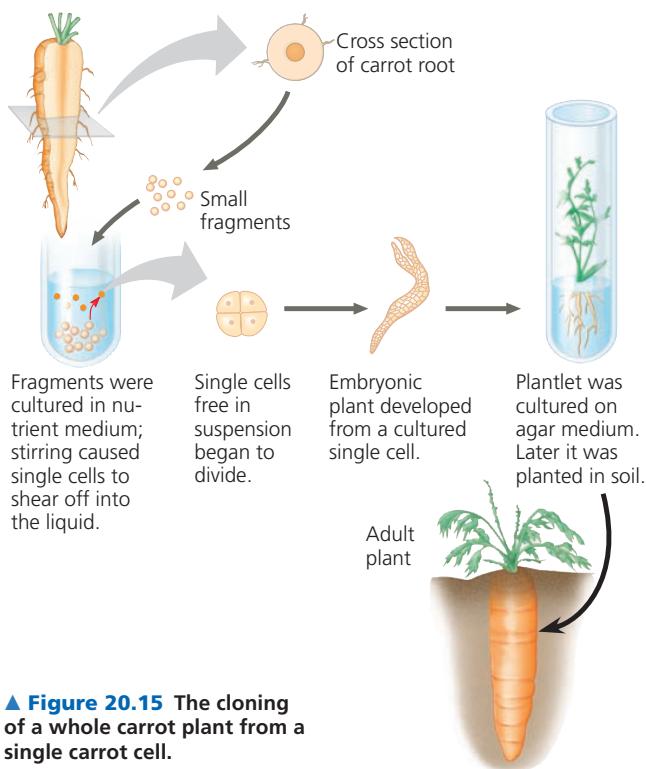
For suggested answers, see Appendix A.

CONCEPT 20.3

Cloned organisms and stem cells are useful for basic research and other applications

Along with advances in DNA technology, scientists have been developing and refining methods for cloning whole multicellular organisms from single cells. In this context, cloning produces one or more organisms that are genetically identical to the “parent” that donated the single cell. This is often called *organismal cloning* to differentiate it from gene cloning and, more significantly, from cell cloning—the division of an asexually reproducing cell into a group of genetically identical cells. (The common theme is that the product is genetically identical to the parent. In fact, the word *clone* comes from the Greek *klon*, meaning “twig.”) The current interest in organismal cloning arises primarily from its ability to generate stem cells. A **stem cell** is a relatively unspecialized cell that can both reproduce itself indefinitely and, under appropriate conditions, differentiate into specialized cells of one or more types. Stem cells have great potential for regenerating damaged tissues.

The cloning of plants and animals was first attempted over 50 years ago in experiments designed to answer basic biological questions. For example, researchers wondered if all the cells of an organism have the same genes or whether cells lose genes during the process of differentiation (see Concept 18.4). One way to answer this question is to see whether a differentiated cell can generate a whole organism—in other words, whether cloning an organism is possible. Let’s discuss these early experiments before we consider more recent progress in organismal cloning and procedures for producing stem cells.



▲ Figure 20.15 The cloning of a whole carrot plant from a single carrot cell.

Cloning Plants: Single-Cell Cultures

The successful cloning of whole plants from single differentiated cells was accomplished during the 1950s by F. C. Steward and his students at Cornell University, who worked with carrot plants (Figure 20.15). They found that differentiated cells taken from the root (the carrot) and incubated in culture medium could grow into normal adult plants, each genetically identical to the parent plant. These results showed that differentiation does not necessarily involve irreversible changes in the DNA. In plants, at least, mature cells can “dedifferentiate” and then give rise to all the specialized cell types of the organism. Any cell with this potential is said to be **totipotent**.

Plant cloning is used extensively in agriculture. For plants such as orchids, cloning is the only commercially practical means of reproducing plants. In other cases, cloning has been used to reproduce a plant with valuable characteristics, such as resistance to plant pathogens. In fact, you yourself may be a plant cloner: If you have ever grown a new plant from a cutting, you have practiced cloning!

Cloning Animals: Nuclear Transplantation

Differentiated cells from animals generally do not divide in culture, much less develop into the multiple cell types of a new organism. Therefore, early researchers had to use a different approach to answer the question: Are differentiated animal cells totipotent? Their approach was to remove the nucleus of an unfertilized or fertilized egg and replace it

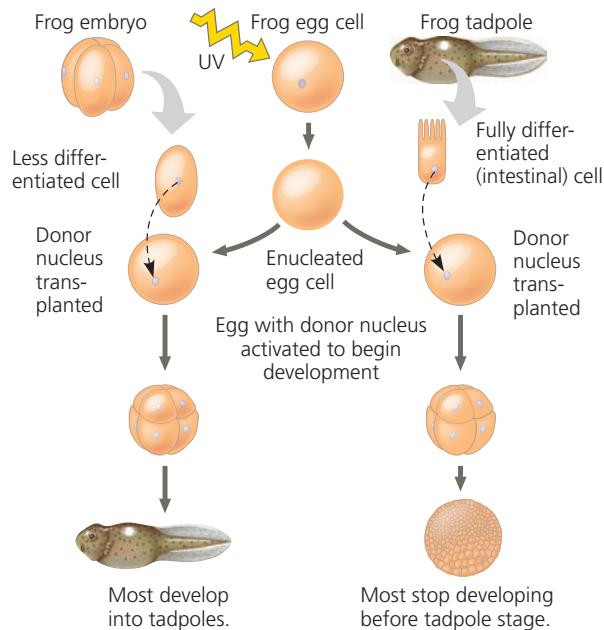
with the nucleus of a differentiated cell, a procedure called *nuclear transplantation*. If the nucleus from the differentiated donor cell retains its full genetic capability, then it should be able to direct development of the recipient cell into all the tissues and organs of an organism.

Such experiments were conducted on one species of frog (*Rana pipiens*) by Robert Briggs and Thomas King in the 1950s and on another (*Xenopus laevis*) by John Gurdon in the 1970s (Figure 20.16). These researchers transplanted a

▼ Figure 20.16 Inquiry

Can the nucleus from a differentiated animal cell direct development of an organism?

Experiment John Gurdon and colleagues at Oxford University, in England, destroyed the nuclei of frog (*Xenopus laevis*) eggs by exposing the eggs to ultraviolet light. They then transplanted nuclei from cells of frog embryos and tadpoles into the enucleated eggs.



Results When the transplanted nuclei came from an early embryo, the cells of which are relatively undifferentiated, most of the recipient eggs developed into tadpoles. But when the nuclei came from the fully differentiated intestinal cells of a tadpole, fewer than 2% of the eggs developed into normal tadpoles, and most of the embryos stopped developing at a much earlier stage.

Conclusion The nucleus from a differentiated frog cell can direct development of a tadpole. However, its ability to do so decreases as the donor cell becomes more differentiated, presumably because of changes in the nucleus.

Source: J. B. Gurdon et al., The developmental capacity of nuclei transplanted from keratinized cells of adult frogs, *Journal of Embryology and Experimental Morphology* 34:93–112 (1975).

WHAT IF? If each cell in a four-cell embryo was already so specialized that it was not totipotent, what results would you predict for the experiment on the left side of the figure?

nucleus from an embryonic or tadpole cell into an enucleated (nucleus-lacking) egg of the same species. In Gurdon's experiments, the transplanted nucleus was often able to support normal development of the egg into a tadpole. However, he found that the potential of a transplanted nucleus to direct normal development was inversely related to the age of the donor: the older the donor nucleus, the lower the percentage of normally developing tadpoles (see Figure 20.16).

From these results, Gurdon concluded that something in the nucleus *does* change as animal cells differentiate. In frogs and most other animals, nuclear potential tends to be restricted more and more as embryonic development and cell differentiation progress. These were foundational experiments that ultimately led to stem cell technology, and Gurdon received the 2012 Nobel Prize in Medicine for this work.

Reproductive Cloning of Mammals

In addition to cloning frogs, researchers have long been able to clone mammals by transplanting nuclei or cells from a variety of early embryos into enucleated eggs. But it was not known whether a nucleus from a fully differentiated cell could be reprogrammed successfully to act as a donor nucleus. In 1997, however, researchers at the Roslin Institute in Scotland captured newspaper headlines when they announced the birth of Dolly, a lamb cloned from an adult sheep by nuclear transplantation from a differentiated cell (**Figure 20.17**). These researchers achieved the necessary dedifferentiation of donor nuclei by culturing mammary cells in nutrient-poor medium. They then fused these cells with enucleated sheep eggs. The resulting diploid cells divided to form early embryos, which were implanted into surrogate mothers. Out of several hundred embryos, one successfully completed normal development, and Dolly was born.

Later analyses showed that Dolly's chromosomal DNA was indeed identical to that of the nucleus donor. (Her mitochondrial DNA came from the egg donor, as expected.) At the age of 6, Dolly suffered complications from a lung disease usually seen only in much older sheep and was euthanized. Dolly's premature death, as well as an arthritic condition, led to speculation that her cells were in some way not quite as healthy as those of a normal sheep, possibly reflecting incomplete reprogramming of the original transplanted nucleus.

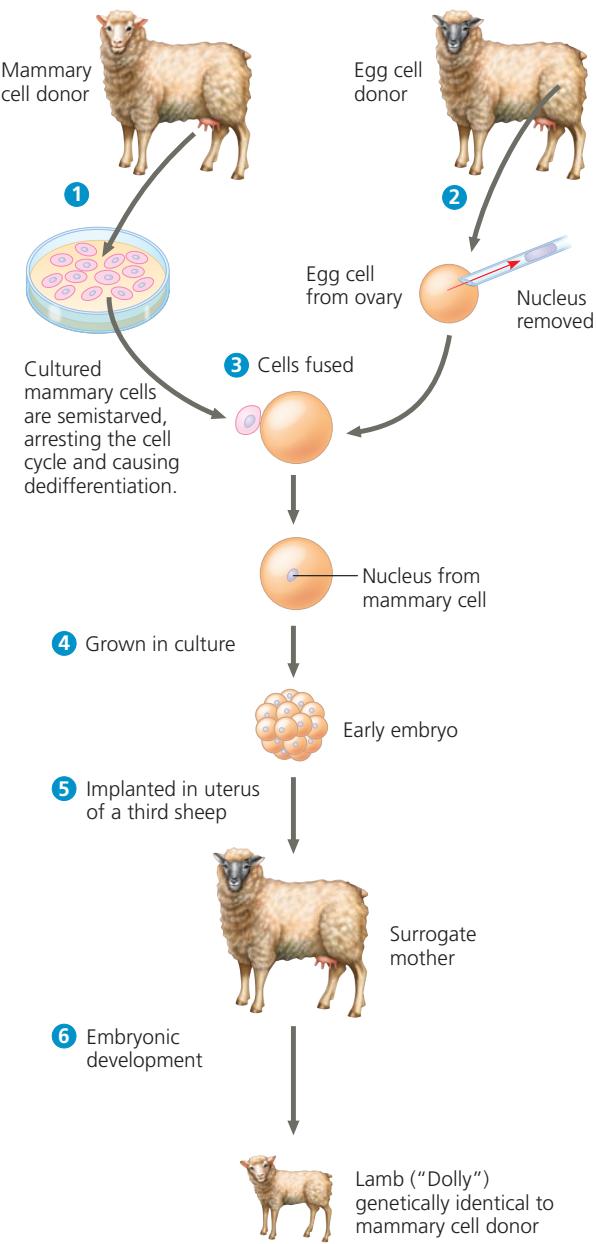
Since that time, researchers have cloned numerous other mammals, including mice, cats, cows, horses, pigs, dogs, and monkeys. In most cases, their goal has been the production of new individuals; this is known as *reproductive cloning*. We have already learned a lot from such experiments. For example, cloned animals of the same species do *not* always look or behave identically. In a herd of cows cloned from the same line of cultured cells, certain cows are dominant in behavior and others are more submissive. Another example of nonidentity in clones is the first cloned cat, named CC for

▼ Figure 20.17 Research Method

Reproductive Cloning of a Mammal by Nuclear Transplantation

Application This method produces cloned animals with nuclear genes identical to those of the animal supplying the nucleus.

Technique Shown here is the procedure used to produce Dolly, the first reported case of a mammal cloned using the nucleus of a differentiated cell.



Results The genetic makeup of the cloned animal is identical to that of the animal supplying the nucleus but differs from that of the egg donor and surrogate mother. (The latter two are "Scottish blackface" sheep, with dark faces.)



▲ Figure 20.18 CC, the first cloned cat, and her single parent.

Rainbow (left) donated the nucleus in a cloning procedure that resulted in CC (right). However, the two cats are not identical: Rainbow is a classic calico cat with orange patches on her fur and has a “reserved personality,” while CC has a gray and white coat and is more playful.

Carbon Copy (**Figure 20.18**). She has a calico coat, like her single female parent, but the color and pattern are different because of random X chromosome inactivation, which is a normal occurrence during embryonic development (see Figure 15.8). And identical human twins, which are naturally occurring “clones,” are always slightly different. Clearly, environmental influences and random phenomena play a significant role during development.

Faulty Gene Regulation in Cloned Animals

In most nuclear transplantation studies thus far, only a small percentage of cloned embryos develop normally to birth. And like Dolly, many cloned animals exhibit defects. Cloned mice, for instance, are prone to obesity, pneumonia, liver failure, and premature death. Scientists assert that even cloned animals that appear normal are likely to have subtle defects.

In recent years, we have begun to uncover some reasons for the low efficiency of cloning and the high incidence of abnormalities. In the nuclei of fully differentiated cells, a small subset of genes is turned on and expression of the rest is repressed. This regulation often is the result of epigenetic changes in chromatin, such as acetylation of histones or methylation of DNA (see Figure 18.7). During the nuclear transfer procedure, many of these changes must be reversed in the later-stage nucleus from a donor animal for genes to be expressed or repressed appropriately in early stages of development. Researchers have found that the DNA in cells from cloned embryos, like that of differentiated cells, often has more methyl groups than does the DNA in equivalent cells from normal embryos of the same species. This finding suggests that the reprogramming of donor nuclei requires

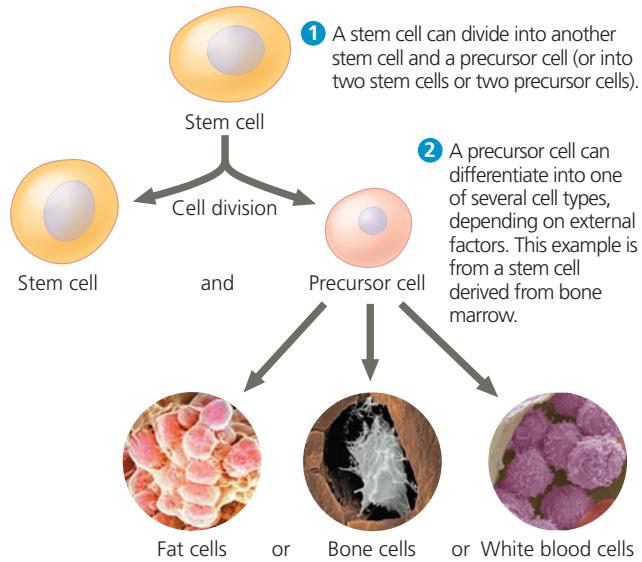
more accurate and complete chromatin restructuring than occurs during cloning procedures. Because DNA methylation helps regulate gene expression, misplaced or extra methyl groups in the DNA of donor nuclei may interfere with the pattern of gene expression necessary for normal embryonic development. In fact, the success of a cloning attempt may depend in large part on whether or not the chromatin in the donor nucleus can be artificially modified to resemble that of a newly fertilized egg.

Stem Cells of Animals

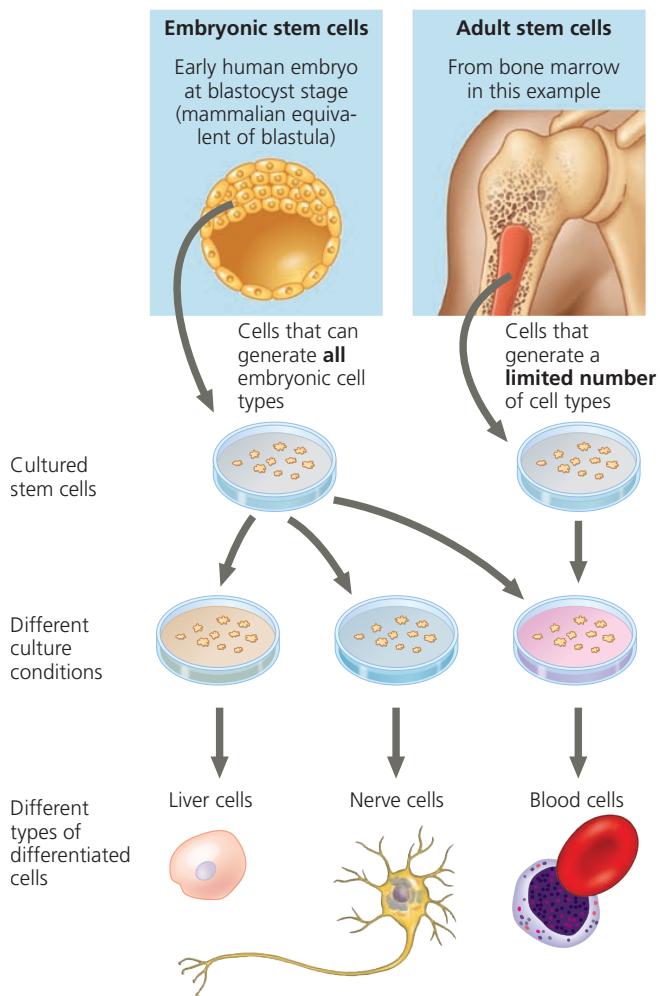
Progress in cloning mammalian embryos, including primates, has heightened speculation about the cloning of humans, which has not yet been achieved. The main reason researchers are trying to clone human embryos is not for reproduction, but for the production of stem cells to treat human diseases. Recall that a stem cell is a relatively unspecialized cell that can both reproduce itself indefinitely and, under appropriate conditions, differentiate into specialized cells of one or more types (**Figure 20.19**). Thus, stem cells are able both to replenish their own population and to generate cells that travel down specific differentiation pathways.

Embryonic and Adult Stem Cells

Many early animal embryos contain stem cells capable of giving rise to differentiated cells of any type. Stem cells can be isolated from early embryos at a stage called the blastula stage or its human equivalent, the blastocyst stage. In culture, these *embryonic stem (ES) cells* reproduce indefinitely; and depending on culture conditions, they can be made to differentiate



▲ Figure 20.19 How stem cells maintain their own population and generate differentiated cells.



▲ Figure 20.20 Working with stem cells. Animal stem cells, which can be isolated from early embryos or adult tissues and grown in culture, are self-perpetuating, relatively undifferentiated cells. Embryonic stem cells are easier to grow than adult stem cells and can theoretically give rise to *all* types of cells in an organism. The range of cell types that can arise from adult stem cells is not yet fully understood.

into a wide variety of specialized cells (Figure 20.20), including even eggs and sperm.

The adult body also has stem cells, which serve to replace nonreproducing specialized cells as needed. In contrast to ES cells, **adult stem cells** are not able to give rise to all cell types in the organism, though they can generate multiple types. For example, one of the several types of stem cells in bone marrow can generate all the different kinds of blood cells (see Figure 20.20), and another can differentiate into bone, cartilage, fat, muscle, and the linings of blood vessels. To the surprise of many, the adult brain has been found to contain stem cells that continue to produce certain kinds of nerve cells there. Researchers have also reported finding stem cells in skin, hair, eyes, and dental pulp. Although adult animals have only tiny numbers of stem cells, scientists are

learning to identify and isolate these cells from various tissues and, in some cases, to grow them in culture. With the right culture conditions (for instance, the addition of specific growth factors), cultured stem cells from adult animals have been made to differentiate into multiple types of specialized cells, although none are as versatile as ES cells.

Research with embryonic or adult stem cells is a source of valuable data about differentiation and has enormous potential for medical applications. The ultimate aim is to supply cells for the repair of damaged or diseased organs: for example, insulin-producing pancreatic cells for people with type 1 diabetes or certain kinds of brain cells for people with Parkinson's disease or Huntington's disease. Adult stem cells from bone marrow have long been used as a source of immune system cells in patients whose own immune systems are nonfunctional because of genetic disorders or radiation treatments for cancer.

The developmental potential of adult stem cells is limited to certain tissues. ES cells hold more promise than adult stem cells for most medical applications because ES cells are **pluripotent**, capable of differentiating into many different cell types. The only way to obtain ES cells thus far, however, has been to harvest them from human embryos, which raises ethical and political issues.

ES cells are currently obtained from embryos donated by patients undergoing infertility treatments or from long-term cell cultures originally established with cells isolated from donated embryos. If scientists were able to clone human embryos to the blastocyst stage, they might be able to use such clones as the source of ES cells in the future. Furthermore, with a donor nucleus from a person with a particular disease, they might be able to produce ES cells that match the patient and are thus not rejected by his or her immune system when used for treatment. When the main aim of cloning is to produce ES cells to treat disease, the process is called *therapeutic cloning*. Although most people believe that reproductive cloning of humans is unethical, opinions vary about the morality of therapeutic cloning.

Induced Pluripotent Stem (iPS) Cells

Resolving the debate now seems less urgent because researchers have learned to turn back the clock in fully differentiated cells, reprogramming them to act like ES cells. The accomplishment of this feat, which posed formidable obstacles, was announced in 2007, first by labs using mouse skin cells and then by additional groups using cells from human skin and other organs or tissues. In all these cases, researchers transformed the differentiated cells into a type of ES cell by using a retrovirus to introduce extra, cloned copies of four "stem cell" master regulatory genes. The "deprogrammed" cells are known as *induced pluripotent stem (iPS)* cells because, in using this fairly simple laboratory technique to return them to their undifferentiated state, pluripotency

has been restored. The experiments that first transformed human differentiated cells into iPS cells are described in **Figure 20.21**. Shinya Yamanaka received the 2012 Nobel Prize in Medicine for this work, shared with John Gurdon, whose work you read about in Figure 20.16.

By many criteria, iPS cells can perform most of the functions of ES cells, but there are some differences in gene expression and other cellular functions, such as cell division. At least until these differences are fully understood, the study of ES cells will continue to make important contributions to the development of stem cell therapies. (In fact, it is likely that ES cells will always be a focus of basic research as well.) In the meantime, work is proceeding using the iPS cells that have been experimentally produced.

There are two major potential uses for human iPS cells. First, cells from patients suffering from diseases can be reprogrammed to become iPS cells, which can act as model cells for studying the disease and potential treatments. Human iPS cell lines have already been developed from individuals with type 1 diabetes, Parkinson's disease, and at least a dozen other diseases. Second, in the field of regenerative medicine, a patient's own cells could be reprogrammed into iPS cells and then used to replace nonfunctional tissues, such as insulin-producing cells of the pancreas.

Recently, in another surprising development, researchers have been able to identify genes that can directly reprogram a differentiated cell into another type of differentiated cell without passing through a pluripotent state. In the first reported example, one type of cell in the pancreas was transformed into another type. However, the two types of cells do not need to be very closely related: Another research group has been able to directly reprogram a skin fibroblast into a nerve cell. Development techniques that direct iPS cells or even fully differentiated cells to become specific cell types for regenerative medicine is an area of intense research, one that has already seen some success. The iPS cells created in this way could eventually provide tailor-made "replacement" cells for patients without using any human eggs or embryos, thus circumventing most ethical objections.

CONCEPT CHECK 20.3

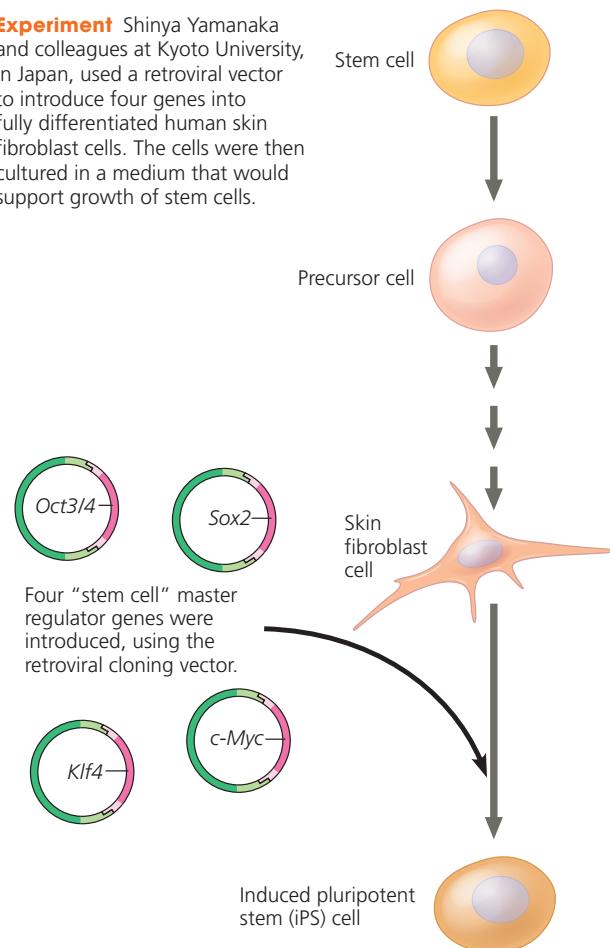
- Based on current knowledge, how would you explain the difference in the percentage of tadpoles that developed from the two kinds of donor nuclei in Figure 20.16?
- If you were to clone a carrot using the technique shown in Figure 20.15, would all the progeny plants ("clones") look identical? Why or why not?
- MAKE CONNECTIONS** Compare an individual carrot cell in Figure 20.15 with the fully differentiated muscle cell in Figure 18.18 in terms of their potential to develop into different cell types.

For suggested answers, see Appendix A.

▼ Figure 20.21 Inquiry

Can a fully differentiated human cell be "deprogrammed" to become a stem cell?

Experiment Shinya Yamanaka and colleagues at Kyoto University, in Japan, used a retroviral vector to introduce four genes into fully differentiated human skin fibroblast cells. The cells were then cultured in a medium that would support growth of stem cells.



Results Two weeks later, the cells resembled embryonic stem cells in appearance and were actively dividing. Their gene expression patterns, gene methylation patterns, and other characteristics were also consistent with those of embryonic stem cells. The iPS cells were able to differentiate into heart muscle cells, as well as other cell types.

Conclusion The four genes induced differentiated skin cells to become pluripotent stem cells, with characteristics of embryonic stem cells.

Source: K. Takahashi et al., Induction of pluripotent stem cells from adult human fibroblasts by defined factors, *Cell* 131:861–872 (2007).

WHAT IF? Patients with diseases such as heart disease, diabetes, or Alzheimer's could have their own skin cells reprogrammed to become iPS cells. Once procedures have been developed for converting iPS cells into heart, pancreatic, or nervous system cells, the patients' own iPS cells might be used to treat their disease. When organs are transplanted from a donor to a diseased recipient, the recipient's immune system may reject the transplant, a condition with serious and often fatal consequences. Would using iPS cells be expected to carry the same risk? Why or why not? Given that these cells are actively dividing, undifferentiated cells, what risks might this procedure carry?

CONCEPT 20.4

The practical applications of DNA-based biotechnology affect our lives in many ways

DNA technology is in the news almost every day. Most often, the topic is a new and promising application in medicine, but this is just one of numerous fields benefiting from DNA technology and genetic engineering.

Medical Applications

One important use of DNA technology is the identification of human genes whose mutation plays a role in genetic diseases. These discoveries may lead to ways of diagnosing, treating, and even preventing such conditions. DNA technology is also contributing to our understanding of “non-genetic” diseases, from arthritis to AIDS, since a person’s genes influence susceptibility to these diseases. Furthermore, diseases of all sorts involve changes in gene expression within the affected cells and often within the patient’s immune system. By using DNA microarray assays (see Figure 20.13) or other techniques to compare gene expression in healthy and diseased tissues, researchers hope to find many of the genes that are turned on or off in particular diseases. These genes and their products are potential targets for prevention or therapy.

Diagnosis and Treatment of Diseases

A new chapter in the diagnosis of infectious diseases has been opened by DNA technology, in particular the use of PCR and labeled nucleic acid probes to track down pathogens. For example, because the sequence of the RNA genome of HIV is known, RT-PCR can be used to amplify, and thus detect, HIV RNA in blood or tissue samples (see Figure 20.12). RT-PCR is often the best way to detect an otherwise elusive infective agent.

Medical scientists can now diagnose hundreds of human genetic disorders by using PCR with primers that target the genes associated with these disorders. The amplified DNA product is then sequenced to reveal the presence or absence of the disease-causing mutation. Among the genes for human diseases that have been identified are those for sickle-cell disease, hemophilia, cystic fibrosis, Huntington’s disease, and Duchenne muscular dystrophy. Individuals afflicted with such diseases can often be identified before the onset of symptoms, even before birth (see Figure 14.19). PCR can also be used to identify symptomless carriers of potentially harmful recessive alleles.

As you learned earlier, genome-wide association studies have pinpointed SNPs (single nucleotide polymorphisms) that are linked to disease-causing alleles (see Figure 20.13).

Individuals can be tested by PCR and sequencing for a SNP that is correlated with the abnormal allele. The presence of particular SNPs is correlated with increased risk for conditions such as heart disease, Alzheimer’s, and some types of cancer. Companies that offer individual genetic testing for risk factors like these are looking for previously identified, linked SNPs. It may be helpful for individuals to learn about their health risks, with the understanding that such genetic tests merely reflect correlations and do not make predictions.

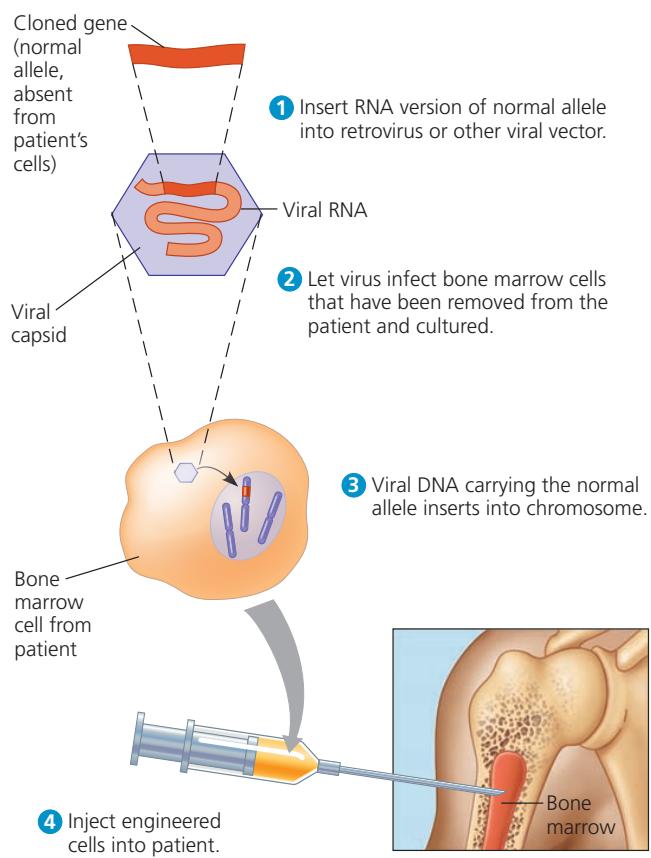
The techniques described in this chapter have also prompted improvements in disease treatments. By analyzing the expression of many genes in breast cancer patients, researchers have been able to refine their understanding of the different subtypes of breast cancer (see Figure 18.27). Knowing the expression levels of particular genes can help physicians determine the likelihood that the cancer will recur, thus helping them design an appropriate treatment. Given that some low-risk patients have a 96% survival rate over a ten-year period with no treatment, gene expression analysis allows doctors and patients access to valuable information when they are considering treatment options.

Many envision a future of “personalized medicine” where each person’s genetic profile can inform them about diseases or conditions for which they are especially at risk and help them make treatment choices. As we will discuss later in the chapter, a *genetic profile* is currently taken to mean a set of genetic markers such as SNPs. Ultimately, however, it will likely mean the complete DNA sequence of an individual—once sequencing becomes inexpensive enough. (See the interview with Charles Rotimi before Chapter 13.) Our ability to sequence a person’s genome rapidly and inexpensively is advancing faster than our development of appropriate treatments for the conditions we are characterizing. Still, the identification of genes involved in these conditions provides us with good targets for therapeutic interventions.

Human Gene Therapy

Gene therapy—the introduction of genes into an afflicted individual for therapeutic purposes—holds great potential for treating the relatively small number of disorders traceable to a single defective gene. In theory, a normal allele of the defective gene could be inserted into the somatic cells of the tissue affected by the disorder.

For gene therapy of somatic cells to be permanent, the cells that receive the normal allele must be cells that multiply throughout the patient’s life. Bone marrow cells, which include the stem cells that give rise to all the cells of the blood and immune system, are prime candidates. **Figure 20.22** outlines one procedure for gene therapy of an individual whose bone marrow cells do not produce a vital enzyme because of a single defective gene. One type of severe combined immunodeficiency (SCID) is caused by this kind of defect. If the treatment is successful, the patient’s bone marrow cells will begin producing the missing protein, and the patient may be cured.



▲ Figure 20.22 Gene therapy using a retroviral vector. A retrovirus that has been rendered harmless is used as a vector in this procedure, which exploits the ability of a retrovirus to insert a DNA transcript of its RNA genome into the chromosomal DNA of its host cell (see Figure 19.8). If the foreign gene carried by the retroviral vector is expressed, the cell and its descendants will possess the gene product. Cells that reproduce throughout life, such as bone marrow cells, are ideal candidates for gene therapy.

The procedure shown in Figure 20.22 has been used in gene therapy trials for SCID. In a trial begun in France in 2000, ten young children with SCID were treated by the same procedure. Nine of these patients showed significant, definitive improvement after two years, the first indisputable success of gene therapy. However, three of the patients subsequently developed leukemia, a type of blood cell cancer, and one of them died. Researchers have concluded it is likely that the insertion of the retroviral vector occurred near a gene that triggers the proliferation of blood cells. Using a viral vector that does not come from a retrovirus, clinical researchers have treated at least three other genetic diseases somewhat successfully with gene therapy: a type of progressive blindness (see Concept 50.3), a degenerative disease of the nervous system, and a blood disorder involving the β -globin gene. The successful trials involve very few patients but are still cause for cautious optimism.

Gene therapy raises many technical issues. For example, how can the activity of the transferred gene be controlled

so that cells make appropriate amounts of the gene product at the right time and in the right place? How can we be sure that the insertion of the therapeutic gene does not harm some other necessary cell function? As more is learned about DNA control elements and gene interactions, researchers may be able to answer such questions.

In addition to technical challenges, gene therapy provokes ethical questions. Some critics believe that tampering with human genes in any way is immoral or unethical. Other observers see no fundamental difference between the transplantation of genes into somatic cells and the transplantation of organs. You might wonder whether scientists are considering engineering human germ-line cells in the hope of correcting a defect in future generations. At present, no one in the mainstream scientific community is pursuing this goal—it is considered much too risky. Such genetic engineering is routinely done in laboratory mice, though, and the technical problems relating to similar genetic engineering in humans will eventually be solved. Under what circumstances, if any, should we alter the genomes of human germ lines? Would this inevitably lead to the practice of eugenics, a deliberate effort to control the genetic makeup of human populations? While we may not have to resolve these questions right now, considering them is worthwhile because they will probably arise at some point in the future.

Pharmaceutical Products

The pharmaceutical industry derives significant benefit from advances in DNA technology and genetic research, applying them to the development of useful drugs to treat diseases. Pharmaceutical products are synthesized using methods of either organic chemistry or biotechnology, depending on the nature of the product.

Synthesis of Small Molecules for Use as Drugs Determining the sequence and structure of proteins crucial for tumor cell survival has led to the identification of small molecules that combat certain cancers by blocking the function of these proteins. One drug, imatinib (trade name Gleevec), is a small molecule that inhibits a specific receptor tyrosine kinase (see Figure 11.8). The overexpression of this receptor, resulting from a chromosomal translocation, is instrumental in causing chronic myelogenous leukemia (CML; see Figure 15.16). Patients in the early stages of CML who are treated with imatinib have exhibited nearly complete, sustained remission from the cancer. Drugs that work in a similar way have also been used with success to treat a few types of lung and breast cancers. This approach is feasible only for cancers for which the molecular basis is fairly well understood.

In many cases of such drug-treated tumors, though, cells later arise that are resistant to the new drug. In one study, the whole genome of the tumor cells was sequenced both before and after the appearance of drug resistance. Comparison of the sequences showed genetic changes that allowed

the tumor cells to “get around” the drug-inhibited protein. So we can see that cancer cells demonstrate the principles of evolution: Certain tumor cells have a random mutation that allows them to survive in the presence of a particular drug, and as a consequence of natural selection in the presence of the drug, these are the cells that survive and reproduce.

Protein Production in Cell Cultures Pharmaceutical products that are proteins are commonly synthesized on a large scale using cell cultures. You learned earlier in the chapter about DNA cloning and gene expression systems for producing large quantities of a chosen protein that is present naturally in only minute amounts. The host cells used in such expression systems can even be engineered to secrete a protein as it is made, thereby simplifying the task of purifying it by traditional biochemical methods.

Among the first pharmaceutical products manufactured in this way were human insulin and human growth hormone (HGH). Some 2 million people with diabetes in the United States depend on insulin treatment to control their disease. Human growth hormone has been a boon to children born with a form of dwarfism caused by inadequate amounts of HGH. Another important pharmaceutical product produced by genetic engineering is tissue plasminogen activator (TPA). If administered shortly after a heart attack, TPA helps dissolve blood clots and reduces the risk of subsequent heart attacks.

For the past 25 years, scientists have also been working on producing proteins in plant cell cultures. They have enjoyed recent success using carrot cells in culture to make an enzyme involved in fat breakdown that is used for treatment of a rare human disease. Plant cells are easily grown in culture (see Figure 20.15), requiring less precise conditions than animal cells. Also, they are unlikely to get contaminated by viruses that could infect animals, a situation that has on occasion held up production of the enzyme for some time. This successful accomplishment is likely to be further extended to other therapeutic proteins in the future.

Protein Production by “Pharm” Animals In some cases, instead of using cell systems to produce large quantities of protein products, pharmaceutical scientists can use whole animals. They can introduce a gene from an animal of one genotype into the genome of another individual, often of a different species. This individual is then called a **transgenic** animal. To do this, they first remove eggs from a female of the recipient species and fertilize them *in vitro*. Meanwhile, they have cloned the desired gene from the donor organism. They then inject the cloned DNA directly into the nuclei of the fertilized eggs. Some of the cells integrate the foreign DNA, the *transgene*, into their genome and are able to express the foreign gene. The engineered embryos that arise from these zygotes are then surgically implanted in a surrogate mother. If an embryo develops successfully, the result is a transgenic animal that expresses its new, “foreign” gene.



▲ **Figure 20.23** **Goats as “pharm” animals.** This transgenic goat carries a gene for a human blood protein, antithrombin, which she secretes in her milk. Patients with a rare hereditary disorder in which this protein is lacking suffer from formation of blood clots in their blood vessels. Easily purified from the goat’s milk, the protein is used to prevent blood clots in these patients during surgery or childbirth.

Assuming that the introduced gene encodes a protein desired in large quantities, these transgenic animals can act as pharmaceutical “factories.” For example, a transgene for a human blood protein such as antithrombin, which prevents blood clots, can be inserted into the genome of a goat in such a way that the transgene’s product is secreted in the animal’s milk (**Figure 20.23**). The protein is then purified from the milk (which is easier than purification from a cell culture).

Human proteins produced in transgenic farm animals for use in humans may differ in some ways from the naturally produced human proteins, possibly because of subtle differences in protein modification. Therefore, such proteins must be tested very carefully to ensure that they (or contaminants from the farm animals) will not cause allergic reactions or other adverse effects in patients who receive them.

Forensic Evidence and Genetic Profiles

In violent crimes, body fluids or small pieces of tissue may be left at the scene or on the clothes or other possessions of the victim or assailant. If enough blood, semen, or tissue is available, forensic laboratories can determine the blood type or tissue type by using antibodies to detect specific cell-surface proteins. However, such tests require fairly fresh samples in relatively large amounts. Also, because many people have the same blood or tissue type, this approach can only exclude a suspect; it cannot provide strong evidence of guilt.

DNA testing, on the other hand, can identify the guilty individual with a high degree of certainty, because the DNA sequence of every person is unique (except for identical twins). Genetic markers that vary in the population can be analyzed for a given person to determine that individual’s unique set of genetic markers, or **genetic profile**. (This term is preferred over “DNA fingerprint” by forensic scientists, who want to emphasize the heritable aspect of these

markers rather than the fact that they produce a pattern on a gel that, like a fingerprint, is visually recognizable.) The FBI started applying DNA technology to forensics in 1988, using a method involving gel electrophoresis and nucleic acid hybridization to detect similarities and differences in DNA samples. This method required much smaller samples of blood or tissue than earlier methods—only about 1,000 cells.

Today, forensic scientists use an even more sensitive method that takes advantage of variations in length of genetic markers called **short tandem repeats (STRs)**. These are tandemly repeated units of two- to five-nucleotide sequences in specific regions of the genome. The number of repeats present in these regions is highly variable from person to person (polymorphic), and even for a single individual, the two alleles of an STR may differ from each other. For example, one individual may have the sequence ACAT repeated 30 times at one genome locus and 15 times at the same locus on the other homolog, whereas another individual may have 18 repeats at this locus on each homolog. (These two genotypes can be expressed by the two repeat numbers: 30,15 and 18,18.) PCR is used to amplify particular STRs, using sets of primers that are labeled with different-colored fluorescent tags; the length of the region, and thus the number of repeats, can then be determined by electrophoresis. The PCR step allows use of this method even when the DNA is in poor condition or available only in minute quantities. A tissue sample containing as few as 20 cells can be sufficient for PCR amplification.

In a murder case, for example, this method can be used to compare DNA samples from the suspect, the victim, and a small amount of blood found at the crime scene. The forensic scientist tests only a few selected portions of the DNA—usually 13 STR markers. However, even this small set of markers can provide a forensically useful genetic profile because the probability that two people (who are not identical twins) would have exactly the same set of STR markers is vanishingly small. The Innocence Project, a nonprofit organization dedicated to overturning wrongful convictions, uses STR analysis of archived samples from crime scenes to revisit old cases. As of 2013, more than 300 innocent people have been released from prison as a result of forensic and legal work by this group (**Figure 20.24**).

Genetic profiles can also be useful for other purposes. A comparison of the DNA of a mother, her child, and the purported father can conclusively settle a question of paternity. Sometimes paternity is of historical interest: Genetic profiles provided strong evidence that Thomas Jefferson or one of his close male relatives fathered at least one of the children of his slave Sally Hemings. Genetic profiles can also identify victims of mass casualties. The largest such effort occurred after the attack on the World Trade Center in 2001; more than 10,000 samples of victims' remains were compared with DNA samples from personal items, such as toothbrushes, provided by families. Ultimately, forensic scientists

- (a)** In 1984, Earl Washington was convicted and sentenced to death for the 1982 rape and murder of Rebecca Williams. His sentence was commuted to life in prison in 1993 due to new doubts about the evidence. In 2000, STR analysis by forensic scientists associated with the Innocence Project showed conclusively that he was innocent. This photo shows Washington just before his release in 2001, after 17 years in prison.



Source of sample	STR marker 1	STR marker 2	STR marker 3
Semen on victim	17,19	13,16	12,12
Earl Washington	16,18	14,15	11,12
Kenneth Tinsley	17,19	13,16	12,12

- (b)** In STR analysis, selected STR markers in a DNA sample are amplified by PCR, and the PCR products are separated by electrophoresis. The procedure reveals how many repeats are present for each STR locus in the sample. An individual has two alleles per STR locus, each with a certain number of repeats. This table shows the number of repeats for three STR markers in three samples: from semen found on the victim, from Washington, and from another man (Kenneth Tinsley), who was in prison because of an unrelated conviction. These and other STR data (not shown) exonerated Washington and led Tinsley to plead guilty to the murder.

▲ **Figure 20.24** STR analysis used to release an innocent man from prison.

succeeded in identifying almost 3,000 victims using these methods.

Just how reliable is a genetic profile? The greater the number of markers examined in a DNA sample, the more likely it is that the profile is unique to one individual. In forensic cases using STR analysis with 13 markers, the probability of two people having identical DNA profiles is somewhere between one chance in 10 billion and one in several trillion. (For comparison, the world's population is between 7 and 8 billion.) The exact probability depends on the frequency of those markers in the general population. Information on how common various markers are in different ethnic groups is critical because these marker frequencies may vary considerably among ethnic groups and between a particular ethnic group and the population as a whole. With the increasing availability of frequency data, forensic scientists can make extremely accurate statistical calculations. Thus, despite problems that can still arise from insufficient data, human error, or flawed evidence, genetic profiles are now accepted as compelling evidence by legal experts and scientists alike.

Environmental Cleanup

Increasingly, the remarkable ability of certain microorganisms to transform chemicals is being exploited for environmental cleanup. If the growth needs of such microbes make them unsuitable for direct use, scientists can now transfer the genes for their valuable metabolic capabilities into other microorganisms, which can then be used to treat environmental problems. For example, many bacteria can extract heavy metals, such as copper, lead, and nickel, from their environments and incorporate the metals into compounds such as copper sulfate or lead sulfate, which are readily recoverable. Genetically engineered microbes may become important in both mining (especially as ore reserves are depleted) and cleaning up highly toxic mining wastes. Biotechnologists are also trying to engineer microbes that can degrade chlorinated hydrocarbons and other harmful compounds. These microbes could be used in wastewater treatment plants or by manufacturers before the compounds are ever released into the environment.

Agricultural Applications

Scientists are working to learn more about the genomes of agriculturally important plants and animals. For a number of years, they have been using DNA technology in an effort to improve agricultural productivity. The selective breeding of both livestock (animal husbandry) and crops has exploited naturally occurring mutations and genetic recombination for thousands of years.

As we described earlier, DNA technology enables scientists to produce transgenic animals, which speeds up the selective breeding process. The goals of creating a transgenic animal are often the same as the goals of traditional breeding—for instance, to make a sheep with better quality wool, a pig with leaner meat, or a cow that will mature in a shorter time. Scientists might, for example, identify and clone a gene that causes the development of larger muscles (muscles make up most of the meat we eat) in one breed of cattle and transfer it to other cattle or even to sheep. However, problems such as low fertility or increased susceptibility to disease are not uncommon among farm animals carrying genes from other species. Animal health and welfare are important issues to consider when developing transgenic animals.

Agricultural scientists have already endowed a number of crop plants with genes for desirable traits, such as delayed ripening and resistance to spoilage and disease, as well as drought. The most commonly used vector for introducing new genes into plant cells is a plasmid, called the *Ti plasmid*, from the soil bacterium *Agrobacterium tumefaciens*. This plasmid integrates a segment of its DNA into the chromosomal DNA of its host plant cells (see Figure 35.25). To make transgenic plants, researchers engineer the plasmid to carry genes of interest and introduce it into cells. For many plant

species, a single tissue cell grown in culture can give rise to an adult plant (see Figure 20.15). Thus, genetic manipulations can be performed on an ordinary somatic cell and the cell then used to generate an organism with new traits.

Genetic engineering is rapidly replacing traditional plant-breeding programs, especially for useful traits, such as herbicide or pest resistance, determined by one or a few genes. Crops engineered with a bacterial gene making the plants resistant to an herbicide can grow while weeds are destroyed, and genetically engineered crops that can resist destructive insects reduce the need for chemical insecticides. In India, the insertion of a salinity resistance gene from a coastal mangrove plant into the genomes of several rice varieties has resulted in rice plants that can grow in water three times as salty as seawater. The research foundation that carried out this feat of genetic engineering estimates that one-third of all irrigated land has high salinity owing to overirrigation and intensive use of chemical fertilizers, representing a serious threat to the food supply. Thus, salinity-resistant crop plants would be enormously valuable worldwide.

Safety and Ethical Questions Raised by DNA Technology

Early concerns about potential dangers associated with recombinant DNA technology focused on the possibility that hazardous new pathogens might be created. What might happen, for instance, if cancer cell genes were transferred into bacteria or viruses? To guard against such rogue microbes, scientists developed a set of guidelines that were adopted as formal government regulations in the United States and some other countries. One safety measure is a set of strict laboratory procedures designed to protect researchers from infection by engineered microbes and to prevent the microbes from accidentally leaving the laboratory. In addition, strains of microorganisms to be used in recombinant DNA experiments are genetically crippled to ensure that they cannot survive outside the laboratory. Finally, certain obviously dangerous experiments have been banned.

Today, most public concern about possible hazards centers not on recombinant microbes but on **genetically modified (GM) organisms** used as food. A GM organism is one that has acquired by artificial means one or more genes from another species or even from another variety of the same species. Some salmon, for example, have been genetically modified by addition of a more active salmon growth hormone gene. However, the majority of the GM organisms that contribute to our food supply are not animals, but crop plants.

GM crops are widespread in the United States, Argentina, and Brazil; together these countries account for over 80% of the world's acreage devoted to such crops. In the United States, most corn, soybean, and canola crops are genetically modified, and GM products are not required to be labeled at present. However, the same foods are an ongoing

subject of controversy in Europe, where the GM revolution has met with strong opposition. Many Europeans are concerned about the safety of GM foods and the possible environmental consequences of growing GM plants. In the year 2000, negotiators from 130 countries agreed on a Biosafety Protocol that requires exporters to identify GM organisms present in bulk food shipments and allows importing countries to decide whether the products pose environmental or health risks. (Although the United States declined to sign the agreement, it went into effect anyway because the majority of countries were in favor of it.) Since then, European countries have, on occasion, refused crops from the United States and other countries, leading to trade disputes. Although a small number of GM crops have been grown on European soil, these products have generally failed in local markets, and the future of GM crops in Europe is uncertain.

Advocates of a cautious approach toward GM crops fear that transgenic plants might pass their new genes to close relatives in nearby wild areas. We know that lawn and crop grasses, for example, commonly exchange genes with wild relatives via pollen transfer. If crop plants carrying genes for resistance to herbicides, diseases, or insect pests pollinated wild ones, the offspring might become “super weeds” that are very difficult to control. Another worry concerns possible risks to human health from GM foods. Some people fear that the protein products of transgenes might lead to allergic reactions. Although there is some evidence that this could happen, advocates claim that these proteins could be tested in advance to avoid producing ones that cause allergic reactions. (For further discussion of plant biotechnology and GM crops, see Concept 38.3.)

Today, governments and regulatory agencies throughout the world are grappling with how to facilitate the use of biotechnology in agriculture, industry, and medicine while ensuring that new products and procedures are safe. In the United States, such applications of biotechnology are evaluated for potential risks by various regulatory agencies, including the Food and Drug Administration, the

Environmental Protection Agency, the National Institutes of Health, and the Department of Agriculture. Meanwhile, these same agencies and the public must consider the ethical implications of biotechnology.

Advances in biotechnology have allowed us to obtain complete genome sequences for humans and many other species, providing a vast treasure trove of information about genes. We can ask how certain genes differ from species to species, as well as how genes and, ultimately, entire genomes have evolved. (These are the subjects of Chapter 21.) At the same time, the increasing speed and falling cost of sequencing the genomes of individuals are raising significant ethical questions. Who should have the right to examine someone else's genetic information? How should that information be used? Should a person's genome be a factor in determining eligibility for a job or insurance? Ethical considerations, as well as concerns about potential environmental and health hazards, will likely slow some applications of biotechnology. There is always a danger that too much regulation will stifle basic research and its potential benefits. However, the power of DNA technology and genetic engineering—our ability to profoundly and rapidly alter species that have been evolving for millennia—demands that we proceed with humility and caution.

CONCEPT CHECK 20.4

1. What is the advantage of using stem cells for gene therapy?
2. List at least three different properties that have been acquired by crop plants via genetic engineering.
3. **WHAT IF?** As a physician, you have a patient with symptoms that suggest a hepatitis A infection, but you have not been able to detect viral proteins in the blood. Knowing that hepatitis A is an RNA virus, what lab test could you perform to support your diagnosis? Explain the results that would support your hypothesis.

For suggested answers, see Appendix A.

20 Chapter Review

SUMMARY OF KEY CONCEPTS

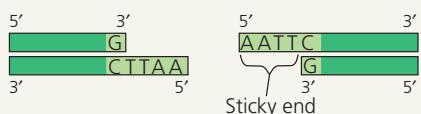
CONCEPT 20.1

DNA sequencing and DNA cloning are valuable tools for genetic engineering and biological inquiry (pp. 409–417)

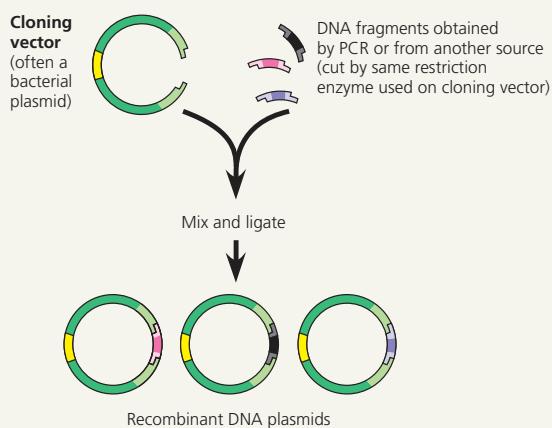
- **Nucleic acid hybridization**, the base pairing of one strand of a nucleic acid to the complementary sequence on a strand from another nucleic acid molecule, is widely used in DNA technology.

- **DNA sequencing** can be carried out using the *dideoxy chain termination* method in automated sequencing machines.
- Next-generation (high-throughput) techniques for sequencing DNA are based on *sequencing by synthesis*: DNA polymerase is used to synthesize a stretch of DNA from a single-stranded template, and the order in which nucleotides are added reveals the sequence.
- **Gene cloning** (or DNA cloning) produces multiple copies of a gene (or DNA fragment) that can be used in analyzing and manipulating DNA and can yield useful new products or organisms with beneficial traits.

- In **genetic engineering**, bacterial **restriction enzymes** are used to cut DNA molecules within short, specific nucleotide sequences (**restriction sites**), yielding a set of double-stranded **restriction fragments** with single-stranded **sticky ends**:



- The sticky ends on **restriction fragments** from one DNA source can base-pair with complementary sticky ends on fragments from other DNA molecules. Sealing the base-paired fragments with **DNA ligase** produces **recombinant DNA** molecules.
- DNA restriction fragments of different lengths can be separated by **gel electrophoresis**.
- The **polymerase chain reaction (PCR)** can produce many copies of (amplify) a specific target segment of DNA *in vitro*, using primers that bracket the desired sequence and a heat-resistant DNA polymerase.
- To clone a eukaryotic gene:



Recombinant plasmids are returned to host cells, each of which divides to form a clone of cells.

- Several technical difficulties hinder the expression of cloned eukaryotic genes in bacterial host cells. The use of cultured eukaryotic cells as host cells, coupled with appropriate **expression vectors**, helps avoid these problems.

? Describe how the process of gene cloning results in a cell clone containing a recombinant plasmid.

CONCEPT 20.2

Biologists use DNA technology to study gene expression and function (pp. 417–422)

- Several techniques use hybridization of a **nucleic acid probe** to detect the presence of specific mRNAs.
- In situ hybridization** and RT-PCR can detect the presence of a given mRNA in a tissue or an RNA sample, respectively.
- DNA microarrays** are used to identify sets of genes co-expressed by a group of cells. Alternatively, their cDNAs can be sequenced (RNA-seq).
- For a gene of unknown function, experimental inactivation of the gene (a gene knockout) and observation of the resulting phenotypic effects can provide clues to its function. In humans, **genome-wide association studies** use **single nucleotide polymorphisms (SNPs)** as genetic markers for alleles that are associated with particular conditions.

? What useful information is obtained by detecting expression of specific genes?

CONCEPT 20.3

Cloned organisms and stem cells are useful for basic research and other applications (pp. 422–427)

- The question of whether all the cells in an organism have the same genome prompted the first attempts at organismal cloning.
- Single differentiated cells from plants are often **totipotent**: capable of generating all the tissues of a complete new plant.
- Transplantation of the nucleus from a differentiated animal cell into an enucleated egg can sometimes give rise to a new animal.
- Certain embryonic **stem cells** (ES cells) from animal embryos and particular adult stem cells from adult tissues can reproduce and differentiate both in the lab and in the organism, offering the potential for medical use. ES cells are **pluripotent** but difficult to acquire. Induced pluripotent stem (iPS) cells, generated by reprogramming differentiated cells, resemble ES cells in their capacity to differentiate. Some differentiated cells have been directly reprogrammed to become different cell types. These cells and iPS cells hold promise for medical research and regenerative medicine.

? Describe how a researcher could carry out (1) organismal cloning, (2) production of ES cells, and (3) generation of iPS cells, focusing on how the cells are reprogrammed and using mice as an example. (The procedures are basically the same in humans and mice.)

CONCEPT 20.4

The practical applications of DNA-based biotechnology affect our lives in many ways (pp. 428–433)

- DNA technology, including the analysis of genetic markers such as SNPs, is increasingly being used in the diagnosis of genetic and other diseases and offers potential for better treatment of genetic disorders (or even permanent cures through **gene therapy**), as well as more informed cancer therapies. DNA technology is used with cell cultures in the large-scale production of protein hormones and other proteins with therapeutic uses. Some therapeutic proteins are being produced in **transgenic** “pharm” animals.
- Analysis of genetic markers such as **short tandem repeats (STRs)** in DNA isolated from tissue or body fluids found at crime scenes leads to a **genetic profile**. Use of genetic profiles can provide definitive evidence that a suspect is innocent or strong evidence of guilt. Such analysis is also useful in parent-hood disputes and in identifying the remains of crime victims.
- Genetically engineered microorganisms can be used to extract minerals from the environment or degrade various types of toxic waste materials.
- The aims of developing transgenic plants and animals are to improve agricultural productivity and food quality.
- The potential benefits of genetic engineering must be carefully weighed against the potential for harm to humans or the environment.

? What factors affect whether a given genetic disease would be a good candidate for successful gene therapy?

TEST YOUR UNDERSTANDING

LEVEL 1: KNOWLEDGE/COMPREHENSION

- In DNA technology, the term *vector* can refer to
 - the enzyme that cuts DNA into restriction fragments.
 - the sticky end of a DNA fragment.
 - a SNP marker.
 - a plasmid used to transfer DNA into a living cell.

- Which of the following tools of DNA technology is incorrectly paired with its use?
 - electrophoresis—separation of DNA fragments
 - DNA ligase—cutting DNA, creating sticky ends of restriction fragments
 - DNA polymerase—polymerase chain reaction to amplify sections of DNA
 - reverse transcriptase—production of cDNA from mRNA
- Plants are more readily manipulated by genetic engineering than are animals because
 - plant genes do not contain introns.
 - more vectors are available for transferring recombinant DNA into plant cells.
 - a somatic plant cell can often give rise to a complete plant.
 - plant cells have larger nuclei.
- A paleontologist has recovered a bit of tissue from the 400-year-old preserved skin of an extinct dodo (a bird). To compare a specific region of the DNA from a sample with DNA from living birds, which of the following would be most useful for increasing the amount of dodo DNA available for testing?
 - SNP analysis
 - polymerase chain reaction (PCR)
 - electroporation
 - gel electrophoresis
- DNA technology has many medical applications. Which of the following is *not* done routinely at present?
 - production of hormones for treating diabetes and dwarfism
 - production of microbes that can metabolize toxins
 - introduction of genetically engineered genes into human gametes
 - prenatal identification of genetic disease alleles

LEVEL 2: APPLICATION/ANALYSIS

- Which of the following would *not* be true of cDNA produced using human brain tissue as the starting material?
 - It could be amplified by the polymerase chain reaction.
 - It was produced from pre-mRNA using reverse transcriptase.
 - It could be labeled and used as a probe to detect genes expressed in the brain.
 - It lacks the introns of the pre-mRNA.
- Expression of a cloned eukaryotic gene in a bacterial cell involves many challenges. The use of mRNA and reverse transcriptase is part of a strategy to solve the problem of
 - post-transcriptional processing.
 - post-translational processing.
 - nucleic acid hybridization.
 - restriction fragment ligation.
- Which of the following sequences in double-stranded DNA is most likely to be recognized as a cutting site for a restriction enzyme?

a. AAGG	c. ACCA
TTCC	TGGT
b. GGCC	d. AAAA
CCGG	TTTT

LEVEL 3: SYNTHESIS/EVALUATION

- MAKE CONNECTIONS** Imagine you want to study one of the human crystallins, proteins present in the lens of the eye (see Figure 1.8). To obtain a sufficient amount of the protein of interest, you decide to clone the gene that codes for it. Assume you know the sequence of this gene. How would you go about this?

- DRAW IT** You are cloning an aardvark gene, using a bacterial plasmid as a vector. The green diagram shows the plasmid, which contains the restriction site for the enzyme used in Figure 20.6. Above the plasmid is a segment of linear aardvark DNA that was synthesized using PCR. Diagram your cloning procedure, showing what would happen to these two molecules during each step. Use one color for the aardvark DNA and its bases and another color for those of the plasmid. Label each step and all 5' and 3' ends.

EVOLUTION CONNECTION

Ethical considerations aside, if DNA-based technologies became widely used, how might they change the way evolution proceeds, as compared with the natural evolutionary mechanisms that have operated for the past 4 billion years?

SCIENTIFIC INQUIRY

You hope to study a gene that codes for a neurotransmitter protein produced in human brain cells. You know the amino acid sequence of the protein. Explain how you might (a) identify what genes are expressed in a specific type of brain cell, (b) identify (and isolate) the neurotransmitter gene, (c) produce multiple copies of the gene for study, and (d) produce large quantities of the neurotransmitter for evaluation as a potential medication.

WRITE ABOUT A THEME: INFORMATION

In a short essay (100–150 words), discuss how the genetic basis of life plays a central role in biotechnology.

SYNTHESIZE YOUR KNOWLEDGE



The water in the Yellowstone National Park hot springs shown here is around 160°F (70°C). Biologists assumed that no species of organisms could live in water above about 130°F (55°C), so they were surprised to find several species of bacteria there, now called *thermophiles* ("heat-lovers"). You've learned in this chapter how an enzyme from one species, *Thermus aquaticus*, made feasible one of the most important DNA-based techniques used in labs today. What was the enzyme, and what was the value of its being isolated from a thermophile? Can you think of reasons other enzymes from this bacterium (or other thermophiles) might also be valuable?

For selected answers, see Appendix A.

MasteringBiology®

Students Go to **MasteringBiology** for assignments, the eText, and the Study Area with practice tests, animations, and activities.

Instructors Go to **MasteringBiology** for automatically graded tutorials and questions that you can assign to your students, plus Instructor Resources.