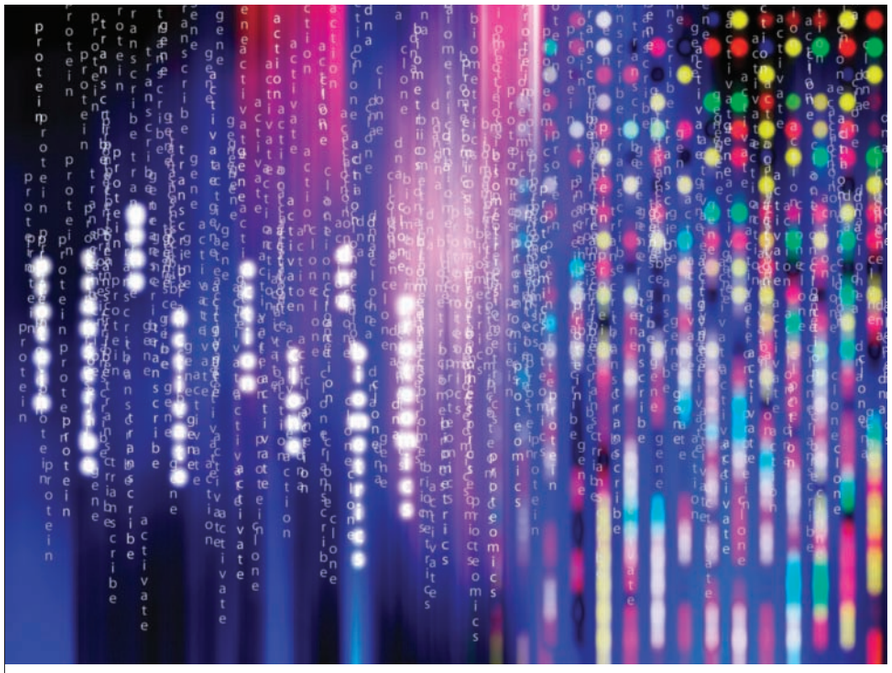


Pasieka/SPL/Photo Researchers, Inc.

Protein microarray, a key tool of proteomics, the study of the complete set of proteins that can be expressed by an organism's genome. Each colored dot is a protein, with a specific color for each protein being studied.

18 DNA Technologies and Genomics

Why It Matters

In early October 1994, 32-year-old Shirley Duguay, a mother of five, disappeared from her home on Prince Edward Island, Canada. Within a few days, her car was found abandoned; bloodstains inside matched her blood type. Several months later the Royal Canadian Mounted Police (RCMP) found Duguay's body in a shallow grave. Among the chief suspects in the murder was her estranged common-law husband, Douglas Beamish, who was living nearby with his parents.

While searching for Duguay, the RCMP discovered a plastic bag containing a man's leather jacket with the victim's blood on it. Beamish's friends and family acknowledged that Beamish had a similar jacket, but none could or would positively identify it. In the lining of the jacket investigators found 27 white hairs, which forensic scientists identified as cat hairs. The RCMP remembered that Beamish's parents had a white cat named Snowball **(Figure 18.1).** Could they prove that the cat hair in the jacket was Snowball's?

A Mountie investigator used the Internet to find two experts on cat genomes, Marilyn Menotti-Raymond and Stephen J. O'Brien of the Laboratory of Genome Diversity at the U.S. National Cancer Institute. Menotti-Raymond and O'Brien analyzed DNA from the root of one of

Figure 18.1
Snowball, the key to a murder case.
(© Marilyn Menotti-Raymond, The National Cancer Institute–Frederick)

the cat hairs taken from the jacket and from a blood sample taken from Snowball. They then used a technique called the polymerase chain reaction (PCR) to amplify 10 specific regions of the cat genome, each of which varies among cats in the number of copies of a short (two-nucleotide) repeated sequence. They found that the hair and blood samples matched perfectly, providing strong evidence that the hair came from Snowball.

Beamish was tried for the murder of Shirley Duguay. The evidence presented by Menotti-Raymond and O'Brien helped convict him, and he was sentenced to 18 years in prison.

The researchers' analysis of the cat hair and blood is an example of DNA fingerprinting. (Though human DNA fingerprinting evidence is now extensively used in court cases, the Beamish case was the first to admit nonhuman animal DNA fingerprinting data as evidence.) DNA fingerprinting is an application of **DNA technologies**, techniques to isolate, purify, analyze, and manipulate DNA sequences. Scientists use DNA technologies both for basic research into the biology of organisms and for applied research. The use of DNA technologies to alter genes for practical purposes is called **genetic engineering.**

Genetic engineering is the latest addition to the broad area known as *biotechnology,* which is any technique applied to biological systems or living organisms to make or modify products or processes for a specific purpose. Thus, biotechnology includes manipulations that do not involve DNA technologies, such as the use of yeast to brew beer and bake bread and the use of bacteria to make yogurt and cheese.

In this chapter, we focus on how biologists isolate genes and manipulate them for basic and applied research. You will learn about the basic DNA technologies and their applications to research in biology, to genetic engineering, and to the analysis of genomes. You will also learn about some of the risks and controversies surrounding genetic engineering and about some of the scientific, social, and ethical questions related to its application.

We begin our discussion with a description of methods used to obtain genes in large quantities, an essential step for their analysis or manipulation.

18.1 DNA Cloning

Remember from Chapter 17 that a *clone* is a line of genetically identical cells or individuals derived from a single ancestor. DNA cloning is a method for producing many copies of a piece of DNA, such as a gene of interest; that is, a gene that a researcher wants to study or manipulate. Scientists clone DNA for many reasons. For example, a researcher might be interested in how a particular human gene functions. Each human cell contains only two copies of most genes, amounting to a very small fraction of the total amount of DNA in a diploid cell. In its natural state in the genome, then, the gene is extremely difficult to study. However, through DNA cloning, a researcher can produce a sample large enough for scientific experimentation.

Cloned genes are used in basic research to find out about their biological functions. For example, researchers can determine the DNA sequence of a cloned gene, giving them the ultimate information about its structure. Also, by manipulating the gene and inducing mutations in it, they can gain information about its function and about how its expression is regulated. Cloned genes can be expressed in bacteria, and the proteins encoded by the cloned genes can be produced in quantity and purified. Those proteins can be used in basic research or, in the case of genes encoding proteins of pharmaceutical or clinical importance, they can be used in applied research.

An overview of one common method for cloning a gene of interest from a genome is shown in **Figure 18.2;** the method uses bacteria (commonly, *Escherichia coli*) and plasmids, the small circular DNA molecules that replicate separately from the bacterial chromosome (see Section 17.1). The researcher extracts DNA containing a gene of interest from cells and cuts it into fragments. The fragments are inserted into plasmids producing *recombinant DNA molecules*—**recombinant DNA** is DNA from two or more different sources joined together. The recombinant plasmids are introduced into bacteria; each bacterium receives a different plasmid. The bacterium continues growing and dividing, and as it does, the plasmid continues to replicate. It is through replication of the plasmid that amplification of the piece of DNA inserted into the plasmid occurs. The final step, then, is to identify the bacteria containing the plasmid with the gene of interest and isolate it for further study.

Bacterial Enzymes Called Restriction Endonucleases Form the Basis of DNA Cloning

The key to DNA cloning is the specific joining of two DNA molecules from different sources, such as a genomic DNA fragment and a bacterial plasmid (see Figure 18.2). This specific joining of DNA is made possible, in part, by bacterial enzymes called **restriction endonucleases** (also called **restriction enzymes**), discovered in the late 1960s. Restriction enzymes recognize short, specific DNA sequences called *restriction sites,* typically four to eight base pairs long, and cut the

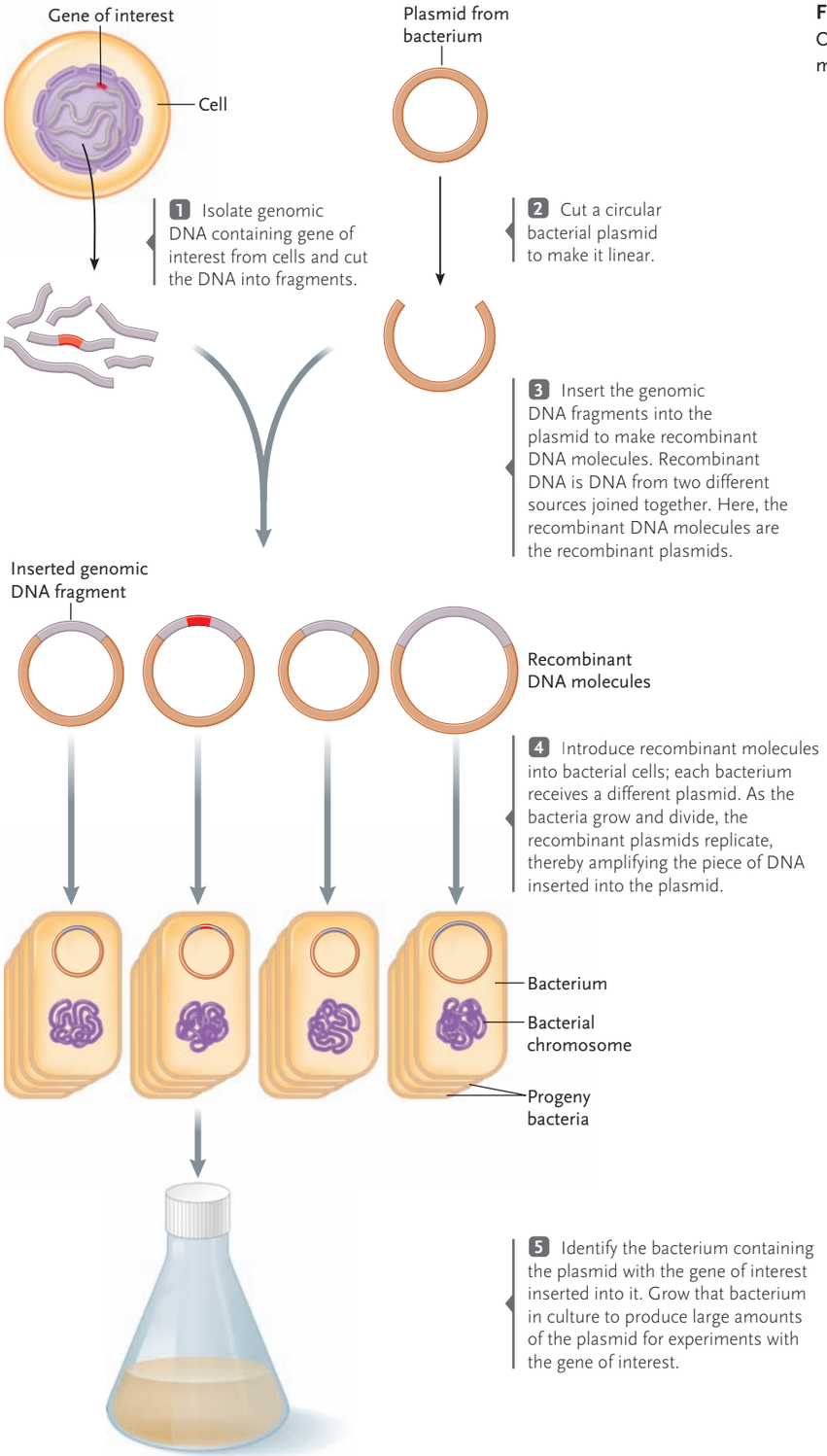


Figure 18.2
Overview of cloning DNA fragments in a bacterial plasmid.

DNA at specific locations within those sequences. The DNA fragments produced by cutting a long DNA molecule with a restriction enzyme are known as **restriction fragments.**

The "restriction" in the name of the enzymes refers to their normal role inside bacteria, in which the enzymes defend against viral attack by breaking down (restricting) the DNA molecules of infecting viruses. The bacterium protects the restriction sites in its own DNA from cutting by modifying bases in those sites enzymatically, thereby blocking the action of its restriction enzyme.

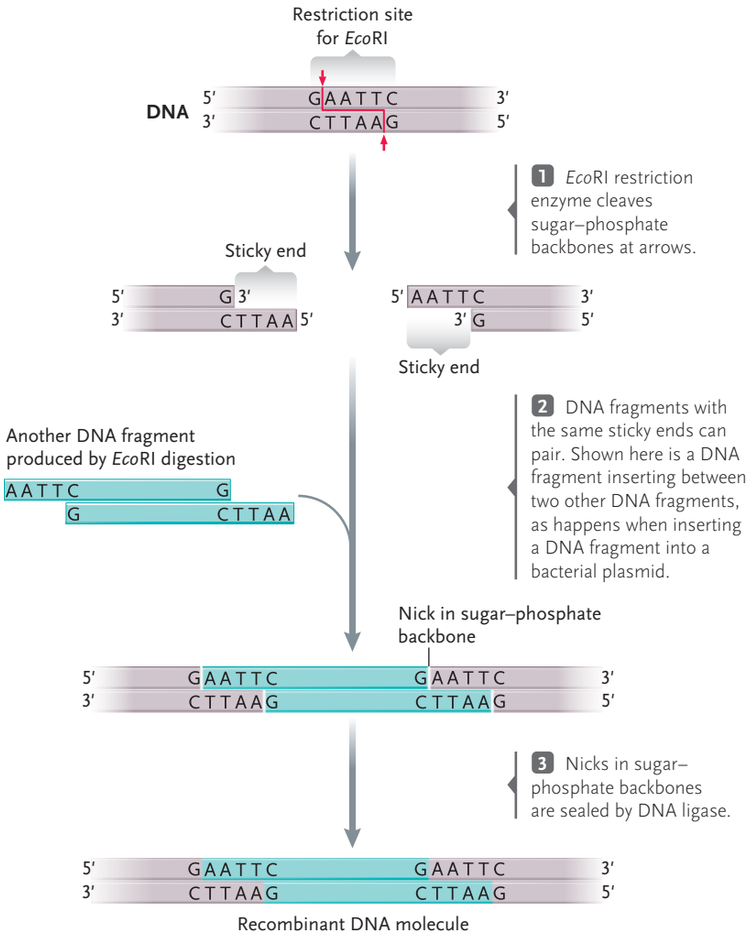


Figure 18.3
The restriction site for the restriction enzyme *Eco*RI, and the generation of a recombinant DNA molecule by complementary base pairing of DNA fragments produced by digestion with the same restriction enzyme.

Hundreds of different restriction enzymes have been identified, each one cutting DNA at a specific restriction site. As illustrated by the restriction site of *Eco*RI **(Figure 18.3),** most restriction sites are symmetrical in that the sequence of nucleotides read in the 5′→3′ direction on one strand is same as the sequence read in the 5′→3′ direction on the complementary strand. The restriction enzymes most used in cloning—such as *Eco*RI—cleave the sugar–phosphate backbones of DNA to produce DNA fragments with single-stranded ends (step 1). The ends are called **sticky ends** because the short single-stranded regions can form hydrogen bonds with complementary sticky ends on any other DNA molecules cut with the same enzyme. For example, step 2 shows the insertion of a DNA molecule with sticky ends produced by *Eco*RI between two other DNA molecules with the same sticky ends. The pairings leave nicks in the sugar–phosphate backbones of the DNA strands that are sealed by *DNA ligase,* an enzyme that has the same function in DNA replication (step 3; see Section 14.3). The result is DNA from two different sources joined together—a recombinant DNA molecule.

Bacterial Plasmids Illustrate the Use of Restriction Enzymes in Cloning

The bacterial plasmids used for cloning are examples of cloning vectors—DNA molecules into which a DNA fragment can be inserted to form a recombinant DNA molecule for cloning. Bacterial plasmid cloning vectors are not naturally in bacteria; they are plasmids modified to have special features. Commonly, plasmid cloning vectors are engineered to contain two genes that are useful in the final steps of a cloning experiment for sorting bacteria that have recombinant plasmids from those that do not. The amp^R gene encodes an enzyme that breaks down the antibiotic ampicillin; when the plasmid is introduced into *E. coli* and the amp^R gene is expressed, the bacteria become resistant to ampicillin. The $lacZ^+$ gene encodes β-galactosidase (recall the *lac* operon from Section 16.1), which hydrolyzes the sugar lactose, as well as a number of synthetic substrates. Restriction sites are located within the $lacZ^+$ gene, but do not alter the gene's function. For a given cloning experiment, one of these restriction sites is chosen.

Cloning a Gene of Interest. **Figure 18.4** expands on the overview of Figure 18.2 to show the steps used to clone a gene of interest using a plasmid cloning vector and restriction enzymes. Genomic DNA isolated from the organism in which the gene is found is cut with a restriction enzyme, and a plasmid cloning vector is cut within the $lacZ^+$ gene with the same restriction enzyme (steps 1 and 2). Mixing the DNA fragments and cut plasmid together with DNA ligase produces various joined molecules as the sticky ends pair and the enzyme seals them together. Some of these molecules are recombinant plasmids consisting of, in each case, a DNA fragment inserted into the plasmid cloning vector; others are nonrecombinant plasmids resulting from the cut plasmid resealed into a circle without an inserted fragment (step 3). In addition, ligase joins together pieces of genomic DNA with no plasmid involved. Only the recombinant plasmids are important in the cloning of the gene of interest; we sort out the other two undesired molecules in later steps.

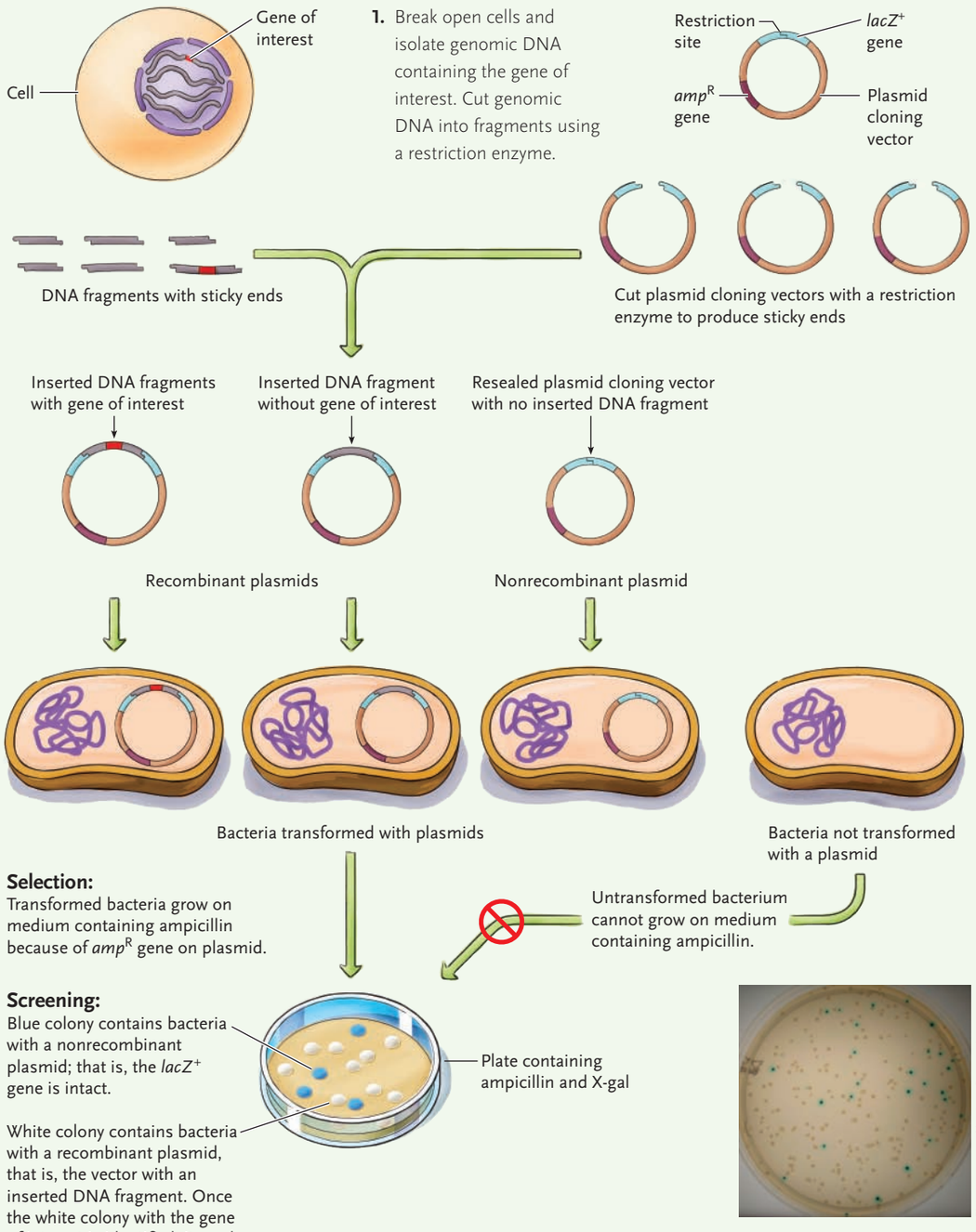
Next, the DNA molecules are transformed—introduced—into ampicillin-sensitive, $lacZ^-$ *E. coli* (which cannot make β-galactosidase), and the transformed bacteria are spread on a plate of agar growth medium containing ampicillin and the β-galactosidase substrate X-gal (steps 4 and 5). (Section 17.1 describes techniques for transformation of DNA into bacteria.) Only bacteria with a plasmid can grow and form colonies because expression of the plasmid's amp^R gene makes the bacteria resistant to ampicillin (see Figure 18.4 results). Within each cell of a colony, the plasmids have replicated until a hundred or so are present.

Figure 18.4 Research Method

Cloning a Gene of Interest in a Plasmid Cloning Vector

PURPOSE: Cloning a gene produces many copies of a gene of interest that can be used, for example, to determine the DNA sequence of the gene, to manipulate the gene in basic research experiments, to understand its function, and to produce the protein encoded by the gene.

PROTOCOL:



INTERPRETING THE RESULTS: Cloning and blue-white screening produce clearly identifiable white colonies containing recombinant plasmids. Most of the white colonies will contain plasmids that do not contain the gene of interest. Further screening will be done to identify the particular white colony that contains a plasmid with the gene of interest (Figure 18.5). Once identified, the colony can be cultured to produce large quantities of the plasmid for analysis or manipulation of the gene.

The X-gal in the medium distinguishes between bacteria that have been transformed with recombinant plasmids and nonrecombinant plasmids by *blue-white screening* (see Figure 18.4 results). If a colony produces β-galactosidase, it converts X-gal to a blue product and the colony turns blue, but if a colony does not produce the enzyme, X-gal is unchanged and the colony remains white. Colonies containing nonrecombinant plasmids have an intact *lacZ*$^+$ gene, produce the enzyme, and turn blue. Colonies containing recombinant plasmids are white because those plasmids each contain a DNA fragment inserted into the *lacZ*$^+$ gene, so they do not produce the enzyme. The white colonies are examined to find the one containing a recombinant plasmid with the gene of interest.

Two researchers, Paul Berg and Stanley Cohen, were prime movers in the development of DNA cloning techniques using restriction enzymes and bacterial plasmids. Berg and Cohen received a Nobel Prize in 1980 for their research, which pushed DNA technology to the forefront of biological investigations.

Identifying the Clone Containing the Gene of Interest. How is a clone containing the gene of interest identified amongst the population of clones? The gene of interest has a unique DNA sequence, which is the basis for one commonly used identification technique. In this technique, called **DNA hybridization**, the gene of interest is identified in the set of clones when it base pairs with a short, single-stranded complementary DNA or RNA molecule called a *nucleic acid probe* **(Figure 18.5).** The probe is typically labeled with a radioactive or a nonradioactive tag, so investigators can detect it. In our example, if we know the sequence of part of the gene of interest, we can use that information to design and synthesize a nucleic acid probe. Or, we can take advantage of DNA sequence similarities of evolutionarily related organisms. For instance, we could make a probe for a human gene based on the sequence of an equivalent mouse gene. Once a colony containing plasmids with the gene of interest has been identified, that colony can be used to produce large quantities of the cloned gene.

DNA Libraries Contain Collections of Cloned DNA Fragments

As you have seen, the starting point for cloning a gene of interest is a large set of plasmid clones carrying fragments representing all of the DNA of an organism's genome. A collection of clones that contains a copy of every DNA sequence in a genome is called a **genomic library.** A genomic library can be made using plasmid cloning vectors or any other kind of cloning vector. The number of clones in a genomic library increases with the size of the genome. For example, a yeast genomic library of plasmid clones consists of hundreds of plasmids, whereas a human genomic library of plasmid clones consists of thousands of plasmids.

A genomic library is a resource containing all of the DNA of an organism cut into pieces. Just as for a book library, where you can search through the same set of books on various occasions to find different passages of interest, you can search through the same genomic library on various occasions to find and isolate different genes or other DNA sequences.

Researchers also commonly use another kind of DNA library that is made starting with mRNA molecules isolated from a cell. To convert single-stranded mRNA to double-stranded DNA for cloning (RNA cannot be cloned) first they use the enzyme *reverse transcriptase* (made by retroviruses) to make a single-stranded DNA that is complementary to the mRNA. Then they degrade the mRNA strand with an enzyme, and use DNA polymerase to make a second DNA strand that is complementary to the first. The result is a **complementary DNA (cDNA).** After adding restriction sites to each end, they insert the cDNA into a cloning vector as described for the genomic library. The entire collection of cloned cDNAs made from the mRNAs isolated from a cell is a **cDNA library.**

Not all genes are active in every cell. Therefore, a cDNA library is limited in that it includes copies of only the genes that were active in the cells used as the starting point for creation of the library. This limitation can be an advantage, however, in identifying genes active in one cell type and not another. cDNA libraries are useful, therefore, for providing clues to the changes in gene activity that are responsible for cell differentiation and specialization. An ingenious method for comparing the cDNAs libraries produced by different cell types—the DNA chip—is described later in this chapter.

cDNA libraries provide a critical advantage to genetic engineers who wish to insert eukaryotic genes into bacteria, particularly when the bacteria are to be used as "factories" for making the protein encoded in the gene. The genes in eukaryotic nuclear DNA typically contain many *introns,* spacer sequences that interrupt the amino acid-coding sequence of a gene (see Section 15.3). Because bacterial DNA does not contain introns, bacteria are not equipped to process eukaryotic genes correctly. However, the cDNA copy of a eukaryotic mRNA already has the introns removed, so bacteria can transcribe and translate it accurately to make eukaryotic proteins.

Genomic and cDNA libraries both depend on cloning in a living cell to produce multiple copies of the DNA of interest. Next we look at a highly automated method of making copies of a targeted piece of DNA in a genome.

The Polymerase Chain Reaction (PCR) Amplifies DNA in Vitro

Producing multiple DNA copies by cloning requires a series of techniques and considerable time. A much more rapid process, **polymerase chain reaction (PCR),**

Figure 18.5 Research Method

DNA Hybridization to Identify a DNA Sequence of Interest

PURPOSE: Hybridization with a specific DNA probe allows researchers to detect a specific DNA sequence, such as a gene, within a population of DNA molecules. Here, DNA hybridization is used to screen a collection of bacterial colonies to identify those containing a recombinant plasmid with a gene of interest.

PROTOCOL:

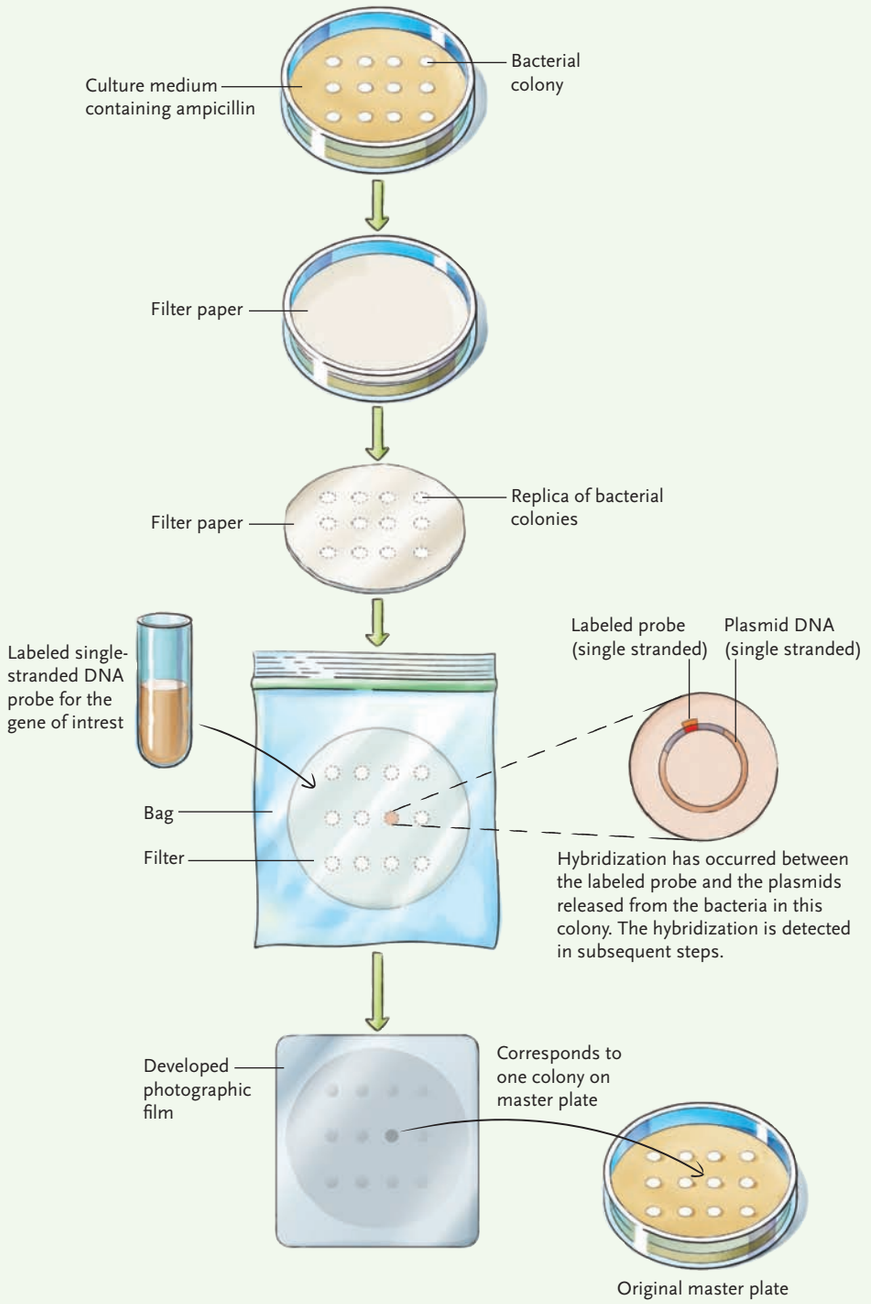
1. Prepare master plates of white colonies detected in the blue-white screening step of Figure 18.4. These colonies contain bacteria with recombinant plasmids. Hundreds or thousands of colonies can be screened for the gene of interest by using many master plates.

2. Lay a special filter paper on the plate to pick up some cells from each colony. This produces a replica of the colony pattern on the filter.

3. Treat the filter to break open the cells and to denature the released DNA to single strands. The single-stranded DNA sticks to the filter in the same position as the colony from which it was derived.

4. Add a labeled single-stranded probe (DNA or RNA) for the gene of interest and incubate. The label can be radioactive or nonradioactive. If a recombinant plasmid's inserted DNA fragment is complementary to the probe, the two will hybridize, that is, form base pairs. Wash off excess labeled probe.

5. Detect the hybridization event by looking for the labeled tag on the probe. If the probe was radioactively labeled, place the filter against photographic film. The decaying radioactive compound exposes the film, giving a dark spot when the film is developed. Correlate the position of any dark spot on the film to the original colony pattern on the master plate. Isolate the colony and use it to produce large quantities of the gene of interest.



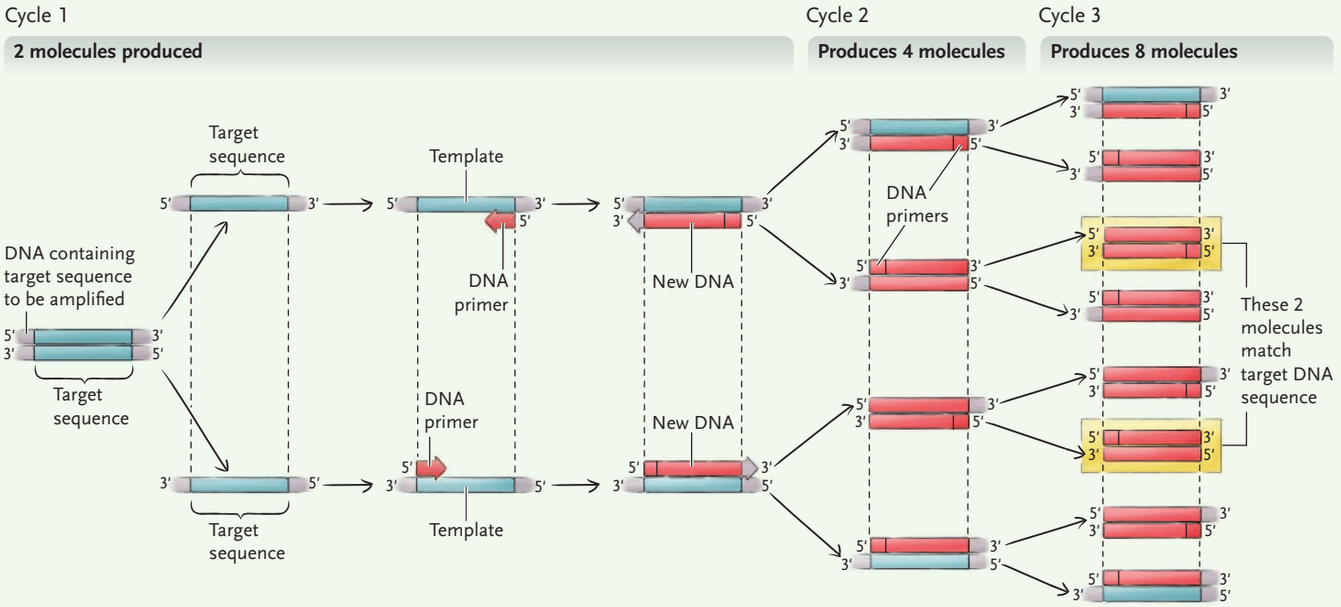
INTERPRETING THE RESULTS: DNA hybridization with a labeled probe enables a researcher to home in on a sequence of interest. If the probe is for a particular gene, it allows the specific identification of a colony containing bacteria with recombinant plasmids carrying that gene. The specificity of the method depends directly on the probe used. The same collection of bacterial clones can be used again and again to search for recombinant plasmids carrying different genes or different plasmids of interest simply by changing the probe used in the experiment.

Figure 18.6 Research Method

The Polymerase Chain Reaction (PCR)

PURPOSE: To amplify—produce large numbers of copies of—a target DNA sequence in the test tube without cloning.

PROTOCOL: A polymerase chain reaction mixture has four key elements: **(1)** the DNA with the target sequence to be amplified; **(2)** a pair of DNA primers, one complementary to one end of the target sequence and the other complementary to the other end of the target sequence; **(3)** the four nucleoside triphosphate precursors for DNA synthesis (dATP, dTTP, dGTP, and dCTP); and **(4)** DNA polymerase. Since PCR uses high temperatures for some of the steps, a heat-stable DNA polymerase is used, typically one isolated from a microorganism that grows in a high-temperature area such as a thermal pool or near a deep-sea vent.



1. Denaturation: Heat DNA containing target sequence to denature it to single strands.
2. Annealing: Cool the mixture to allow the two primers to anneal to their complementary sequences at the two ends of the target sequence.
3. Heat to the optimal temperature for DNA polymerase to extend the primers, using the four nucleoside triphosphate precursors to make complementary copies of the two template strands. This completes cycle 1 of PCR; the end result is two molecules.
4. Repeat the same steps of denaturation, annealing of primers, and extension in cycle 2, producing a total of four molecules.
5. Repeat the same steps in cycle 3, producing a total of eight molecules. Two of the eight match the exact length of the target DNA sequence (highlighted in yellow).

INTERPRETING THE RESULTS: After three cycles, PCR produces a pair of molecules matching the target sequence. Subsequent cycles amplify these molecules to the point where they outnumber all other molecules in the reaction by many orders of magnitude.

produces an extremely large number of copies of a specific DNA sequence from a DNA mixture without having to clone the sequence in a host organism. The process is called *amplification* because it increases the amount of DNA to the point where it can be analyzed or manipulated easily. Developed in 1983 by Kary B. Mullis and F. Faloona at Cetus Corporation (Emeryville, CA), PCR has become one of the most important tools in modern molecular biology, finding wide application in all areas of biology. Mullis received a Nobel Prize in 1993 for his role in the development of PCR.

How PCR is performed is shown in **Figure 18.6.** PCR essentially is DNA replication, but a special case in which a DNA polymerase replicates just a portion of a DNA molecule rather than the whole molecule. PCR takes advantage of a characteristic common to all DNA polymerases: these enzymes add nucleotides only to the end of an existing chain called the *primer* (see Section 14.3). For replication to take place, a primer therefore must be in place, base-paired to the template chain at which replication is to begin. By cycling 20 to 30 times through a series of steps, PCR amplifies the target sequence, producing millions of copies.

Since the primers used in PCR are designed to bracket only the sequence of interest, the cycles replicate only this sequence from a mixture of essentially any DNA molecules. Thus PCR not only finds the "needle in the haystack" among all the sequences in a mixture, but also makes millions of copies of the "needle"—the DNA sequence of interest. Usually no further purification of the amplified sequence is necessary.

The characteristics of PCR allow extremely small DNA samples to be amplified to concentrations high enough for analysis. PCR is used, for example, to produce enough DNA for analysis from the root of a single human hair, or from a small amount of blood, semen, or saliva, such as the traces left at the scene of a crime. It is also used to extract and multiply DNA sequences from skeletal remains; ancient sources such as mammoths, Neanderthals, and Egyptian mummies; and, in rare cases, from amber-entombed fossils, fossil bones, and fossil plant remains.

A successful outcome of PCR is shown by analyzing a sample of the amplified DNA using **agarose gel electrophoresis** to see if the copies are the same length as the target **(Figure 18.7).** Gel electrophoresis is a technique by which DNA, RNA, or protein molecules are separated in a gel subjected to an electric field. The type of gel and the conditions used vary with the experiment, but in each case the gel functions as a molecular sieve to separate the macromolecules based on size, electrical charge, or other properties. For separating large DNA molecules, such as those typically produced by PCR, a gel made of agarose, a natural molecule isolated from seaweed, is used because of its large pore size.

For PCR experiments, the size of the amplified DNA is determined by comparing the position of the DNA band with the positions of DNA fragments of known size separated on the gel at the same time. If that size matches the predicted size for the target DNA, PCR is deemed successful. In some cases, such as DNA from ancient sources, a size prediction may not be possible; in this case, agarose gel electrophoresis analysis simply indicates whether there was DNA in the sample that could be amplified.

The advantages of PCR have made it the technique of choice for researchers, law enforcement agencies, and forensic specialists whose primary interest is in the amplification of specific DNA fragments up to a practical maximum of a few thousand base pairs. Cloning remains the technique of choice for amplification of longer fragments. The major limitation of PCR relates to the primers. In order to design a primer for PCR, the researcher must first have sequence information about the target DNA. By contrast, cloning can be used to amplify DNA of unknown sequence.

Study Break

1. What features do restriction enzymes have in common? How do they differ?
2. Plasmid cloning vectors are one type of cloning vector that can be used with *E. coli* as a host organism. What features of a plasmid cloning vector make it useful for constructing and cloning recombinant DNA molecules?
3. What is a cDNA library, and from what cellular material is it derived? How does a cDNA library differ from a genomic library?
4. What information and materials are needed to amplify a region of DNA using PCR?

18.2 Applications of DNA Technologies

The ability to clone pieces of DNA—genes, especially—and to amplify specific segments of DNA by PCR revolutionized biology. These and other DNA technologies are now used for research in all areas of biology, including cloning genes to determine their structure, function, and regulation of expression; manipulating genes to determine how their products function in cellular or developmental processes; and identifying differences in DNA sequences among individuals in ecological studies. The same DNA technologies also have practical applications, including medical and forensic detection, modification of animals and plants, and the manufacture of commercial products. In this section, case studies provide examples of how the techniques are used to answer questions and solve problems.

DNA Technologies Are Used in Molecular Testing for Many Human Genetic Diseases

Many human genetic diseases are caused by defects in enzymes or other proteins that result from mutations at the DNA level. Once scientists have identified the specific mutations responsible for human genetic diseases, they can often use DNA technologies to develop molecular tests for those diseases. One example is sickle-cell disease (see *Why It Matters* in Chapter 12,

Figure 18.7 Research Method

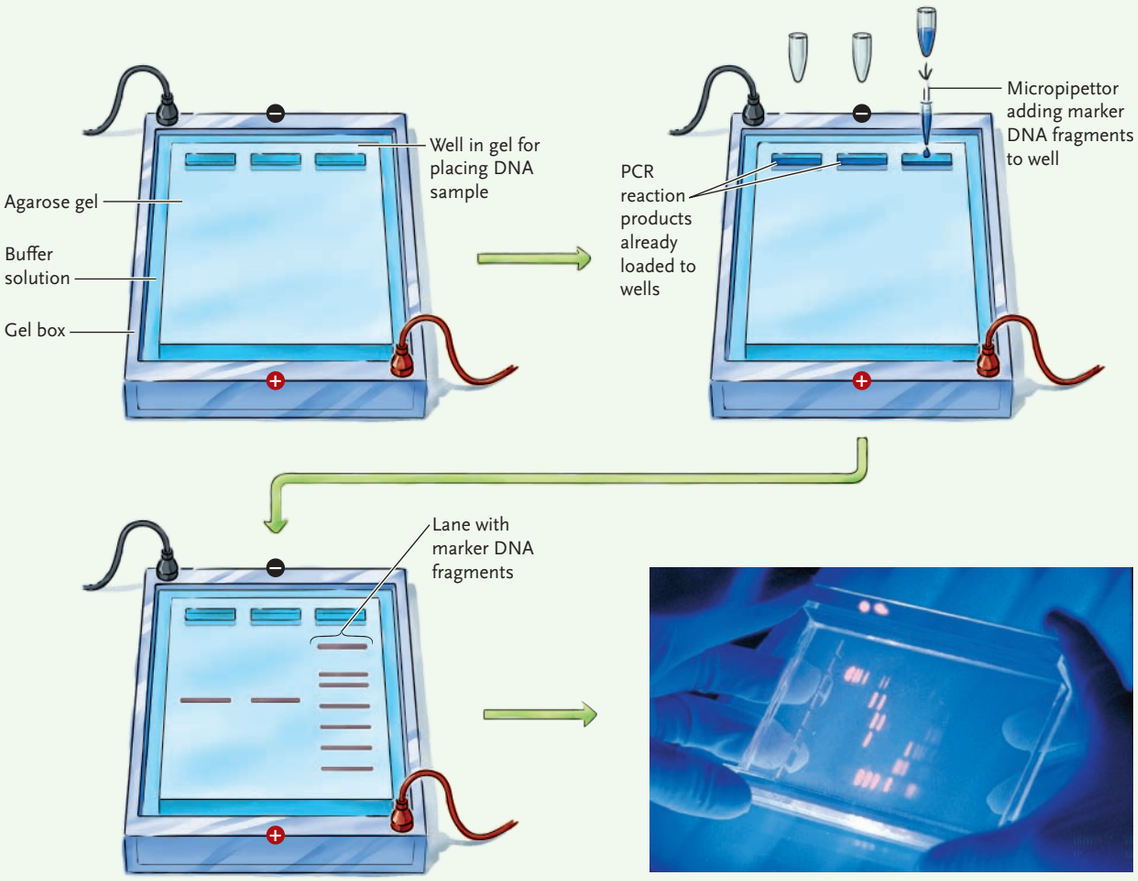
Separation of DNA Fragments by Agarose Gel Electrophoresis

PURPOSE: Gel electrophoresis separates DNA molecules, RNA molecules, or proteins according to their sizes, electrical charges, or other properties through a gel in an electric field. Different gel types and conditions are used for different molecules and types of applications. A common gel for separating large DNA fragments is made of agarose.

PROTOCOL:

1. Prepare a gel consisting of a thin slab of agarose and place it in a gel box in between two electrodes. The gel has wells for placing the DNA samples to be analyzed. Add buffer to cover the gel.

2. Load DNA sample solutions, such as PCR products, into wells of the gel, alongside a well loaded with marker DNA fragments of known sizes.



Damon Biotech, Inc.

3. Apply an electric current to the gel; DNA fragments are negatively charged, so they migrate to the positive pole. Shorter DNA fragments migrate faster than longer DNA fragments. At the completion of the separation, DNA fragments of the same length have formed bands in the gel. At this point, the bands are invisible.

4. Stain the gel with a dye that binds to DNA. The dye fluoresces under UV light, enabling the DNA bands to be seen and photographed. An actual gel showing separated DNA bands stained and visualized this way is shown.

INTERPRETING THE RESULTS: Agarose gel electrophoresis separates DNA fragments according to their length. The lengths of the DNA fragments being analyzed are determined by measuring their migration distances and comparing those distances to a calibration curve of the migration distances of the marker bands, which have known length. For PCR, agarose gel electrophoresis shows whether DNA of the correct length was amplified. For restriction enzyme digests, this technique shows whether fragments are produced as expected.

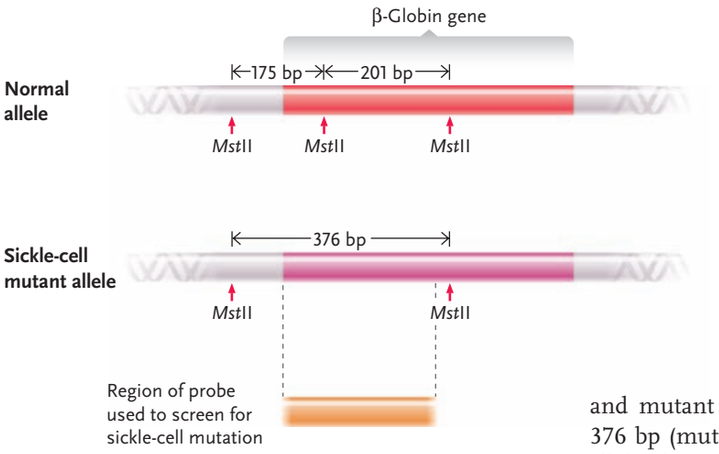


Figure 18.8
Restriction site differences between the normal and sickle-cell mutant alleles of the β-globin gene. The figure shows a DNA segment that can be used as a probe to identify these alleles in subsequent analysis (see Figure 18.9).

Section 12.2, and Section 13.4). People with this disease are homozygous for a DNA mutation that affects hemoglobin, the oxygen carrying molecule of the blood. Hemoglobin consists of two copies each of the α-globin and β-globin polypeptides. The mutation, which is in the β-globin gene, alters one amino acid in the polypeptide. As a consequence, the function of hemoglobin is significantly impaired in individuals homozygous for the mutation (who have sickle-cell anemia), and mildly impaired in individuals heterozygous for the mutation (who have sickle-cell trait).

The sickle-cell mutation changes a restriction site in the DNA **(Figure 18.8).** Three restriction sites for *Mst*II are associated with the normal β-globin gene, two within the coding sequence of the gene and one upstream of the gene. The sickle-cell mutation eliminates the middle site of the three. Cutting the β-globin gene with *Mst*II produces two DNA fragments from the normal gene and one fragment from the mutated gene (see Figure 18.8). Restriction-enzyme-generated DNA fragments of different lengths from the same region of the genome such as in this example are known as **restriction fragment length polymorphisms** (RFLPs, pronounced "riff-lips").

RFLPs typically are analyzed using **Southern blot analysis** (named after its inventor, researcher Edward Southern) **(Figure 18.9).** In this technique, genomic DNA is digested with a restriction enzyme, and the DNA fragments are separated using agarose gel electrophoresis. The fragments are then transferred—blotted—to a filter paper, and a labeled probe is used to identify a DNA sequence of interest from among the many thousands of fragments on the filter paper.

Analyzing DNA for the sickle-cell mutation by *Mst*II digestion and Southern blot analysis is straightforward (see Figure 18.9). An individual with sickle-cell disease will have one DNA band of 376 bp detected by the probe (lane A), a healthy individual will have two DNA bands of 175 and 201 bp (lane B), and an individual with sickle-cell trait (heterozygous for normal and mutant alleles) will have three DNA bands of 376 bp (mutant allele), and 201 and 175 bp (normal allele) (lane C). The same probe detects all three RFLP fragments by binding to all or part of the sequence.

Restriction enzyme digestion and Southern blot analysis may be used to test for a number of other human genetic diseases, including phenylketonuria and Duchenne muscular dystrophy. In some cases, restriction enzyme digestion is combined with PCR for a quicker, easier analysis. The gene or region of the gene with the restriction enzyme variation is first amplified using PCR, and the amplified DNA is then cut with the diagnostic restriction enzyme. Amplification produces enough DNA so that separation by size on an agarose gel produces clearly visible bands, positioned according to fragment length. Researchers can then determine whether the fragment lengths match a normal or abnormal RFLP pattern. This method eliminates the need for a probe or for Southern blotting.

DNA Fingerprinting Is Used to Identify Human Individuals as well as Individuals of Other Species

Just as each human has a unique set of fingerprints, each also has unique combinations and variations of DNA sequences (with the exception of identical twins) known as *DNA fingerprints*. **DNA fingerprinting** is a technique used to distinguish between individuals of the same species using DNA samples. Invented by Sir Alec Jeffreys in 1985, DNA fingerprinting has become a mainstream technique for distinguishing human individuals, notably in forensics and paternity testing. And, as *Why It Matters* indicated, the technique is applicable to all kinds of organisms. We focus on humans in the following discussion.

DNA Fingerprinting Principles. In DNA fingerprinting, scientists use molecular techniques, most typically PCR, to analyze DNA variations at various loci in the genome. In the United States, 13 loci in noncoding regions of the genome are the standards for PCR analysis. Each locus is an example of a *short tandem repeat* (STR) sequence, meaning that it has a short sequence of DNA repeated

Figure 18.9 Research Method

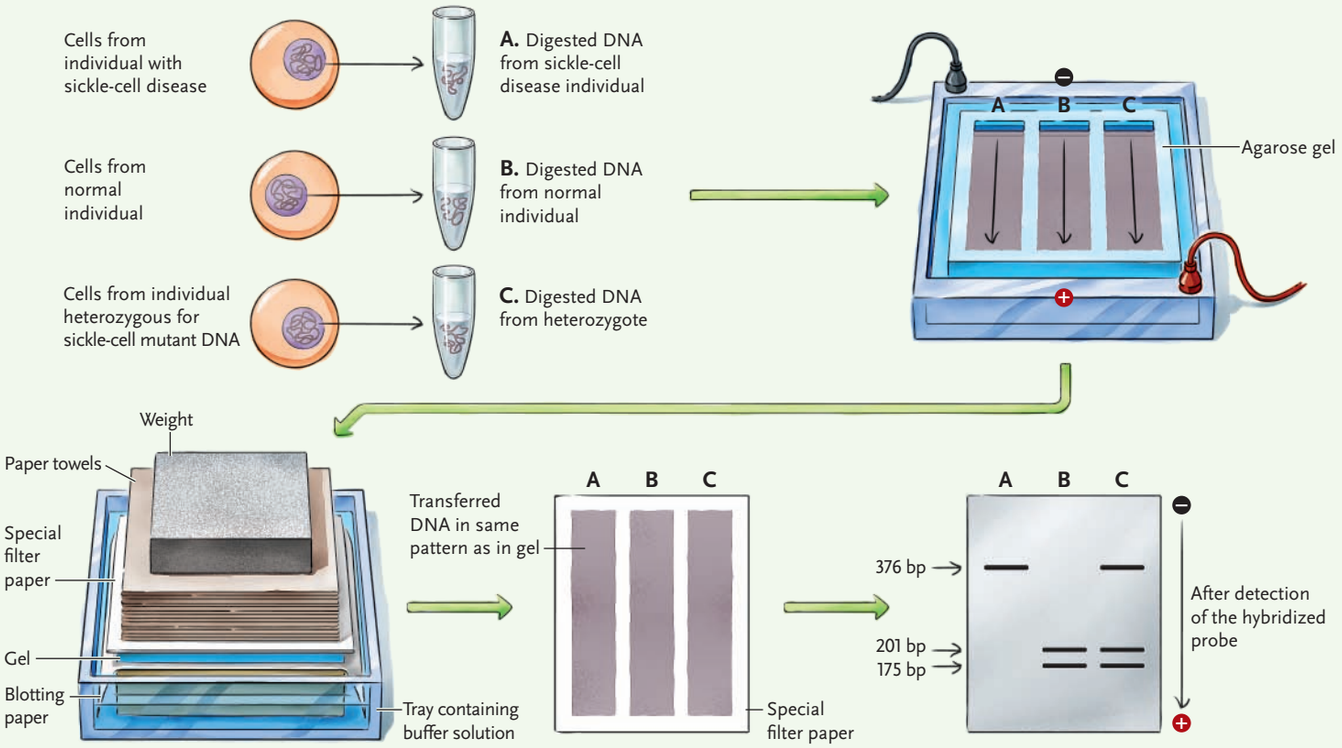
Southern Blot Analysis

PURPOSE: The Southern blot technique allows researchers to identify DNA fragments of interest after separating DNA fragments on a gel. One application is to compare different samples of genomic DNA cut with a restriction enzyme to detect specific restriction fragment length polymorphisms. Here the technique is used to distinguish between individuals with sickle-cell disease, individuals with sickle-cell trait, and normal individuals.

PROTOCOL:

1. Isolate genomic DNA and digest with a restriction enzyme. Here, genomic DNA is isolated from three individuals: A, sickle-cell disease (homozygous for the sickle-cell mutant allele); B, normal (homozygous for the normal allele); and C, sickle-cell trait (heterozygote for sickle-cell mutant allele). Digest the DNA with *Mst*II.

2. Separate the DNA fragments by agarose gel electrophoresis. The thousands of differently sized DNA fragments generated results in a smear of DNA down the length of each lane in the gel, which can be seen after staining the DNA. (Gel electrophoresis and gel staining are shown in Figure 18.7).



3. Hybridization with a labeled DNA probe to identify DNA fragments of interest cannot be done directly with an agarose gel. Edward Southern devised a method to transfer the DNA fragments from a gel to a special filter paper. First, treat the gel with a solution to denature the DNA to single strands. Next, place the gel on a piece of blotting paper with ends of the paper in the buffer solution and place the special filter paper on top of the gel. Capillary action wicks the buffer solution in the tray up the blotting paper, through the gel and special filter paper, and into the weighted stack of paper towels on top of the gel. The movement of the solution transfers—blots—the single-stranded DNA fragments to the filter paper, where they stick. The pattern of DNA fragments is the same as it was in the gel.

4. To home in on a particular region of the genome, use DNA hybridization with a labeled probe. That is, incubate a labeled, single-stranded probe with the filter and, after washing off excess probe, detect hybridization of the probe with DNA fragments on the filter. For a radioactive probe, the filter is placed against photographic film, which, after development will show a band or bands where the probe hybridized. In this experiment, the probe is a cloned piece of DNA from the area shown in Figure 18.8 that can bind to all three of the *Mst*II fragments of interest.

INTERPRETING THE RESULTS: The hybridization result indicates that the probe has identified a very specific DNA fragment or fragments in the digested genomic DNA. The RFLPs for the β-globin gene can be seen in Figure 18.8. DNA from the sickle-cell disease individual cut with *Mst*II results in a single band of 376 bp detected by the probe, while DNA from the normal individual results in two bands of 201, and 175 bp. DNA from a sickle-cell trait heterozygote results in three bands of 376 bp (from the sickle-cell mutant allele), and 201 and 175 bp (the latter two from the normal allele). This type of analysis in general is useful for distinguishing normal and mutant alleles of genes where the mutation involved alters a restriction site.

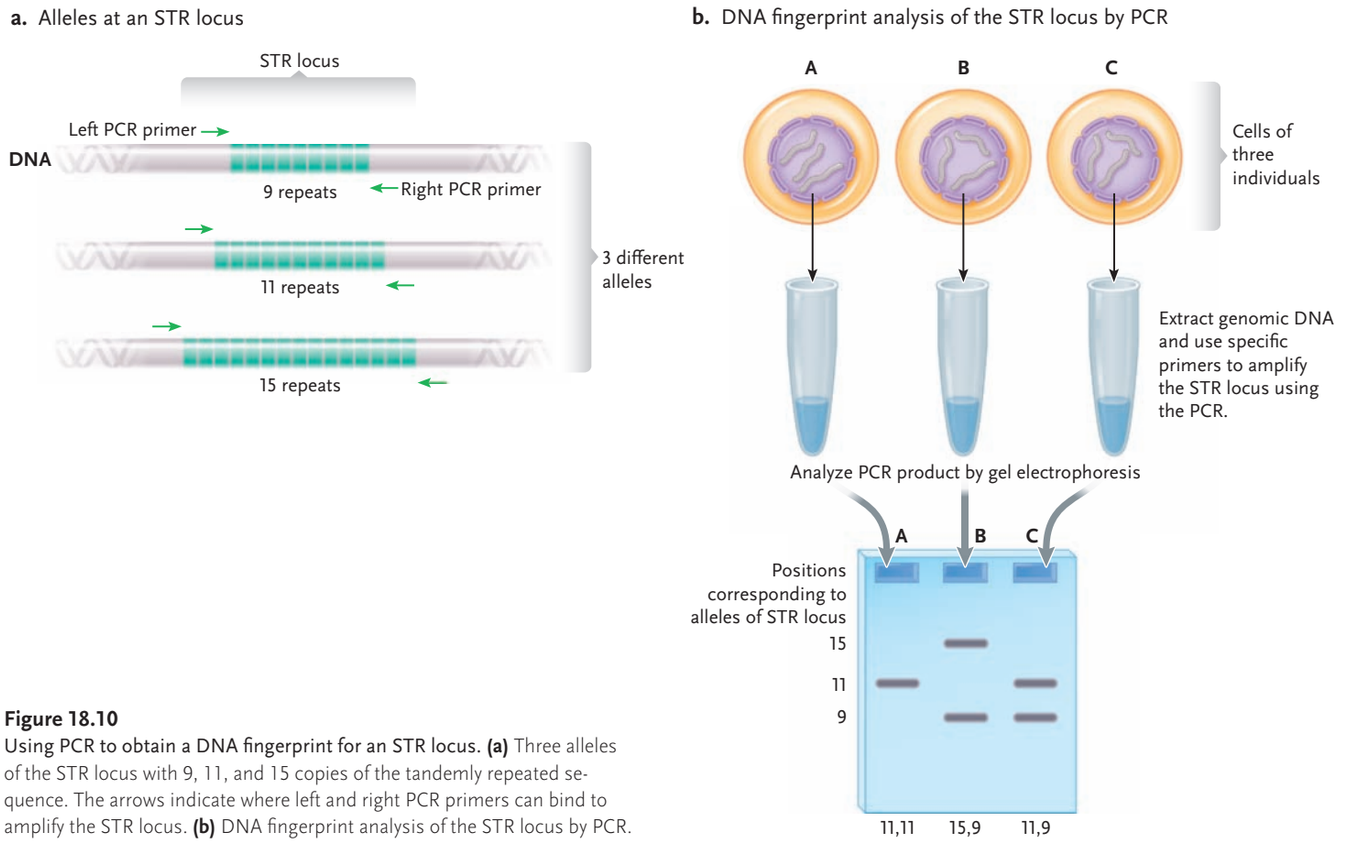


Figure 18.10
Using PCR to obtain a DNA fingerprint for an STR locus. **(a)** Three alleles of the STR locus with 9, 11, and 15 copies of the tandemly repeated sequence. The arrows indicate where left and right PCR primers can bind to amplify the STR locus. **(b)** DNA fingerprint analysis of the STR locus by PCR.

in series, with each repeat about 3 to 5 bp. Each locus has a different repeated sequence, and the number of repeats varies among individuals in a population. For example, one STR locus has the sequence AGAT repeated between 8 and 20 times. As a further source of variation, a given individual is either homozygous or heterozygous for an STR allele; perhaps you are homozygous for the eleven-repeat allele, or heterozygous for a nine-repeat allele and a fifteen-repeat allele. Likely your DNA fingerprint for this locus is different from most of the others in your class. Because each individual has an essentially unique combination of alleles (identical twins are the exception), analysis of multiple STR loci can discriminate between DNA of different individuals.

Figure 18.10 illustrates how PCR is used to obtain a DNA fingerprint for a theoretical STR locus with three alleles of 9, 11, and 15 tandem repeats (Figure 18.10a). Using primers that flank the STR locus, the locus is amplified from genomic DNA using PCR, and the PCR products are analyzed by gel electrophoresis (Figure 18.10b). The number of bands on the gel and the sizes of the DNA in the bands show the STR alleles that were amplified. One band indicates that the individual was homozygous for an STR allele with a particular number of repeats, while two bands indicates the individual is heterozygous for two STR alleles with different numbers of repeats. In the result shown in Figure 18.10b, the A individual is homozygous for a 11-repeat allele (designated 11,11), B is heterozygous for a 15-repeat allele and a 9-repeat allele (15,9), and C is heterozygous for the 11-repeat allele and the 9-repeat allele (11,9).

DNA Fingerprinting in Forensics. DNA fingerprints are routinely used to identify criminals or eliminate innocent persons as suspects in legal proceedings. For example, a DNA fingerprint prepared from a hair found at the scene of a crime, or from a semen sample, might be compared with the DNA fingerprint of a suspect to link the suspect with the crime. Or, a DNA fingerprint of blood found on a suspect's clothing or possessions might be compared with the DNA fingerprint of a victim. Typically, the evidence is presented in terms of the probability that the particular DNA sample could have come from a random individual. Hence the media report probability values, such as one in several million, or in several billion, that a person other than the accused could have left his or her DNA at the crime scene.

Although courts initially met with legal challenges to the admissibility of DNA fingerprints, experience has shown that they are highly dependable as a line of evidence if DNA samples are collected and prepared with care and if a sufficient number of polymorphic loci are examined. There is always concern, though, about the possibility of contamination of the sample with DNA from another source during the path from

crime scene to forensic lab analysis. Moreover, in some cases criminals themselves have planted fake DNA samples at crime scenes to confuse the investigation.

There are many examples of the use of DNA fingerprinting to identify a criminal. For example, in a case in England, the DNA fingerprints of more than 4000 men were made during an investigation of the rape and murder of two teenage girls. The results led to the release of a man wrongly imprisoned for the crimes and to the confession and conviction of the actual killer. And the application of DNA fingerprinting techniques to stored forensic samples has led to the release of a number of persons wrongly convicted for rape or murder.

DNA Fingerprinting in Testing Paternity and Establishing Ancestry. DNA fingerprints are also widely used as evidence of paternity because parents and their children share common alleles in their DNA fingerprints. That is, each child receives one allele of each locus from one parent and the other allele from the other parent. A comparison of DNA fingerprints for a number of loci can prove almost infallibly whether a child has been fathered or mothered by a given person. DNA fingerprints have also been used for other investigations, such as confirming that remains discovered in a remote region of Russia were actually those of Czar Nicholas II and members of his family, murdered in 1918 during the Russian revolution.

DNA fingerprinting is also widely used in studies of other organisms, including other animals, plants, and bacteria. Examples include testing for pathogenic *E. coli* in food sources such as hamburger meat, investigating cases of wildlife poaching, detecting genetically modified organisms among living organisms or in food, and comparing the DNA of ancient organisms with present-day descendants.

Genetic Engineering Uses DNA Technologies to Alter the Genes of a Cell or Organism

We have seen the many ways scientists use DNA technologies to ask, and answer, questions that were once completely inaccessible. Genetic engineering goes beyond gathering information; it is the use of DNA technologies to modify genes of a cell or organism. The goals of genetic engineering include using prokaryotes, fungi, animals, and plants as factories for the production of proteins needed in medicine and scientific research; correcting hereditary disorders; and improving animals and crop plants of agricultural importance. In many of these areas genetic engineering has already been spectacularly successful. The successes and potential benefits of genetic engineering, however, are tempered by ethical and social concerns about its use, along with the fear that the methods may produce toxic or damaging foods, or release dangerous and uncontrollable organisms to the environment.

Genetic engineering uses DNA technologies of the kind discussed already in this chapter. DNA—perhaps a modified gene—is introduced into target cells of an organism. Organisms that have undergone a gene transfer are called **transgenic**, meaning that they have been modified to contain genetic information—the *transgene*—from an external source.

The following sections discuss examples of applications of genetic engineering to bacteria, animals, and plants, and assess major controversies arising from these projects.

Genetic Engineering of Bacteria to Produce Proteins. Transgenic bacteria have been made, for example, to make proteins for medical applications, break down toxic wastes such as oil spills, produce industrial chemicals such as alcohols, and process minerals. *E. coli* has been the organism of choice for many of these applications of DNA technologies.

Using *E. coli* to make a protein from a foreign source is conceptually straightforward. First, the gene for the protein is cloned from the appropriate organism. Then the gene is inserted in a special type of bacterial plasmid called an *expression vector,* which has a bacterial promoter adjacent to the restriction site used for inserting the gene. The resulting recombinant plasmid is transformed into *E. coli,* which transcribes the gene and translates the resulting mRNA to make the desired protein. The protein is either extracted from the bacterial cells and purified or, if the protein is secreted, it is purified from the culture medium.

For example, *E. coli* bacteria have been genetically engineered to make the human hormone insulin; the commercial product is called humulin. Insulin is required by persons with some forms of diabetes. Humulin is a perfect copy of the human insulin hormone. Many other proteins, including human growth hormone to treat human growth disorders, tissue plasminogen activator to dissolve blood clots that cause heart attacks, and a vaccine against hoof-and-mouth disease of cattle (a highly contagious and sometimes fatal viral disease), have been developed for commercial production in bacteria by similar methods.

Although they offer many benefits, genetically engineered bacteria pose the risk that they may be released accidentally into the environment where any adverse effects are currently unknown. Scientists minimize the danger of accidental release by growing the bacteria in laboratories that follow appropriate biosafety protocols. In addition, the bacterial strains used typically are genetically modified so that they will not survive outside of the growth media used in the laboratory.

Genetic Engineering of Animals. Many animals, including fruit flies, fish, mice, pigs, sheep, goats, and cows, have been altered successfully by genetic engineering. There are many purposes for these altera-

tions, including basic research, correcting genetic disorders in humans and other mammals, and producing pharmaceutically important proteins.

Genetic Engineering Methods for Animals. Several methods are used to introduce a gene of interest into animal cells. The gene may be introduced into *germ-line cells,* which develop into sperm or eggs and thus enable the introduced gene to be passed from generation to generation. Or, the gene may be introduced into *somatic* (body) *cells,* differentiated cells that are not part of lines producing sperm or eggs, in which case the gene is not transmitted from generation to generation.

Germ-line cells of embryos are often used as targets for introducing genes, particularly in mammals **(Figure 18.11).** The treated cells are then cultured in quantity and reintroduced into early embryos. If the technique is successful, some of the introduced cells become founders of cell lines that develop into eggs or sperm with the desired genetic information integrated into their DNA. Individuals produced by crosses using the engineered eggs and sperm then contain the introduced sequences in all their cells. Several genes have been introduced into the germ lines of mice by this approach, resulting in permanent, heritable changes in the engineered individuals.

A related technique involves introducing desired genes into *stem cells,* which are capable of differentiating into almost any adult cell type and tissue. Stem cells that have taken up the gene are then injected into an early embryo, where they differentiate into a variety of tissues along with cells of the embryo itself, including sperm and egg cells. Males and females are then bred, leading to offspring that are either homozygotes, containing two copies of the introduced gene, or heterozygotes, containing one introduced gene and one gene that was native to the embryo receiving the engineered stem cells.

Introduction of genes into stem cells has been performed mostly in mice. One of the highly useful results is the production of a "knockout mouse," a homozygous recessive that receives two copies of a gene altered to a nonfunctional state, and thus has no functional copies. The effect of the missing gene on the knockout mouse is a clue to the normal function of the gene. In some cases, knockout mice are used to model human genetic diseases.

For introducing genes into somatic cells, typically somatic cells are removed from the body, cultured, and then transformed with DNA containing the transgene. The modified cells are then reintroduced into the body where the transgene functions. Because germ cells and their products are not involved, the transgene remains in the individual and is not passed to offspring.

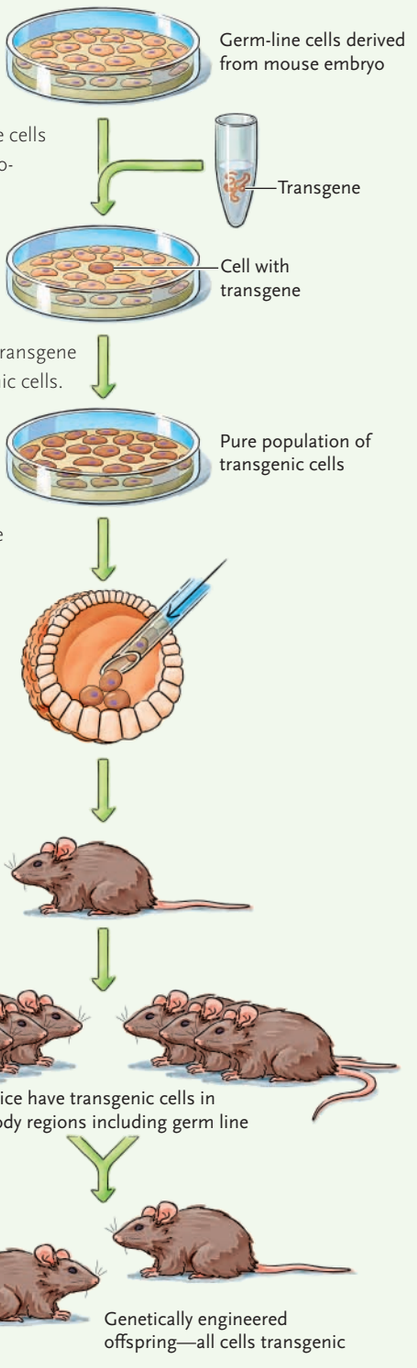
Gene Therapy: Correcting Genetic Disorders. The path to **gene therapy**—correcting genetic disorders—in humans began with experiments using mice. In 1982, Richard Palmiter at the University of Washington,

Figure 18.11 Research Method

Introduction of Genes into Mouse Embryos Using Embryonic Germ-Line Cells

PURPOSE: To make a transgenic animal that can transmit the transgene to offspring. The embryonic germ-line cells that receive the transgene develop into the reproductive cells of the animal.

PROTOCOL:



INTERPRETING THE RESULTS: The result of the breeding is some offspring in which all cells are transgenic—a genetically engineered animal has been produced.

R. L. Brinster, R. E. Hammer, School of Veterinary Medicine, University of Pennsylvania

Figure 18.12
A genetically engineered giant mouse (right) produced by the introduction of a rat growth hormone gene into the animal. A mouse of normal size is on the left.

Ralph Brinster of the University of Pennsylvania, and their colleagues injected a growth hormone gene from rats into fertilized mouse eggs and implanted the eggs into a surrogate mother. She gave birth to some normal-sized mouse pups that grew faster than normal and became about twice the size of their normal litter mates. These *giant mice* **(Figure 18.12)** attracted extensive media attention from around the world.

Palmiter and Brinster next attempted to cure a genetic disorder by gene therapy. In this experiment, they were able to correct a genetic growth hormone deficiency that produces dwarf mice. They introduced a normal copy of the growth hormone gene into fertilized eggs taken from mutant dwarf mice and implanted them into a surrogate mother. The transgenic mouse pups grew to slightly larger than normal, demonstrating that the genetic defect in those mice had been corrected.

This sort of experiment, in which a gene is introduced into germ-line cells of an animal to correct a genetic disorder, is **germ-line gene therapy.** For ethical reasons, germ-line gene therapy is not permitted with humans. Instead, humans are treated with **somatic gene therapy**, in which genes are introduced into somatic cells (as described in the previous section).

The first successful use of somatic gene therapy with a human subject who had a genetic disorder was carried out in the 1990s by W. French Anderson and his colleagues at the National Institutes of Health. The subject was a young girl with *adenosine deaminase deficiency (ADA)*. Without the adenosine deaminase enzyme, white blood cells cannot mature (see Chapter 43); without normally functioning white blood cells, the body's immune response is so deficient that most children with ADA die of infections before reaching puberty. The researchers successfully introduced a functional ADA gene into mature white blood cells isolated from the patient. Those cells were reintroduced into the girl, and expression of the ADA gene provided a temporary cure for her ADA deficiency. The cure was not permanent because mature white blood cells, produced by differentiation of stem cells in the bone marrow, are nondividing cells with a finite life time. Therefore, the somatic gene therapy procedure has to be repeated every few months. Indeed, the subject of this example still receives periodic gene therapy to maintain the necessary levels of the ADA enzyme in her blood. In addition, she receives direct doses of the normal enzyme.

Successful somatic gene therapy has also been achieved for sickle-cell disease. In December 1998, a 13-year-old boy's bone marrow cells were replaced with stem cells from the umbilical cord of an unrelated infant. The hope was that the stem cells would produce healthy bone marrow cells, the source of blood cells. The procedure worked, and the patient has been declared cured of the disease.

However, despite enormous efforts, human somatic gene therapy has not been the panacea people expected. Relatively little progress has been made since the first gene therapy clinical trial for ADA deficiency described, and, in fact, there have been major setbacks. In 1999, for example, a teenage patient in a somatic gene therapy trial died as a result of a severe immune response to the viral vector being used to introduce a normal gene to correct his genetic deficiency. Furthermore, some children in gene therapy trials involving the use of retrovirus vectors to introduce genes into blood stem cells have developed a leukemia-like condition. In short, somatic gene therapy is not yet an effective treatment for human genetic disease, even though the approach has been successful in a number of cases to correct models of human genetic disorders in experimental mammals. Although no commercial human gene therapy product has been approved for use, research and clinical trials continue as scientists try to circumvent the difficulties.

Turning Domestic Animals into Protein Factories. Another successful application of genetic engineering turns animals into pharmaceutical factories for the production of proteins required to treat human diseases or other medical conditions. Most of these *pharming* projects, as they are called, engineer the animals to produce the desired proteins in milk, making the production, extraction, and purification of the proteins harmless to the animals.

One of the first successful applications of this approach was carried out with sheep engineered to produce a protein required for normal blood clotting in humans. The protein, called a *clotting factor,* is deficient in persons with one form of hemophilia, who require frequent injections of the factor to avoid bleeding to death from even minor injuries. Using DNA cloning techniques, researchers joined the gene encoding the normal form of the clotting factor to the promoter sequences of the β-lactoglobin gene, which encodes a protein secreted in milk, and introduced it into fertilized eggs. Those cells were implanted into a surrogate mother, and the transgenic sheep born were allowed

to mature. The β-lactoglobin promoter controlling the clotting factor gene became activated in mammary gland cells of females, resulting in the production of clotting factor. The clotting factor was then secreted into the milk. Production in the milk is harmless to the sheep and yields the protein in a form that can easily be obtained and purified.

Other similar projects are under development to produce particular proteins in transgenic mammals. These include a protein to treat cystic fibrosis, collagen to correct scars and wrinkles, human milk proteins to be added to infant formulas, and normal hemoglobin for use as an additive to blood transfusions.

Producing Animal Clones. Making transgenic mammals is expensive and inefficient. And, because only one copy of the transgene typically becomes incorporated into the treated cell, not all progeny of a transgenic animal inherit that gene. Scientists reasoned that an alternative to breeding a valuable transgenic mammal to produce progeny with the transgene would be to clone the mammal. Each clone would be identical to the original, including the expression of the transgene. That this is possible was shown in 1997 when two Scottish scientists, Ian Wilmut and Keith H. S. Campbell of the Roslin Institute, Edinburgh, announced that that they had successfully cloned a sheep from a single somatic cell derived from an adult sheep **(Figure 18.13)**—the first mammalian clone made. For their experiment, the researchers fused a diploid cell derived from the mammary gland of a 6-year-old adult sheep with an unfertilized egg cell from which the nucleus had been removed. Signals from the egg cytoplasm triggered DNA replication and cell division, producing a cluster of cells derived from the mammary gland cell. The cluster was implanted into the uterus of an adult female sheep, where it developed into an embryo that grew to full term and was delivered as an apparently normal lamb, named Dolly. Their cloning success rate, though, was very low—Dolly represents less than 0.4% of the transgenic cells they made. Dolly developed to sexual maturity and produced four normal offspring. She was euthanized at age 6 after contracting a fatal, virus-induced lung disease that her cloners believe was unrelated to the cloning.

After the successful cloning experiment producing Dolly, many additional mammals have been cloned, including mice, goats, pigs, monkeys, rabbits, dogs, a male calf appropriately named Gene, and a domestic cat called *CC* (for *Copy Cat*).

Cloning farm animals has been so successful that several commercial enterprises now provide cloned copies of champion animals. One example is a clone of an American Holstein cow, Zita, who was the U.S. national champion milk producer for many years. Animal breeders estimate that there are now more than 100 cloned animals on American farms, and breeders plan to produce entire herds if government approval is granted.

PA News Photo Library/AP Wide World Photos

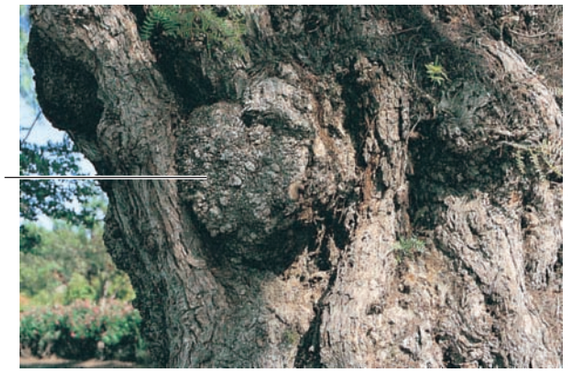
Figure 18.13
Dolly, the cloned sheep.

The cloning of domestic animals has its drawbacks. Many cloning attempts fail, leading to the death of the transplanted embryos. Cloned animals often suffer from health defects from conditions such as birth defects and poor lung development. Genes may be lost during the cloning process or may be expressed abnormally in the cloned animal. For example, molecular studies have shown that the expression of perhaps hundreds of genes in the genomes of clones is regulated abnormally.

Genetic Engineering of Plants. Genetic engineering of plants has led to increased resistance to pests and disease; greater tolerance to heat, drought, and salinity; greater crop yields; faster growth; and resistance to herbicides. Another aim is to produce seeds with higher levels of amino acids. The essential amino acid lysine, for example, is present only in limited quantities in cereal grains such as wheat, rice, oats, barley, and corn; the seeds of legumes such as beans, peas, lentils, soybeans, and peanuts are deficient in the essential amino acids methionine or cysteine. Increasing the amounts of the deficient amino acids in plant seeds by genetic engineering would greatly improve the diet of domestic animals and human populations that rely on seeds as a primary food source. Efforts are also under way to increase the content of vitamins and minerals in crop plants.

Other possibilities for plant genetic engineering include plant pharming to produce pharmaceutical products. Plants are ideal for this purpose, because they are primary producers at the bottom rung of the food chain and can be grown in huge numbers with maximum conservation of the sun's energy captured in photosynthesis.

Some plants, such as *Arabidopsis,* tobacco, potato, cabbage, and carrot, have special advantages for genetic engineering because individual cells can be removed from an adult, altered by the introduction of a desired gene, and then grown in cultures into a multicellular mass of cloned cells called a *callus*. Subsequently, roots,



Stephen Wolfe, *Molecular and Cellular Biology*

Figure 18.14
A crown gall tumor on the trunk of a California pepper tree. The tumor, stimulated by genes introduced from the bacterium *Rhizobium radiobacter*, is the bulbous, irregular growth extending from the trunk.

stems, and leaves develop in the callus, forming a young plant that can then be grown in containers or fields by the usual methods. In the plant, each cell contains the introduced gene. The gametes produced by the transgenic plants can then be used in crosses to produce offspring, some of which will have the transgene, as in the similar experiments with animals.

Methods Used to Insert Genes into Plants. Genes are inserted into plant cells by several techniques. A commonly used method takes advantage of a natural process that causes crown gall disease, which is characterized by bulbous, irregular growths—tumors, essentially—that can develop at wound sites on the trunks and limbs of deciduous trees **(Figure 18.14).** Crown gall disease is caused by the bacterium *Rhizobium radiobacter* (formerly *Agrobacterium tumefaciens,* recently reclassified on the basis of genome analysis). This bacterium contains a large, circular plasmid called the **Ti (tumor inducing) plasmid.** The interaction between the bacterium and a plant cell it infects stimulates the excision of a segment of the Ti plasmid called *T DNA* (for transforming DNA), which then integrates into the plant cell's genome. Genes on the T DNA are then expressed; the products stimulate the transformed cell to grow and divide and therefore to produce a tumor. The tumors provide essential nutrients for the bacterium. The Ti plasmid is used as a vector for making transgenic plants, in much the same way as bacterial plasmids are used as vectors to introduce genes into bacteria **(Figure 18.15).**

Successful Plant Genetic Engineering Projects. An early visual demonstration of the successful use of genetic engineering techniques to produce a transgenic plant is the glowing tobacco plant **(Figure 18.16).** The transgenic plant contained luciferase, the gene for the firefly enzyme. When the plant was soaked in the substrate for the enzyme, it became luminescent.

The most widespread application of genetic engineering of plants involves the production of transgenic crops. Thousands of such crops have been developed and field tested, and many have been approved for commercial use. If you examine the processed plant-based foods at a national supermarket chain, you will likely find that at least two-thirds contain transgenic plants.

In many cases, plants are modified to make them resistant to insect pests, viruses, or herbicides. Crops modified for insect resistance include corn, cotton, and potatoes. The most common approach to making plants resistant to insects is to introduce the gene from the bacterium *Bacillus thuringiensis* that encodes the *Bt* toxin, an organic pesticide. This toxin has been used in powder form to kill insects in agriculture for many years, and now transgenic plants making their own *Bt* toxin are resistant to specific groups of insects that feed on them. Millions of acres of crop plants planted in the United States, amounting to about 70% of the nation's agricultural acreage, are now *Bt*-engineered varieties.

Virus infections cause enormous crop losses worldwide. Transgenic crops that are virus-resistant would be highly valuable to the agricultural community. There is some promise in this area. By some unknown process, transgenic plants expressing certain viral proteins become resistant to infections by whole viruses that contain those same proteins. Two virus-resistant genetically modified crops made so far are papaya and squash.

Several crops have also been engineered to become resistant to herbicides. For example, *glyphosate* (commonly known by its brand name, Roundup) is a highly potent herbicide that is widely used in weed control. The herbicide works by inhibiting a particular enzyme in the chloroplast. Unfortunately, it also kills crops. But transgenic crops have been made in which a bacterial form of the chloroplast enzyme has been added to the plants. The bacterial-derived enzyme is not affected by Roundup, and farmers that use these herbicide-resistant crops can spray fields of crops to kill weeds without killing the crops. Now most of the corn, soybean, and cotton plants grown in the United States and many other countries are the genetically engineered, glyphosate-resistant ("Roundup-ready") varieties.

Crop plants are also being engineered to alter their nutritional qualities. For example, a strain of rice plants has been produced with seeds rich in β-carotene, a precursor of vitamin A, as well as iron **(Figure 18.17).** The new rice, which is given a yellow or golden color by the carotene, may provide improved nutrition for the billions of people that depend on rice as a diet staple. In particular, the rice may help improve the nutrition of children younger than age 5 in southeast Asia, 70% of whom suffer from impaired vision because of vitamin A deficiency. *Insights from the Molecular Revolution* describes an experiment in which rice plants were genetically engineered to develop resistance to a damaging bacterial blight.

Plant pharming is also an active area both in university research labs and at biotechnology companies. Plant pharming involves the engineering of transgenic plants to produce medically valuable products. The ap-

proach is one described earlier: the gene for the product is cloned into a cloning vector adjacent to a promoter, in this case one active in plants, and the recombinant DNA molecule is introduced into plants. Products under development include vaccines for various bacterial and viral diseases, protease inhibitors to treat or prevent virus infections, collagen to treat scars and wrinkles, and aprotinin to reduce bleeding and clotting during heart surgery.

In contrast to animal genetic engineering, genetically altered plants have been widely developed and appear to be here to stay as mainstays of agriculture. But, as the next section discusses, both animal and plant genetic engineering have not proceeded without concerns.

DNA Technologies and Genetic Engineering Are a Subject of Public Concern

When recombinant DNA technology was developed in the early 1970s, researchers quickly recognized that in addition to the anticipated many benefits, there might be deleterious outcomes. One key concern at the time was that a bacterium carrying a recombinant DNA molecule might escape into the environment. Perhaps it could transfer that molecule to other bacteria and produce new, potentially harmful, strains. To address these concerns, the U.S. scientists who developed the technology drew up safety guidelines for recombinant DNA research in the United States. Adopted by the National Institutes of Health (NIH), the guidelines listed the precautions to be used in the laboratory when constructing recombinant DNA molecules and included the design and use of host organisms that could survive only in growth media in the laboratory. Since that time countless thousands of experiments involving recombinant DNA molecules have been done in laboratories around the world. Those experiments have shown that recombinant DNA manipulations can be done safely. Over time, therefore, the recombinant DNA guidelines have become more relaxed. Nonetheless, stringent regulations still exist for certain areas of recombinant DNA research that pose significant risk, such as cloning genes from highly pathogenic bacteria or viruses, or gene therapy experiments. In essence, as the risk increases, the research facility must increase its security and it must obtain more levels of approval by peer scientist groups.

Guidelines for genetic engineering also extend to research in several areas that have been the subject of public concern and debate. While the public is concerned little about genetically engineered microorganisms, for example those cleaning up oil spills and hazardous chemicals, it is concerned about possible problems with **genetically modified organisms (GMOs)** used as food. A GMO is a transgenic organism; the majority of GMOs are crop plants. Issues are the safety of GMO-containing food and the possible adverse effects

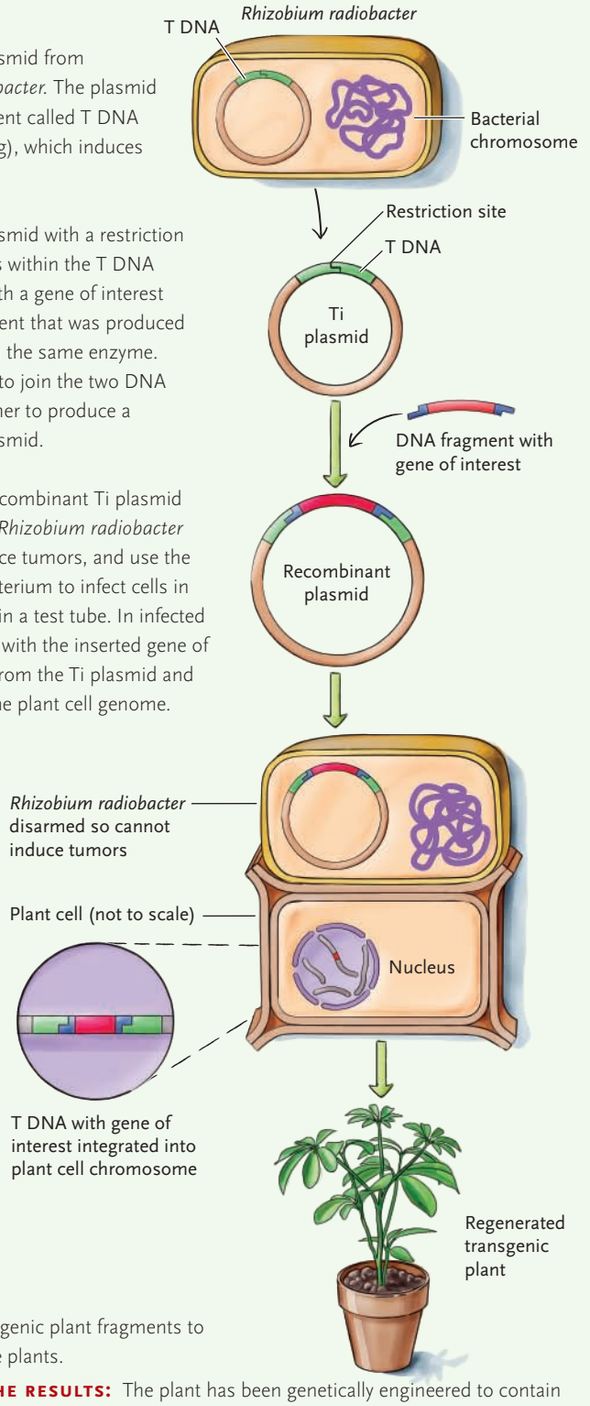
Figure 18.15 Research Method

Using the Ti Plasmid of Rhizobium radiobacter *to Produce Transgenic Plants*

PURPOSE: To make transgenic plants. This technique is one way to introduce a transgene into a plant for genetic engineering purposes.

PROTOCOL:

1. Isolate the Ti plasmid from *Rhizobium radiobacter*. The plasmid contains a segment called T DNA (T = transforming), which induces tumors in plants.
2. Digest the Ti plasmid with a restriction enzyme that cuts within the T DNA segment. Mix with a gene of interest on a DNA fragment that was produced by digesting with the same enzyme. Use DNA ligase to join the two DNA molecules together to produce a recombinant plasmid.
3. Transform the recombinant Ti plasmid into a disarmed *Rhizobium radiobacter* that cannot induce tumors, and use the transformed bacterium to infect cells in plant fragments in a test tube. In infected cells, the T DNA with the inserted gene of interest excises from the Ti plasmid and integrates into the plant cell genome.
4. Culture the transgenic plant fragments to regenerate whole plants.



INTERPRETING THE RESULTS: The plant has been genetically engineered to contain a new gene. The transgenic plant will express a new trait based on that gene, perhaps resistance to an herbicide or the production of an insect toxin according to the goal of the experiment.

Kevin V. Wood

Figure 18.16
A genetically engineered tobacco plant, made capable of luminescence by the introduction of a firefly gene coding for the enzyme luciferase.

of the GMOs to the environment, such as by interbreeding with natural species or by harming beneficial insect species. For example, could introduced genes providing herbicide or insect resistance move from crop plants into related weed species through cross pollination, producing "super weeds" that might be difficult or impossible to control? And *Bt*-expressing corn was originally thought to have adverse effects on monarch butterflies who fed on the pollen. The most recent of a series of independent studies investigating this possibility has indicated that the risk to the butterflies is extremely low.

More broadly, different countries have reacted to GMOs in different ways. In the United States, transgenic crops are widely planted and harvested. Before commercialization, such GMOs are evaluated for potential risk by appropriate government regulatory agencies, including the NIH, Food and Drug Administration (FDA), Department of Agriculture, and Environmental Protection Agency (EPA). Typically, opposition to GMOs has come from particular activist and consumer groups.

Political opposition to GMOs has been greater in Europe, dampening the use of transgenic crop plants in the fields and GMOs in food. In 1999, the European Union (EU) imposed a 6-year moratorium on all GMOs, leading to a bitter dispute with the United States, Canada, and Argentina, the leading growers of transgenic crops. More recently, the EU has revised the GMO regulations in all member states. Basically, the EU has decided that using genetic engineering in agriculture and food production is permissible provided the GMO or food containing it is safe for humans, animals, and the environment. All use of GMOs in the field or in food requires authorization following a careful review process.

On a global level, an international agreement, the **Cartagena Protocol on Biosafety**, "promotes biosafety by establishing practical rules and procedures for the safe transfer [between countries], handling and use of GMOs." Separate procedures have been set up for GMOs that are to be introduced into the environment and those that are to be used as food or feed or for processing. To date, 132 countries have ratified the Protocol. (As of November 2006, the United States is not one of them.)

In sum, the use of DNA technologies in biotechnology has the potential for tremendous benefits to humankind. Such experimentation is not without risk, and so for each experiment, researchers must assess that risk and make a judgment about whether to proceed and, if so, how to do so safely. Furthermore, agreed-upon guidelines and protocols should ensure a level of biosafety for researchers, consumers, politicians, and governments.

We now turn to the analysis of whole genomes.

Dr. Jorge Mayer, Golden Rice Project

Figure 18.17
Rice genetically engineered to contain β-carotene.

Study Break

1. What are the principles of DNA fingerprinting?
2. What is a transgenic organism?
3. What is the difference between using germ-line cells and somatic cells for gene therapy?

18.3 Genome Analysis

The development of DNA technologies for analyzing genes and gene expression revolutionized experimental biology. DNA sequencing techniques (described in this section) have made it possible to analyze the sequences of cloned genes and genes amplified by PCR. Having the complete sequence of a gene aids researchers tremendously in unraveling how the gene functions. But a gene is only part of a genome. Researchers want to know about the organization of genes in a complete genome, and how genes work together in networks to control life. Of particular interest, of course, is the human genome. The complete sequencing of the approximately 3 billion base-pair human genome—the

Insights from the Molecular Revolution

Engineering Rice for Blight Resistance

Rice is common in the diet of the entire human population; for one-third of humanity, more than 2 billion people, it is the primary nutrient source. Worldwide, some 146 million hectares are planted with rice, producing 560 million tons of the grain annually.

This major human staple is threatened by a rice blight caused by the bacterium *Xanthomonas oryzae*. Rice plants infected by the bacterium turn yellow and wilt; in many Asian and African rice fields, as much as half of the crop is lost to the blight. Although some wild forms of rice, which are not usable as crop plants, have a natural resistance to the *Xanthomonas* blight, none of the cultivated varieties is resistant. Attempts to develop resistant strains by crossbreeding crop rice with wild varieties have produced only one resistant type that, unfortunately, is essentially unusable as a food source.

At best, crossbreeding plants to develop resistance can take many years to produce results. As a shortcut, Pamela Ronald and her colleagues at the University of California at Davis decided to try genetic engineering. They were aided by the results of genetic experiments in which a gene, *Xa21*, was found that confers resistance to *Xanthomonas* in a wild rice.

Ronald and her coworkers set out to clone the DNA of the *Xa21* gene. Their first step was to locate it in wild rice chromosomes, using traditional genetic crosses to set up a linkage map. The studies showed that *Xa21* is located near several known genes, including one that was so close that it rarely recombined with *Xa21* in genetic crosses.

The DNA of the known gene was used as a probe to find nearby sequences in a genomic library of the resistant wild rice genome. The probe paired with 16 DNA fragments in the library. To determine whether any of the fragments included the *Xa21* gene, the researchers introduced each fragment individually into rice crop plants that were susceptible to *Xanthomonas* infection, using a gene gun. Out of 1500 plants that incorporated the fragments, 50 plants, all containing the same 9600-base-pair fragment, proved to be resistant to the blight. These plants were the first rice crop plants to be engineered successfully to resist *Xanthomonas* infections.

The 9600-base-pair fragment was then broken into subfragments that were cloned and tested individually for their ability to confer resistance. This technique allowed the researchers to identify and isolate the specific part of the fragment containing the resistance gene. Sequencing the gene revealed that it encodes a typical plasma membrane receptor protein, which in some way triggers an internal cellular response on exposure to *Xanthomonas* or its molecular components. The response alters cell structure or biochemistry to inhibit growth of the bacterium.

Ronald and her coworkers have introduced the *Xa21* gene successfully into three varieties of rice that are widely grown as crops in Asia and Africa. These genetically engineered plants will be field-tested to see if they have the yield, taste, and resistance needed to make them successful crop plants. Ronald has also sent copies of the *Xa21* gene to investigators in Europe, Africa, Asia, and other locations in the United States, so they can experiment with introducing blight resistance into local varieties. If these efforts are successful, the genetically engineered rice plants promise to greatly increase the world output of this economically vital crop.

Human Genome Project (HGP)—began in 1990. The task was completed in 2003 by an international consortium of researchers and by a private company, Celera Genomics. As part of the official HGP, for purposes of comparison the genomes of several important model organisms commonly used in genetic studies were sequenced: *E. coli* (representing prokaryotes), the yeast *Saccharomyces cerevisiae* (representing single-celled eukaryotes), *Drosophila melanogaster* and *Caenorhabditis elegans* (the fruit fly and nematode worm, respectively, representing multicellular animals of moderate genome complexity), and *Mus musculus* (the mouse, representing a mammal of genome complexity comparable to that of humans). In addition, the sequences of the genomes of many organisms beyond this list, including plants, have been completed or are in progress at this time. What researchers are learning from analyzing complete genomes is of enormous importance to our understanding of biology and the evolution of organisms.

DNA Sequencing Techniques Are Based on DNA Replication

DNA sequencing is the key technology for genome sequencing projects. DNA sequencing is also used on a smaller scale, for example, in determining the sequence of individual genes that have been cloned or amplified by PCR.

DNA sequencing was first developed in the late 1970s by Allan M. Maxam, a graduate student, and his mentor, Walter Gilbert of Harvard University; within a few years, another investigator, Frederick Sanger of Cambridge University, designed the method that is most used today. Gilbert and Sanger were awarded a Nobel Prize in 1980.

The Sanger method is based on the properties of nucleotides known as *dideoxyribonucleotides*—the method, therefore, is also called *dideoxy sequencing* **(Figure 18.18).** Dideoxyribonucleotides have a single —H bound to the 3′ carbon of the deoxyribose sugar

Figure 18.18 Research Method

Dideoxy (Sanger) Method for Sequencing DNA

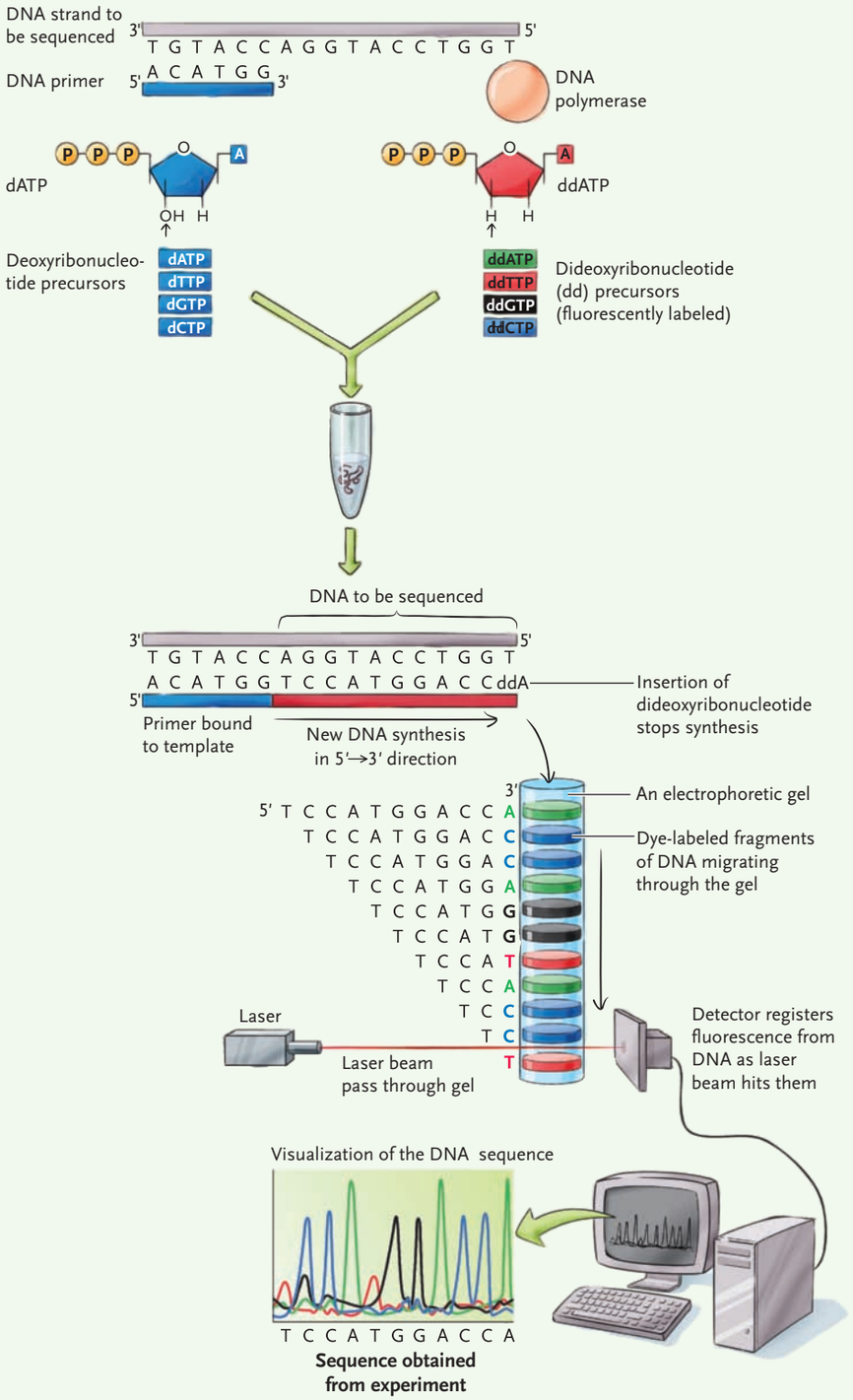
PURPOSE: Obtain the sequence of a piece of DNA, such as in gene sequencing or genome sequencing. The method is shown here with an automated sequencing system.

PROTOCOL:

1. A dideoxy sequencing reaction has the following components: the fragment of DNA to be sequenced (denatured to single strands); a DNA primer that will bind to the 3′ end of the sequence to be determined; a mixture of the four deoxyribonucleotide precursors for DNA synthesis; and a mixture of the four dideoxyribonucleotides (dd) precursors, each labeled with a different fluorescent molecule, and DNA polymerase to catalyze the DNA synthesis reaction.

2. Synthesis of the new DNA strand is in the 5′→3′ direction starting at the 3′ end of the primer. New synthesis continues until a dideoxyribonucleotide is incorporated into the DNA instead of a normal deoxyribonucleotide. For a large population of template DNA strands, the dideoxy sequencing reaction produces a series of new strands, with lengths from one on up. At the 3′ end of each new strand is the labeled dideoxyribonucleotide that terminated the synthesis.

3. The labeled strands produced by the reaction are separated by gel electrophoresis. The principle of separation is the same as for agarose gel electrophoresis described in Figure 18.9. But here it is necessary to discriminate between DNA strands that differ in length by one nucleotide, which agarose gels cannot do. In this case, therefore, a gel made of polyacrylamide is prepared in a capillary tube for separating the DNA fragments. As the bands of DNA fragments move near the bottom of the tube, a laser beam shining through the gel excites the fluorescent labels on each DNA fragment. The fluorescence is registered by a detector with the wavelength of the fluorescence indicating which of the four dideoxyribonucleotides is at the end of the fragment in each case.



INTERPRETING THE RESULTS: The data from the laser system are sent to a computer that interprets which of the four possible fluorescent labels is at the end of each DNA strand. The results show the colors of the labels as the DNA bands passed the detector. They may be seen on the computer screen or in printouts. The sequence of the newly synthesized DNA, which is complementary to the template strand, is read from left (5′) to right (3′). (The sequence shown here begins after the primer.)

instead of the —OH normally appearing at this position in deoxyribonucleotides. DNA polymerases, the replication enzymes, recognize the dideoxyribonucleotides and place them in the DNA just as they do the normal deoxyribonucleotides. However, because a dideoxyribonucleotide has no 3′-OH group available for addition of the next base, replication of a nucleotide chain stops when one of these nucleotides is added to a growing nucleotide chain. (Remember from Section 14.3 that a 3′-OH group must be present at the growing end of a nucleotide chain for the next nucleotide to be added during DNA replication.) In a dideoxy sequencing reaction, researchers use a mixture of dideoxyribonucleotides and normal nucleotides, so that chain termination will occur randomly at each position where a particular nucleotide appears in the population of DNA molecules being replicated. Each chain-termination event generates a newly synthesized DNA strand that ends with the dideoxyribonucleotide; hence, for this particular strand, the base at the 3′ end is known and, because of base-pairing rules, the base on the template strand being sequenced is deduced. Once they know the base at the end of each terminated DNA strand, researchers can work out the complete sequence of the template DNA strand.

The dideoxy sequencing method can be used with any pure piece of DNA, such as a cloned DNA fragment or a fragment amplified by PCR. An unambiguous sequence of about 500 to 750 nucleotides can be obtained from each sequencing experiment.

Structural Genomics Determines the Complete DNA Sequence of Genomes

Genome analysis consists of two main areas: *structural genomics* and *functional genomics*. **Structural genomics** is the actual sequencing of genomes and the analysis of the nucleotide sequences to locate genes and other functionally important sequences within the genome. **Functional genomics** is the study of the functions of genes and of other parts of the genome. In the case of genes, this includes developing an understanding of the regulation of their expression, the proteins they encode, and the role played by the proteins in the organism's metabolic processes.

The most widely used method for sequencing a genome is the *whole-genome shotgun method* **(Figure 18.19).** In this method, the entire genome is broken into thousands to millions of random, overlapping fragments, and each fragment is cloned and sequenced. The genome sequence then is assembled by computer on the basis of the sequence overlaps between fragments.

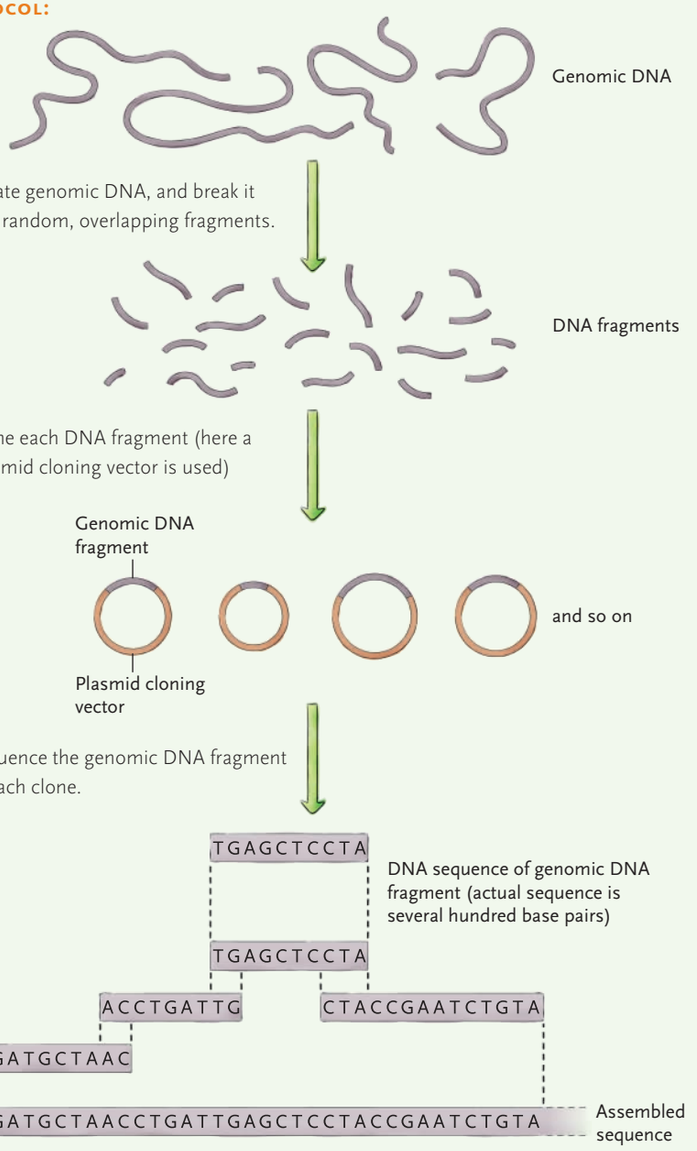
The first genome sequence reported, that of the bacterium *Haemophilus influenzae,* was determined using the whole-genome shotgun method by J. Craig Venter and his associates at Celera Genomics (the developers of the method). Originally it was thought that

Figure 18.19 Research Method

Whole-Genome Shotgun Sequencing

PURPOSE: Obtain the complete sequence of the genome of an organism.

PROTOCOL:



4. Enter the DNA sequences of the fragments into a computer, and use the computer to assemble overlapping sequences into the continuous sequences of each chromosome of the organism. This technique is analogous to taking 10 copies of a book that have been torn randomly into smaller sets of a few pages each and, by matching overlapping pages of the leaflets, assembling a complete copy of the book with the pages in the correct order.

INTERPRETING THE RESULTS: The method generates the complete sequence of the genome of an organism.

the much larger genomes of eukaryotes would be too difficult to sequence using this method. But improvements in sequencing technologies and in the computer algorithms used to identify overlapping sequences have made it easier to assemble the segment sequences into the sequence of a whole genome. Whole genome shotgun sequencing is now the method of choice for sequencing essentially any genome.

Functional Genomics Focuses on the Functions of Genes and Other Parts of the Genome

The genomes of a large number of viruses and more than 180 organisms have been sequenced, and those of more species are continually being added to the total. Among those already sequenced are the cytomegalovirus, bacteria including *E. coli,* various archaean species, and eukaryotes including the brewer's yeast *Saccharomyces cerevisiae,* the protozoan *Plasmodium falciparium* (the malarial parasite), the roundworm *Caenorhabditis elegans,* the plants *Arabidopsis thaliana* and rice, the fruit fly *Drosophila melanogaster,* the chicken, the mouse, the rat, the dog, the chimpanzee, and human.

Analysis of Genome Sequences. The complete genome sequence for an organism is basically a very long string of letters, which means little without further analysis. Discovering the functions of genes and other parts of the genome is one important goal of this analysis. Most research is focused on the genes because they control the functions of cells and, therefore, of organisms. Functional genomics relies on laboratory experiments by molecular biologists and sophisticated computer analyses by researchers in the rapidly growing field of **bioinformatics,** which fuses biology with mathematics and computer science. Bioinformatics is used, for example, to find genes within a genomic sequence, align sequences in databases to determine the degree of matching, predict the structure and function of gene products, and postulate evolutionary relationships for sequences.

Protein-coding genes are of particular interest in genome analysis. Once a genome sequence is determined, researchers use computer algorithms to search both strands of the sequence for these genes. They identify possible protein-coding genes by searching for open reading frames (ORFs), that is, a start codon (ATG, at the DNA level) in frame (separated by a multiple of three nucleotides) with one of the stop codons (TAG, TAA, or TGA at the DNA level). This process is easy for prokaryotic genomes, because the genes have no introns. In eukaryotic protein-coding genes, which typically have introns, more sophisticated algorithms are used to try to identify the junctions between exons and introns in scanning for open reading frames.

Each open reading frame found by computer analysis of a genome can be "translated" by computer to give the amino acid sequence of the protein it could encode. Researchers may then be able to assign a function to the open reading frame by performing a *sequence similarity search,* a computer-based comparison of a DNA or amino acid sequence with databases of sequences of known genes or proteins. That is, if an open reading frame or its protein product resemble those of a previously sequenced gene, the two genes are related in an evolutionary sense and are likely to have similar functions.

Many new features of genetic organization have been discovered, or previous conclusions reinforced, through the findings of genome sequencing. One of the more surprising discoveries is that the eukaryotic genomes sequenced to date contain large numbers of previously unknown genes, many more than scientists expected to find. In *Caenorhabditis,* for example, 12,000 of the 19,000 genes are of unknown function. Identifying these genes and their functions is one of the major challenges of contemporary molecular genetics.

Another revelation is the degree to which different organisms, some of them widely separated in evolutionary origins, contain similar genes. For example, even though the yeast *Saccharomyces* is a fungus separated from our species by millions of years of evolutionary history, about 2300 of its approximately 6000 genes are related to those of mammals, including many genes that control progress through the cell cycle. The similarities are so close that the yeast and human versions of many genes can be interchanged with little or no effect on cell functions in either organism.

The sequences also confirm that eukaryotic genomes contain large numbers of noncoding sequences, most of them in the form of repeated sequences of various lengths and numbers. Most of these sequences, which make up from about 25% to 50% of the total genomic DNA in different eukaryotic species, have no determined function at this point in time.

Features of the Human Genome Sequence. The human genome sequence consists of 3.2 *billion* base pairs. Until the genome was sequenced, researchers expected that human cells might contain as many as 100,000 different protein-coding genes. The best current estimate is 20,000 to 25,000 protein-coding genes. However, although the number of protein-coding genes is unexpectedly small, the total number of different proteins produced in humans is much greater and probably approaches the 100,000 figure originally proposed for genes. The additional proteins arise through such processes as alternative splicing during mRNA processing (see Section 15.3) and differences in protein processing (discussed further in the following).

All the protein-coding sequences occupy less than 2% of the human genome. Introns—the noncoding spacers in genes—occupy another 24% of the ge-

nome. The rest of the DNA, almost three-quarters of the genome, occupies the spaces between genes. Some of this intergenic DNA is functional and includes regulatory sequences such as promoters and enhancers, but much of it, more than 50% of the total genome, consists of repeated sequences that have no known function.

Completing the human genome sequence is only the beginning of human genomics. The next steps are to determine the functions of the unknown genes and of the sequence elements in intergenic regions. This *data mining*, as it is called, may answer fundamental questions about genome organization and the mechanisms controlling genes in development and cell differentiation. Genes related to human health and disease, including cancer, are of particular interest. The analysis of these disease-related genes may suggest methods to predict individual susceptibility to diseases and may possibly lead to means for their diagnosis and treatment.

On an even larger scale, the human genome is being compared with the genomes of other species to determine the molecular basis of differences in anatomy, physiology, and developmental patterns between species. Ultimately, species comparisons may reveal the mutational changes underlying the evolution of our species and many others. This area of genomics is known as *comparative genomics*.

There are bioethics issues concerning the human genome. To address those issues, the U.S. Department of Energy and the NIH have funded studies of the ethical, legal, and social issues surrounding the availability of genetic information from human genome research. Among the questions being looked at: Who should have access to personal genetic information, and how should it be used? To what extent should genetic information be private and confidential? How will genetic tests be evaluated and regulated? How can people be informed sufficiently about the genetic information from genomic analysis so that they can make informed personal medical choices? Does a set of genes predispose a person's behavior, and can the person control that behavior?

Studying Differential Gene Activity in Entire Genomes with DNA Microarrays. As a part of genome research, investigators are interested in comparing which genes are active in different cell types of humans and other organisms, and tracking the changes in total gene activity in the same cell types as development progresses. In some cases, the researcher wants to know whether or not particular genes are being expressed, and in other cases how the level of expression varies in different circumstances. This research has been revolutionized by a technique using **DNA microarrays.** The microarrays are also called **DNA chips** for short because the techniques used to "print" the arrays resemble those used to lay out electronic circuits on a computer chip. The surface of a DNA chip is divided into a microscopic grid of about 60,000 spaces. On each space of the grid, a computerized system deposits a microscopic spot containing about 10,000,000 copies of a DNA probe about 20 nucleotides long.

Studies of gene activity using DNA microarrays involves comparing gene expression under a defined experimental condition with expression under a reference (control) condition. For instance, DNA microarrays can be used to answer basic biological questions, such as: How does gene expression change when a cell goes from a resting state (reference condition) to a dividing state (experimental condition); that is, how is gene expression different in different stages of development? DNA microarrays can also be used to address many questions of medical significance, such as: How are genes differentially expressed in normal cells and cells of various cancers? In these experiments, investigators might focus on which genes are active and inactive under the two conditions, or on how the levels of expression of genes change under the two conditions.

Figure 18.20 shows how a DNA microarray is used to compare gene expression in normal cells and in cancer cells in humans. mRNAs are isolated from each cell type, and cDNAs are made from them, incorporating different fluorescent labels: green for one, red for the other. The two cDNAs are mixed and added to the DNA chip, where they hybridize with any complementary probes. A laser locates and quantifies the green and red fluorescence, enabling a researcher to see which genes are expressed in the cells and, for those that are expressed, to quantify differences in gene expression between the two cell types (see Interpreting the Results in Figure 18.20). The results can help researchers understand how the cancer develops and progresses.

DNA microarrays are also used to screen individuals for particular mutations. To detect mutations, the probes spotted onto the chip include probes for the normal sequence of the genes of interest along with probes for sequences of all known mutations. A fluorescent spot at a site on the chip printed with a probe for a given mutation immediately shows the presence of the mutation in the individual. Such a test is currently used to screen patients for whether they carry any one of a number of mutations of the *breast cancer 1 (BRCA1)* gene known to be associated with the possible development of breast cancer.

Studying the Array of Expressed Proteins Is the Next Level of Genomic Analysis

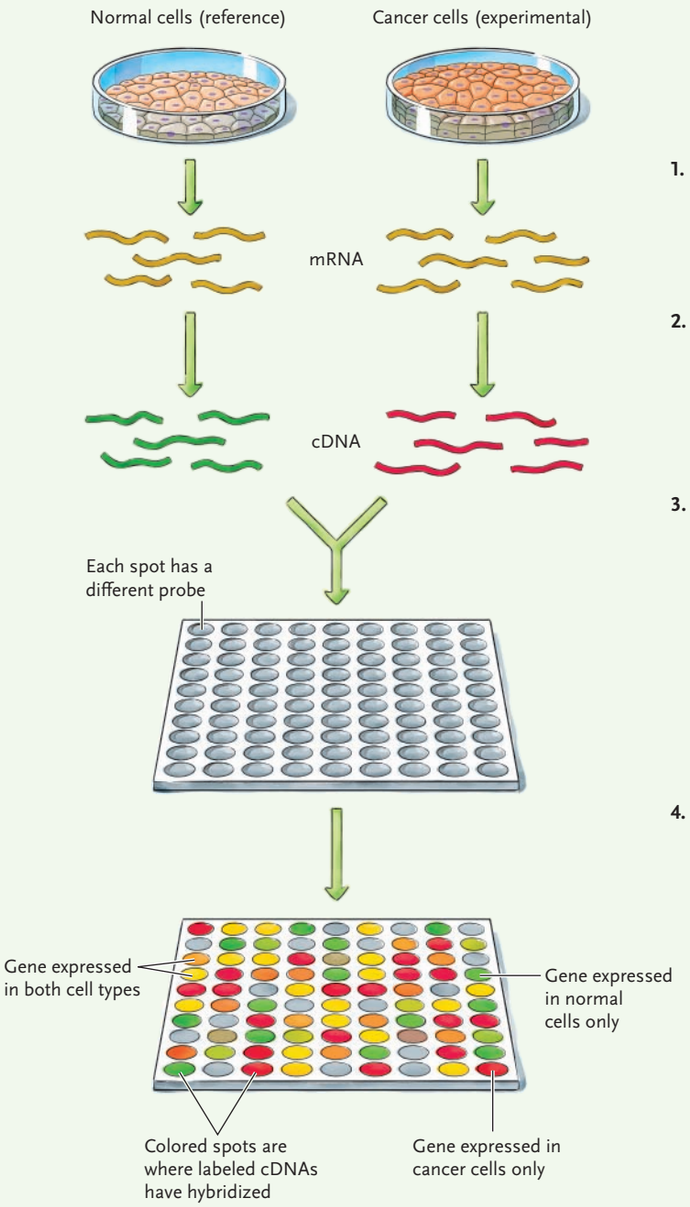
Genome research also includes analysis of the proteins encoded by a genome, for proteins are largely responsible for cell function and, therefore, for all the functions of an organism. The term **proteome** has been coined to refer to the complete set of proteins that can be expressed by an organism's genome. A *cellular proteome* is a subset of those proteins, the collection of

Figure 18.20 Research Method

DNA Microarray Analysis of Gene Expression Levels

PURPOSE: DNA microarrays can be used in various experiments, including comparing the levels of gene expression in two different tissues, as illustrated here. The power of the technique is that the entire set of genes in a genome can be analyzed simultaneously.

PROTOCOL:



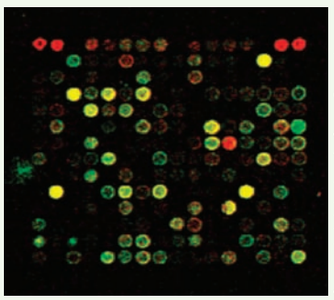
1. Isolate mRNAs from a control cell type (here, normal cells) and an experimental cell type (here, cancer cells).

2. Prepare cDNA libraries from each mRNA sample. For the normal cell (control) library use nucleotides with a green fluorescent label, and for the cancer cell (experimental) library use nucleotides with a red fluorescent label.

3. Denature the cDNAs to single strands, mix them, and pump them across the surface of a DNA microarray containing a set of single-stranded probes representing every protein-coding gene in the human genome. The probes are spotted on the surface, with each spot containing a probe for a different gene. Allow the labeled cDNAs to hybridize with the gene probes on the surface of the chip, and then wash excess cDNAs off.

4. Locate and quantify the fluorescence of the labels on the hybridized cDNAs with a laser detection system.

Actual DNA microarray result



Courtesy Ludvig Institute for Cancer Research

INTERPRETING THE RESULTS: The colored spots on the microarray indicate where the labeled cDNAs have bound to the gene probes attached to the chip and, therefore, which genes were active in normal and/or cancer cells. Moreover, we can quantify the gene expression in the two cell types by the color detected. A purely green spot indicates the gene was active in the normal cell, but not in the cancer cell. A purely red spot indicates the gene was active in the cancer cell, but not in the normal cell. A yellow spot indicates the gene was equally active in the two cell types, and other colors tell us the relative levels of gene expression in the two cell types. For this particular experiment, we would be able to see how many genes have altered expression in the cancer cells, and exactly how their expression was changed.

proteins found in a particular cell type under a particular set of environmental conditions.

The study of the proteome is the field of **proteomics.** The number of possible proteins encoded by the genome is larger than the number of protein-coding genes in the genome, at least in eukaryotes. In eukaryotes, alternative splicing of gene transcripts and variation in protein processing means that expression of a gene may yield more than one protein product. Therefore, proteomics is a more challenging area of research than is genomics.

The two major immediate goals of proteomics are to determine (1) the number and structure of proteins in the proteome and (2) the functional interactions between the proteins. The interactions are particularly important because they help us understand how proteins work together to determine the phenotype of the cell. For instance, if a particular set of interacting proteins characterized a lung tumor cell, then drugs could be developed that specifically target the interactions.

What are the tools of proteomics? For many years it has been possible to separate and identify proteins by gel electrophoresis (using polyacrylamide to make the gels, the same material used for separating DNA fragments in DNA sequencing) or mass spectrometry. However, to study an entire cellular proteome, many more proteins must be analyzed simultaneously than is possible with either of those techniques. A big step in that direction is the development of **protein microarrays (protein chips),** which are similar in concept to DNA microarrays. For example, one type of protein microarray involves binding antibodies prepared against different proteins to different locations on the protein chip. An antibody for a foreign substance such

Unanswered Questions

Can pharmaceuticals be customized to the individual?

One exciting avenue of investigation arising from genomics research is pharmacogenomics. The word is a combination of "pharmacology" and "genomics" and means the study of how the genome of an individual affects the body's responses.

Currently, pharmaceuticals are dispensed under the assumption that all humans are basically the same. Of course, the dose may be adjusted to the weight or age of the individual. However, not everyone is the same, as we know from the fact that every drug provokes an allergic reaction in some people. Indeed, a variety of factors affect a person's response to medicines, including the genome (the expression of the particular array of genes in tissues targeted by the pharmaceutical), diet, health, age, and the environment.

The goal is for the pharmaceutical to be as effective as it can be, with the best available drug therapy given to a patient right from the beginning. Pharmacogenomics offers the promise of customizing drugs by adapting to them to key cellular proteomes of the patient. Conceptually drugs could be chemically matched to the proteins, enzymes, and RNA molecules that are associated with diseases. Such drugs therefore would be much more targeted to specific diseases than present-day drugs are, the result being that the therapeutic benefits will be maximized while side effects will be minimized. And drug dosages will be tailored to an individual's genetic makeup (as manifested in the proteome) so as to take into account how and at what rate a person metabolizes a drug.

There are very few examples of pharmacogenomics in action in the clinic at the moment. Many research groups are currently working at the lab bench to develop applications.

Can gene expression at the genomic level be used to identify specific cancers?

DNA microarrays can be used to analyze the differences in gene expression between two cell or tissue types (see Figure 18.20). In 1999, Eric Lander's research group at Whitehead Institute, MIT, was the first to show that gene expression arrays can distinguish between two types of cancer. The cancers, acute myeloid leukemia (AML) and acute lymphoblastic leukemia (ALL), are very similar in clinical symptoms, but because the treatments are different the diagnosis must be precise. Before Lander's discovery, the cancers could be distinguished only through a series of expensive tests that took precious time. Lander used DNA microarray analysis based on 6800 human genes and accurately distinguished the two types.

DNA microarray analysis has now been developed to identify a number of different types of cancers based on gene expression profiles. Undoubtedly many more cancers will be added to the list as time goes on.

Can genome-wide approaches produce tools to combat pathogens?

Animals are susceptible to attack by a wide range of pathogenic agents, including virulent bacteria and viruses. Humans are vaccinated against some of the important viral pathogens, an action that has saved millions of lives. Still, we are invaded. In healthy individuals, the immune system is activated to combat the invasion. Immunocompromised individuals, however, have weakened immune systems; they risk serious illness or death from infection with pathogens that healthy individuals combat easily. Pathogens are also highly relevant to our lives in another way: agricultural animals and plants are vital for maintaining our food supply, yet they also may be or may become susceptible to pathogen invasion.

Biodefense proteomics is an area of research with the goal of using genome-wide approaches to produce tools to combat pathogens. It is anticipated that the research will discover targets for the next generation of vaccines, therapeutics, and diagnostics. The proteomics approaches being used include characterizing the proteomes of pathogens (for example, for the pathogen itself if it is cellular, or for an infected cell if the pathogen is viral); to determine the mechanisms of microbial pathogenesis; and to develop an understanding of animal immune responses and nonimmune mediate responses that are triggered by microbial pathogens.

Peter J. Russell

as a protein is generated by the immune system of an animal that has been injected with that substance. The antibody is isolated from the blood of that animal and can be used to bind specifically to the protein in experiments. Proteins are isolated from cells and labeled, and then pumped over the surface of the protein microarray. Each labeled protein binds to the antibody for that protein. After washing off excess proteins, the protein microarray is analyzed much as for DNA microarrays to determine where the proteins bound and to quantify that binding. With this technique, a researcher can quantify proteins in different cell types and different tissues. Researchers can also compare proteins under different conditions, such as during differentiation, or with and without a particular disease condition, or with and without a particular drug treatment. In the future, we can expect protein arrays to become routine for studying cellular proteomes.

Systems Biology Is the Study of the Interactions between All the Components of an Organism

Traditional biology research focuses on identifying and studying the functions of individual genes, proteins, and cells. Although such research has provided an enormous body of knowledge—this textbook being an example—it provides only a limited insight into how a whole organism functions at the cellular and molecular levels. For instance, studying separately the individual components of a bicycle does not tell you what the whole bicycle is or what it does.

Systems biology is an area of biology that seeks to overcome the limitations of traditional biology approaches by studying the organism as a whole to unravel the integrated and interacting network of genes, proteins, and biochemical reactions responsible for life. That is, systems biologists work from the premise that those interactions are responsible for an organism's form and function. Present-day research in systems biology has been stimulated by the development of techniques for genomic and proteomic analysis and by the data from those analyses.

Systems biologists use genomics and proteomics techniques, such as those discussed in this section along with information from other sources. They typically obtain very complex data and use sophisticated quantitative analysis to generate models for the interactions within an organism.

Systems biologists study organisms of many kinds. Some focus on humans and have the ambitious goal of transforming the practice of medicine. The vision is to define the interactions between all the components that affect the health of an individual human. It may then be possible to predict more accurately than is currently possible whether a person will develop particular diseases and to personalize treatments for those diseases.

Unquestionably DNA technologies and genomics research have resulted in remarkable achievements so far, probably touching every person on the planet directly or indirectly. They, and proteomics research, also have enormous potential to lead to major advances and achievements in the future.

Study Break

1. How are possible protein-coding genes identified in a genome sequence of a bacterium? Of a mammal?
2. How would you determine how a steroid hormone affects gene expression in human tissue culture cells?

Review

Go to ThomsonNOW™ at www.thomsonedu.com/login to access quizzing, animations, exercises, articles, and personalized homework help.

18.1 DNA Cloning

- Producing multiple copies of genes by cloning is a common first step for studying the structure and function of genes or for manipulating genes. Cloning involves cutting genomic DNA and a cloning vector with the same restriction enzyme, joining the fragments to produce recombinant plasmids, and introducing those plasmids into a living cell such as a bacterium, where replication of the plasmid takes place (Figures 18.2–18.4).
- A clone containing a gene of interest may be identified among a population of clones by using DNA hybridization with a labeled nucleic acid probe (Figure 18.5).
- A genomic library is a collection of clones that contains a copy of every DNA sequence in the genome. A cDNA (complementary DNA) library is the entire collection of cloned cDNAs made from the mRNAs isolated from a cell. A cDNA library contains only sequences from the genes that are active in the cell when the mRNAs are isolated.
- PCR amplifies a specific target sequence in DNA, such as a gene, defined by a pair of primers. PCR increases DNA quantities by successive cycles of denaturing the template DNA, annealing the primers, and extending the primers in a DNA synthesis reaction catalyzed by DNA polymerase; with each cycle, the amount of DNA doubles (Figure 18.6).

Animation: Base pairing of DNA fragments

Animation: Formation of recombinant DNA

Animation: Restriction enzymes

Animation: How to make cDNA

Animation: Use of a radioactive probe

Animation: Polymerase chain reaction (PCR)

Animation: Automated DNA sequencing

18.2 Applications of DNA Technologies

- Recombinant DNA and PCR techniques are used in DNA molecular testing for human genetic disease mutations. One approach exploits restriction site differences between normal and mutant alleles of a gene that create restriction fragment length polymorphisms (RFLPs) detectable by DNA hybridization with a labeled nucleic acid probe (Figures 18.8 and 18.9).
- Human DNA fingerprints are produced from a number of loci in the genome characterized by tandemly repeated sequences that vary in number in all individuals (except identical twins). To produce a fingerprint, the PCR is used to amplify the region of genomic DNA for each locus, and the lengths of the PCR products indicate the alleles an individual has for the repeated sequences at each locus. DNA fingerprints are widely used to establish paternity, ancestry, or criminal guilt (Figure 18.10).
- Genetic engineering is the introduction of new genes or genetic information to alter the genetic makeup of humans, other animals, plants, and microorganisms such as bacteria and yeast. Genetic engineering primarily aims to correct hereditary defects, improve domestic animals and crop plants, and provide proteins for medicine, research, and other applications. (Figures 18.11, 18.12, and 18.15).
- Genetic engineering has enormous potential for research and applications in medicine, agriculture, and industry. Potential risks include unintended damage to living organisms or the environment.

Animation: DNA fingerprinting

Animation: Gene transfer using a Ti plasmid

Animation: Transferring genes into plants

18.3 Genome Analysis

- Genome analysis consists of two main areas: structural genomics, the sequencing of genomes and the identification of the genes the sequences contain, and functional genomics, the study of the function of genes and other parts of the genome.
- Sequencing a genome involves a replication reaction with a DNA template, a DNA primer, the four normal deoxyribonucleotides, and a mixture of four dideoxyribonucleotides, each labeled with a different fluorescent tag, and DNA polymerase. Replication stops at any place in the sequence in which a dideoxyribonucleotide is substituted for the normal deoxyribonucleotide. The lengths of the terminated DNA chains and the label on them indicate the overall sequence of the DNA chain being sequenced (Figure 18.18).
- The whole-genome shotgun method of sequencing a genome involves breaking up the entire genome into random, overlapping fragments, cloning each fragment, determining the sequence of the fragment in each clone, and using computer algorithms to assemble overlapping sequences into the sequence of the complete genome (Figure 18.19).
- Once a gene is sequenced, the sequence of the protein encoded in a prokaryotic gene can be deduced by reading the coding portion of the gene three nucleotides at a time, starting at the AUG codon that indicates the beginning of a coding sequence.
- Complete genome sequences have been obtained for many viruses, a large number of prokaryotes, and many eukaryotes, including the human. The sequences have revealed that all eukaryotes share related gene sequences, and they have also revealed a significant proportion of genes whose functions are not presently known.
- Having the complete genome of an organism makes it possible to study the expression of all of the genes in the genome simultaneously, including comparing gene expression in two different cell types. The DNA microarray (or DNA chip) is typically used for the comparison; this technique can provide information about which genes are active in the two cell types as well as relative levels of expression of those genes (Figure 18.20).
- Proteomics is the study of the complete set of proteins in an organism or in a particular cell type. Protein numbers, protein structure, and protein interactions are all topics of proteomics.
- Systems biology combines data derived from genomics, proteomics, and other sources of information. Using sophisticated quantitative analysis, it seeks to model the total array of interactions responsible for an organism's form and function.

Questions

Self-Test Questions

1. Using cDNA is associated with which of the following?
   a. Introns can be identified and sequenced by this method.
   b. It measures both active and inactive DNA.
   c. Promoter regions can be identified by this method.
   d. One can identify start and stop regions by this method.
   e. One can identify active mRNA and make a complementary DNA sequence to the mRNA.
2. Restriction endonucleases, ligases, plasmids, viral or yeast vectors, electrophoretic gels, and a bacterial gene resistant to an antibiotic are all required for:
   a. dideoxyribonucleotide analysis.
   b. PCR.
   c. DNA cloning.
   d. DNA fingerprinting.
   e. DNA sequencing.
3. The PCR technique is distinguished from other processes discussed in this chapter by the use of:
   a. primers.
   b. DNA.
   c. RNA.
   d. Taq polymerase.
   e. the four nucleoside triphosphates.
4. Restriction fragment length polymorphisms:
   a. are produced by reaction with restriction endonucleases and are detected by Southern blot analysis.
   b. are of the same length for mutant and normal β-globin alleles.
   c. determine the sequence of bases in a DNA fragment.
   d. have in their middle short fragments of DNA that are palindromic.
   e. are used as vectors.
5. DNA fingerprinting:
   a. compares one stretch of the same DNA between two or more people.
   b. measures different lengths of DNA from many repeating noncoding regions for comparison between two or more people.
   c. requires the largest DNA lengths to run the greatest distance on a gel.
   d. requires amplification after the gels are run.
   e. can easily differentiate DNA between identical twins.

6. Dolly, a sheep, was an example of reproductive (germ line) cloning. Required to perform this process was:
   a. implantation of uterine cells from one strain into the mammary gland of another.
   b. the fusion of the mammary cell from one strain with an enucleated egg of another strain.
   c. the fusion of an egg from one strain with the egg of a different strain.
   d. the fusion of an embryonic diploid cell with an adult haploid cell.
   e. the fusion of two nucleated mammary cells from two different strains.
7. All of the following are true for somatic cell gene therapy *except:*
   a. White blood cells can be used.
   b. Somatic cells are cultured, and the desired DNA is introduced into them.
   c. Cells with the introduced DNA are returned to the body.
   d. The technique is still very experimental.
   e. The inserted genes are passed on to the offspring.
8. The sequence of the human genome:
   a. was obtained by sequencing overlapping DNA fragments.
   b. revealed far more genes than expected.
   c. revealed 3 trillion base pairs.
   d. used techniques not applicable to mapping other species.
   e. revealed 250,000 protein-coding genes.
9. Sanger's DNA sequencing technique:
   a. uses dideoxyribonucleotides to make new full-length strands of DNA.
   b. is based on cellular transcription.
   c. requires an RNA template, RNA primer, RNA polymerase, reverse transcriptase, and the dideoxyribonucleotides, ddATP, ddUTP, ddCTP, and ddGTP.
   d. places the RNA template to be sequenced on a gel and then adds the other ingredients from (c).
   e. is based on DNA replication.
10. A microarray could be used to:
   a. sequence DNA from several chromosomes in one individual.
   b. synthesize multiple copies of DNA from several sources.
   c. propagate human germ-line cells for cloning.
   d. compare coding DNA from a patient's normal lung cells with coding DNA from his cancerous lung cells.
   e. determine proteins that are expressed under certain environmental conditions.

Questions for Discussion

1. Do you think that genetic engineering is worth the risk? Who do you think should decide whether genetic engineering experiments and projects should be carried out: scientists, judges, politicians?
2. Do you think that human germ-line cells should be modified by genetic engineering to cure birth defects? To increase intelligence or beauty?
3. Write a paragraph supporting genetic engineering, and one arguing against it. Which argument carries more weight, in your opinion?
4. What should juries know to interpret DNA evidence? Why might juries sometimes ignore DNA evidence?
5. A forensic scientist obtained a small DNA sample from a crime scene. In order to examine the sample, he increased its quantity by the polymerase chain reaction. He estimated that there were 50,000 copies of the DNA in his original sample. Derive a simple formula and calculate the number of copies he will have after 15 cycles of the PCR.
6. A market puts out a bin of tomatoes that have outstanding color, flavor, and texture. A sign posted above them identifies them as genetically engineered produce. Most shoppers pick unmodified tomatoes in an adjacent bin, even though they are pale, mealy, and nearly tasteless. Which tomatoes would you pick? Why?

Experimental Analysis

Suppose a biotechnology company has developed a GMO, a transgenic plant that expresses *Bt* toxin. The company sells its seeds to a farmer under the condition that the farmer may plant the seed, but not collect seed from the plants that grow and use it to produce crops in the subsequent season. The seeds are expensive, and the farmer buys seeds from the company only once. How could the company show experimentally that the farmer has violated the agreement and is using seeds collected from the first crop to grow the next crop?

Evolution Link

Search for the words "comparative genomics" on the Internet to answer this question: How can complete genome sequences provide more accurate information about the evolution of species than sequences of one or a few genes?

How Would You Vote?

Nutritional labeling is required on all packaged food in the United States, but genetically modified food products may be sold without labeling. Should food distributors be required to label all products made from genetically modified plants or livestock? Go to www.thomsonedu.com/login to investigate both sides of the issue and then vote.