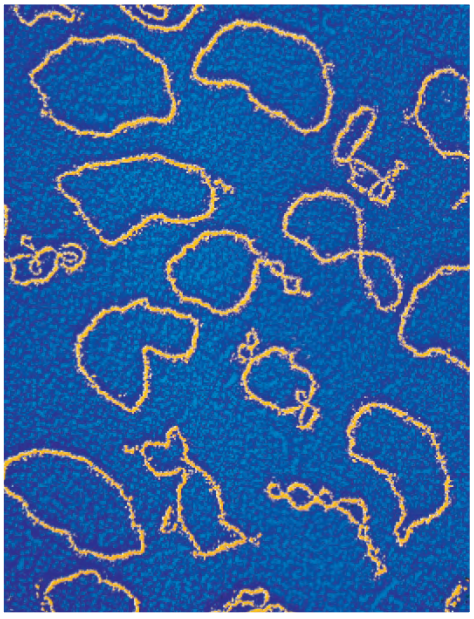
13

GENOMICS, BIOTECHNOLOGY, AND RECOMBINANT DNA



Artificially colored transmission electron micrograph of DNA plasmids from the bacterium *Escherichia coli*. These plasmids are used in genetic engineering.

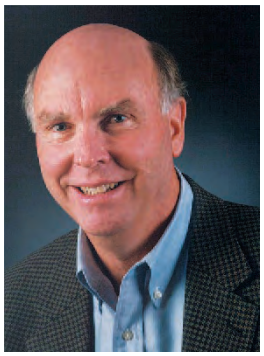
(© Dr. Gopal Murti/SPL/Photo Rsearchers, Inc.)

STUDY OBJECTIVES

1. To look at the techniques of gene cloning 358
2. To examine the techniques of creating restriction maps 377
3. To study the methods of DNA sequencing 383
4. To look at the goals and methods of the Human Genome Project 390
5. To look at the practical benefits and human issues of genetic engineering 397

STUDY OUTLINE

In the spring of 2000, J. Craig Venter, CEO of Celera Genomics, and Francis Collins, director of the National Institutes of Health's Human Genome Research Institute, jointly announced that they and their colleagues had completed the sequence of the human genome. Although there is still work ahead to finish the project, the accomplishment was enormous. To some, it was working out the very secret of life. This accomplishment firmly established the science of **genomics,** the study of the mapping and sequencing of genomes.



J. Craig Venter (1946–).
(Courtesy of Celera Genomics.)

Francis S. Collins (1950–).
(Courtesy of Francis Collins.)

Since the mid-1970s, the field of molecular genetics has undergone explosive growth, noticeable not only to geneticists, but also to medical practitioners and researchers, agronomists, animal scientists, venture capitalists, and the public in general. Medical practitioners and researchers have new treatments for diseases available. Agronomists see the possibility of greatly improved crop yields, and animal scientists have gained the possibility of greatly improving food production from domesticated animals. Geneticists and molecular biologists are gaining major new insights into understanding gene expression and its control.

The new DNA manipulation techniques, centered on the isolation, amplification, sequencing, and expression of genes, are based on the insertion of a particular piece of foreign DNA into a vector—a plasmid or phage. A plasmid is placed into a host cell, either prokaryotic or eukaryotic, which then divides repeatedly, producing numerous copies of the vector with its foreign piece of DNA. A phage simply multiplies in host cells (fig. 13.1). In both cases, the foreign piece of DNA is amplified in number; it can be expressed (transcribed and translated into a protein) when in a plasmid in a host cell. A commonly used host cell is *E. coli.* Following its amplification, the foreign DNA can be purified and its nucleotide sequence determined. When it is expressed, large quantities of the gene product of the foreign DNA can be obtained. The new technology is variously referred to as **gene cloning, recombinant DNA technology,** or **genetic engineering.** In this chapter, we look in detail at the methods and procedures of recombinant DNA technology, including DNA sequencing.

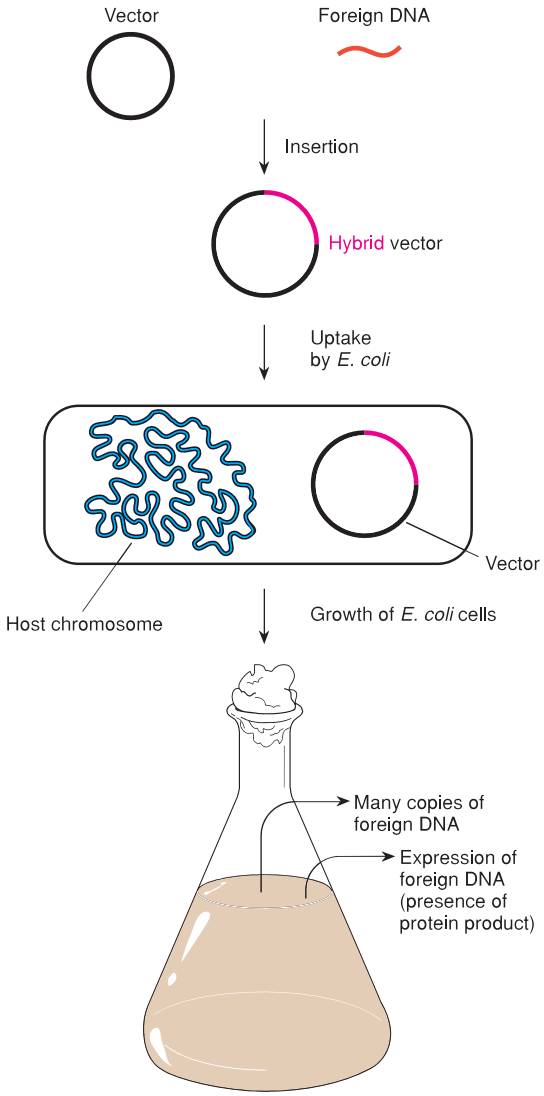


Figure 13.1 Overview of recombinant DNA techniques. A hybrid vector is created, containing an insert of foreign DNA. The vector is then inserted into a host organism. Replication of the host results in many copies of the foreign DNA and, if the gene is expressed, quantities of the gene product. (All DNA shown is double-stranded.)

GENOMIC TOOLS

Restriction Endonucleases

In 1978, Nobel prizes in physiology and medicine were awarded to W. Arber, H. Smith, and D. Nathans for their pioneering work in the study of **restriction endonucleases.** These are enzymes that bacteria use to destroy foreign DNA, presumably, the DNA of invading viruses. The enzymes recognize certain nucleotide sequences **(restriction sites)** found on foreign DNA, usually from four to eight base-pairs long, and then cleave that DNA at or near those sites. (Restriction endonucleases were originally so named because they restricted phage infection among strains of bacteria. Phages that could survive in one strain could not survive in other strains with different restriction enzymes.)

Three types of restriction endonucleases are known. Their groupings are based on the types of sequences they recognize, the nature of the cut made in the DNA, and the enzyme structure. Types I and III restriction endonucleases are not useful for gene cloning because they cleave DNA at sites other than the recognition sites and thus cause random, unpredictable cleavage patterns. Type II endonucleases, however, cleave at the specific sites they recognize, leading to predictable cleavage patterns. The sites type II endonucleases recognize are inverted repeats; they have twofold symmetry. To see the symmetry, you must read outward from a central axis on opposite strands of the DNA. For example, the type II restriction endonuclease *Bam*HI recognizes

5′-GGA | TCC-3′
3′-CCT | AGG-5′

Reading out from the center (vertical line) is AGG on the top strand and AGG on the bottom strand. The sequence is, in a sense, **a palindrome,** a sequence that reads the same from either direction. (Palindrome is from the Greek *palindromos,* which means "to run back." The name Hannah and the numerical sequence 1238321 are palindromes.) In figure 13.2 are some palindromic sequences that type II restriction endonucleases recognize; well over one hundred type II enzymes are known.

The host cell protects its DNA not by being free of these restriction sites, but usually by methylating its DNA in these regions (fig. 13.3). The same sequences that the

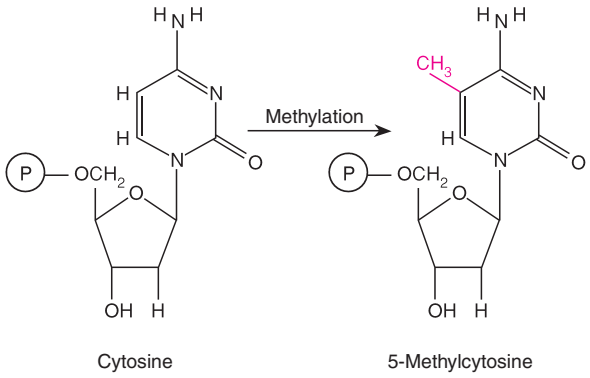


Figure 13.3 A methylase enzyme adds a methyl group to cytosine, converting it to 5-methylcytosine.

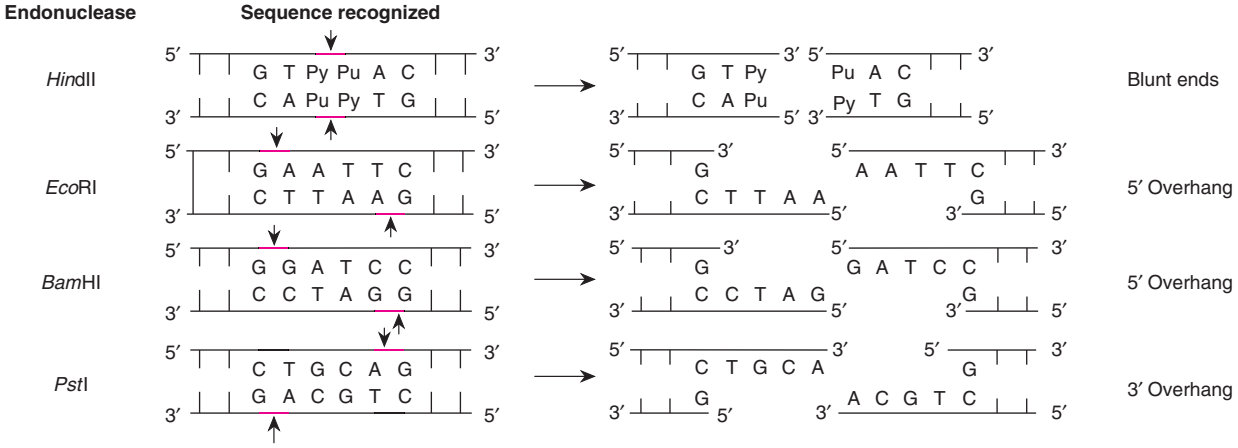


Figure 13.2 Sequences cleaved by various type II restriction endonucleases. Py is any pyrimidine and Pu is any purine. *Arrows* denote places where endonucleases cleave the DNA. In 1971, K. Danna and D. Nathans showed that a restriction endonuclease would consistently cut DNA into pieces of the same size. This precision and repeatability of enzyme action made enzymes useful for further research. Not all restriction endonucleases make staggered cuts with 3′ and 5′ overhangs; some produce blunt ends.

endonucleases attack in the unmethylated condition are protected when methylated. After host DNA replication, new double helices are hemimethylated; that is, the old strand is methylated but the new one is not. In this configuration, the new strand is quickly methylated (fig. 13.4). Foreign DNA, without methyl groups on either strand, is not methylated.

Restriction endonucleases are named after the bacteria from which they were isolated: *Bam*HI from *Bacillus amyloliquefaciens,* strain H; *Eco*RI from *E. coli,* strain RY13; *Hin*dII from *Haemophilus influenzae,* strain Rd; and *Bgl*I from *Bacillus globigii.* From here on, we will refer to type II restriction endonucleases simply as restriction enzymes.

Restriction enzymes cut the DNA in two different ways. For example, *Hin*dII cuts the recognition sequence down the middle, leaving "blunt" ends on the DNA (see fig. 13.2). We will discuss how pieces of DNA with blunt ends can be used in cloning. The staggered cuts made, for example, by *Bam*HI leave "sticky" ends (a 5′ overhang) that can reanneal spontaneously as hydrogen bonds form between the complementary bases (see fig. 13.2). The ability to reanneal these sticky ends, first demonstrated by S. Cohen, H. Boyer, and colleagues in 1971, opened up the field of gene cloning.

Prokaryotic Vectors

With current technology it is routine to join together, in vitro, DNAs from widely different sources. In figure 13.5, we see how a circular DNA molecule cleaved by a specific restriction enzyme can recircularize if it is cleaved in only one place, or how different molecules with the same free ends can anneal to form hybrid molecules. Only the action of a DNA ligase is needed to make the molecules complete (see chapter 9).

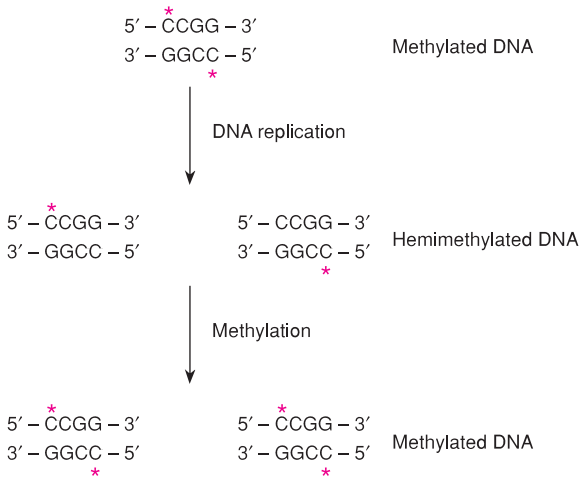


Figure 13.4 Host DNA is methylated in the *Hpa*II restriction site. *Asterisks* indicate methyl groups on cytosines. After DNA replication, the DNA is hemimethylated; the new strands have no methyl groups. Hemimethylated DNA is then fully methylated by cellular enzymes.

One of the pieces of DNA involved in the annealing can be a plasmid, a piece of DNA that can replicate in a cell independently of the cellular chromosome. The **recombinant plasmid** (fig. 13.6) can be transferred into a cell. (A recombinant plasmid is also known as a **hybrid plasmid, hybrid vehicle, hybrid vector,** or **chimeric plasmid.** The latter is after the *chimera,* a mythological monster with a lion's head, a goat's body, and a serpent's tail.) Many procedures exist that can introduce this recombinant plasmid into a host cell. For example, a bacterial cell can be made permeable to this, or any, plasmid by the addition of a dilute solution of calcium chloride. Once inside the cell, the foreign DNA is replicated each time the plasmid DNA replicates.

Note that in the process of inserting a piece of foreign DNA, the restriction site is duplicated, with one copy at either end of the insert. This property makes it easy to remove the cloned insert at some future time, if needed, since restriction sites enclose it (fig. 13.6).

Cloning with Restriction Enzymes

A few conditions must be met in order to succeed in cloning DNAs from different sources. A plasmid vehicle should be cleaved at only one point by the endonuclease. If it is cleaved at more than one point, it will fragment during the experiment. However, some phage vehicles must be cleaved at two points so that the foreign DNA can replace a length of the phage DNA rather than simply being inserted. Common vehicles, derivatives of phage λ, have been named **Charon phages** (pronounced "karon") after the mythical boatman of the River Styx. (See chapter 14 for a detailed discussion of phage λ.)

During normal phage infection (see chapter 7), only DNA the size of a phage genome is packaged into λ heads. Thus, for λ to be a useful vector, the foreign DNA must replace part of its DNA. We note that λ can function quite well as a hybrid vehicle with a 15,000 base-pair (15 kilobases, or 15 kb) section replaced by foreign DNA because that section of phage DNA is used for integration into the *E. coli* chromosome, a nonessential phage function. That is, the phage can infect a bacterium, replicate inside the bacterium, and burst out without the integration region. Genetic engineers have created a λ DNA molecule with the nonessential region missing and an *Eco*RI cleavage site in its place. Only hybrid DNA can thus be incorporated

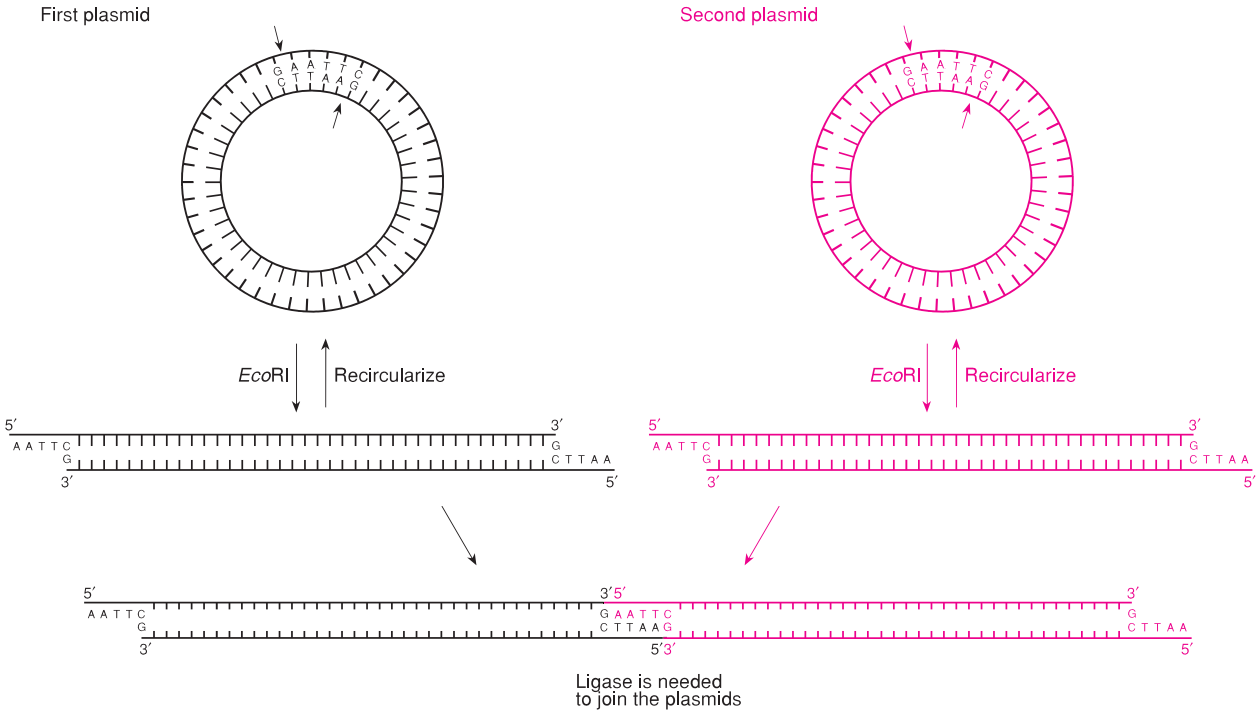


Figure 13.5 Circular plasmid DNA with a palindrome recognized by *Eco*RI. After the DNA is cleaved by the endonuclease, it has two exposed ends that can join to recircularize the molecule or unite two or more linear molecules of DNA cleaved by the same restriction endonuclease. The final nicks are closed with DNA ligase. S. Cohen, H. Boyer, and their colleagues first joined plasmids with this technique in 1971.

into phage heads because the diminished phage DNA, without an insert, is too small to be properly packaged.

One disadvantage of cloning with normal *E. coli* plasmids is that they are unstable if the foreign DNA is very large, greater than about 15 kb. That is, if a large chromosomal segment is cloned, the plasmid tends to lose parts of the clone as the plasmid replicates. Primarily for this reason, geneticists began using phage λ as a vector (see fig. 7.21) because these phages could successfully maintain foreign DNA as large as 24 kb.

The phage chromosome is about 50 kb of DNA; within the phage head it is linear, and within the cell it is circular. The DNA to fill the phage head is recognized during infection because it has a small segment of single-stranded DNA called a *cos* site (twelve bases; derived from the term "cohesive ends") at either end. Reannealing the *cos* sites allows λ chromosomes to circularize when they enter a host cell; cutting the DNA at the *cos* site opens the circle into a linear molecule (fig. 13.7). Geneticists have taken advantage of these *cos* sites to clone even larger segments of foreign DNA because it turns out that even 24 kb is not adequate to study some eukaryotic genes or gene groups. Many eukaryotic genes are very large because of their introns and transcriptional control segments. DNA up to 50 kb can be cloned if *cos* segments are attached to either end with a plasmid origin of DNA replication and a selectable antibiotic gene. These *cos*-site-containing plasmids are called **cosmids** (fig. 13.7). Cosmids not only allow the cloning of very large pieces of DNA, they actually select for large segments of foreign DNA because small cosmids are not incorporated into phage heads. Thus, foreign DNA ranging from 2.5 to 50 kb in size can be cloned using plasmids, Charon phages, or cosmids. (Much larger pieces of DNA, about a million bases, can be cloned in yeast, as we will describe later.)

Selecting for Hybrid Vectors

In the methods we have described, restriction enzymes separately cut both vector and foreign DNA. The two are then mixed in the presence of ligase. The many products that are created can be divided generally into three categories: vectors with foreign DNA, vectors without foreign DNA, and fragments. In a later section, we will discuss methods of finding a particular piece of foreign DNA in a vector. Here, we point out how vectors with inserts of any kind are selected.

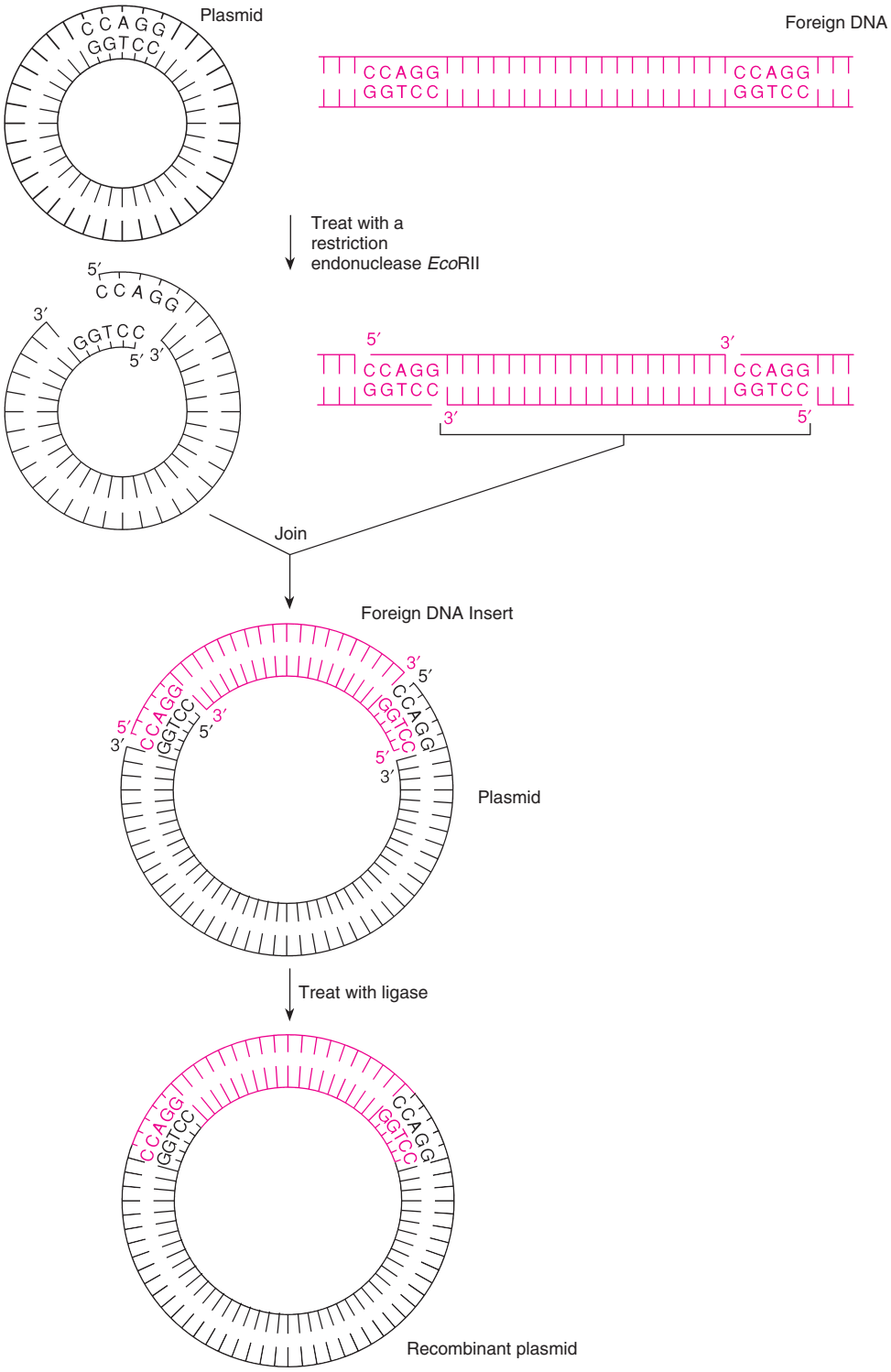


Figure 13.6 Formation of a recombinant plasmid. The same restriction endonuclease, in this case *Eco*RII, is used to cleave both host and foreign DNA. Some of the time, cleaved ends will come together to form a plasmid with an insert of the foreign DNA. Ligase seals the nicks. P. Berg was the first scientist to clone a piece of foreign DNA when he inserted the genome of the SV40 virus into phage λ in 1973.

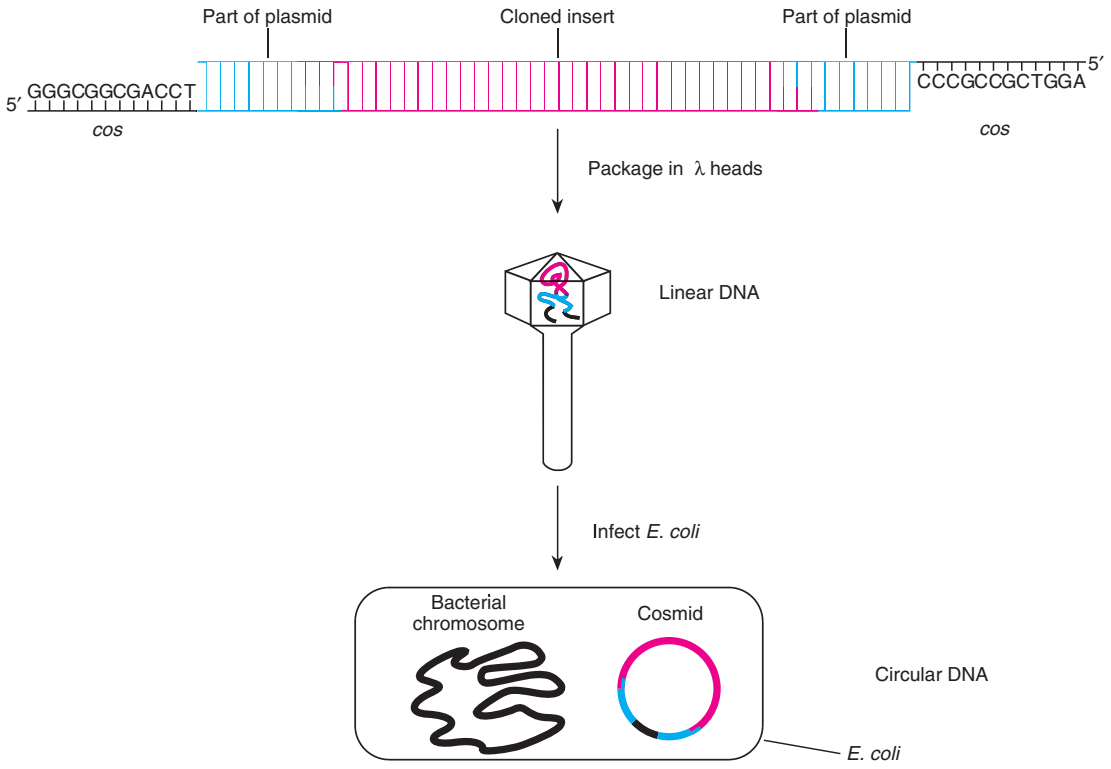


Figure 13.7 A cosmid is a plasmid with *cos* sites that can be transferred into bacteria within phage lambda heads, a very efficient method of infection. The *cos* sites are single-stranded; they reanneal to a *circle* when inside the host. (The heavy lines of the linear DNA, bacterial chromosome, and cosmid are double-stranded DNA.)

Charon phages are selected simply by their ability to infect *E. coli* cells. As we mentioned, after manipulation, only λ DNA with a foreign insert is packaged because of the size requirement. Plasmids that contain foreign DNA can be selected through screening for antibiotic resistance. For example, a widely used cloning plasmid is named pBR322. (Plasmids are often named with the initials of their developers. The vector pBR322 was first described in a paper published in 1977 by authors F. Bolivar and R. Rodriguez, hence pBR.) Plasmid pBR322 contains genes for tetracycline and ampicillin resistance and various restriction sites. There is, for example, a *Bam*HI site in the tetracycline-resistance gene (fig. 13.8). After the ligating procedure, plasmids with and without foreign DNA will be present. *E. coli* cells are then exposed to this DNA mixture in the presence of calcium chloride; after taking up the DNA, the *E. coli* cells are plated on a medium without antibiotics. Replica-plating is done onto plates with one or both antibiotics. Colonies resistant to both antibiotics are composed of cells with plasmids having no inserts; those resistant only to ampicillin have a plasmid with an insert. Colonies resistant to neither antibiotic have cells with no plasmids.

Blunt-End Ligation

Restriction endonuclease treatment may not suffice for cloning; an endonuclease may cut in the wrong place, say in the middle of a desired gene, or the foreign DNA may have been isolated by other methods, such as physical shearing. In these cases, several other methods of cloning can be used.

The most common method of joining foreign and vehicle molecules that do not have sticky ends is called **blunt-end ligation;** the phage enzyme, T4 DNA ligase, can join blunt-ended DNA. Blunt ends can be generated when segments of DNA to be cloned are created by physically breaking the DNA or by using certain restriction endonucleases, such as *Hin*dII (see fig. 13.2), that form blunt ends. Since the ligase is nonspecific about which blunt ends it joins, many different, unwanted products result from its action. Restriction enzymes that produce sticky ends are preferred for cloning.

A variation of blunt-end ligation uses **linkers**—short, artificially synthesized pieces of DNA containing a restriction endonuclease recognition site. When these linkers are attached to blunt pieces of DNA and then

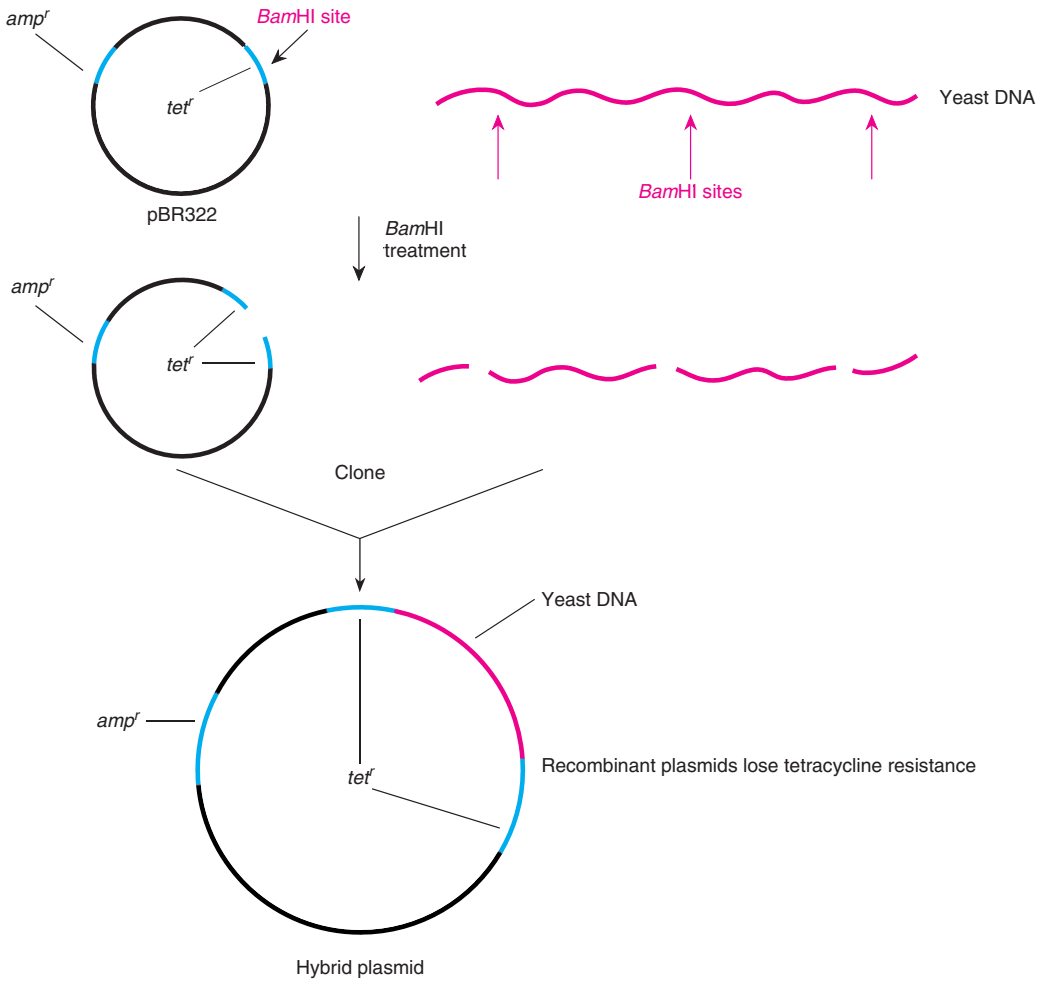


Figure 13.8 *E. coli* plasmid pBR322. This plasmid carries two genes, amp^r and tet^r, that confer resistance to ampicillin and tetracycline, respectively. A *Bam*HI restriction site occurs within the tet^r gene. A cloned fragment within the tet^r gene therefore destroys the tetracycline resistance. (Heavy *black, blue,* and *red* lines represent double-stranded DNA.)

treated with the appropriate restriction endonuclease, sticky ends are created. In figure 13.9, the linkers are twelve base-pair (bp) segments of DNA with an *Eco*RI site in the middle. They are attached to the DNA to be cloned with T4 DNA ligase. Subsequent treatment with *Eco*RI will result in DNA with *Eco*RI sticky ends.

DNA for cloning can be obtained generally in two ways: (1) a desired gene or DNA segment can be synthesized or isolated or (2) the genome of an organism can be broken into small pieces and the small pieces can be randomly cloned **(shotgun cloning).** Then the desired DNA segment must be "fished" out from among the various clones created. Let us look first at synthesizing or isolating a desirable gene before cloning it, and then look at the process of locating a desired gene after it has been cloned.

Cloning a Particular Gene

Creating DNA to Clone

To clone a particular gene (or DNA segment), a scientist must have a purified double-stranded piece of DNA containing that gene. There are numerous ways to obtain that DNA; several entail creating or isolating a single-stranded messenger RNA that is then enzymatically converted into double-stranded DNA. The problem is then reduced to obtaining the desired messenger RNA.

The messenger RNA for a particular gene can be obtained in several different ways, depending on the particular gene. If large quantities of the RNA from a particular cell are available, the RNA can be isolated directly. For example, mammalian erythrocytes have abundant quanti-

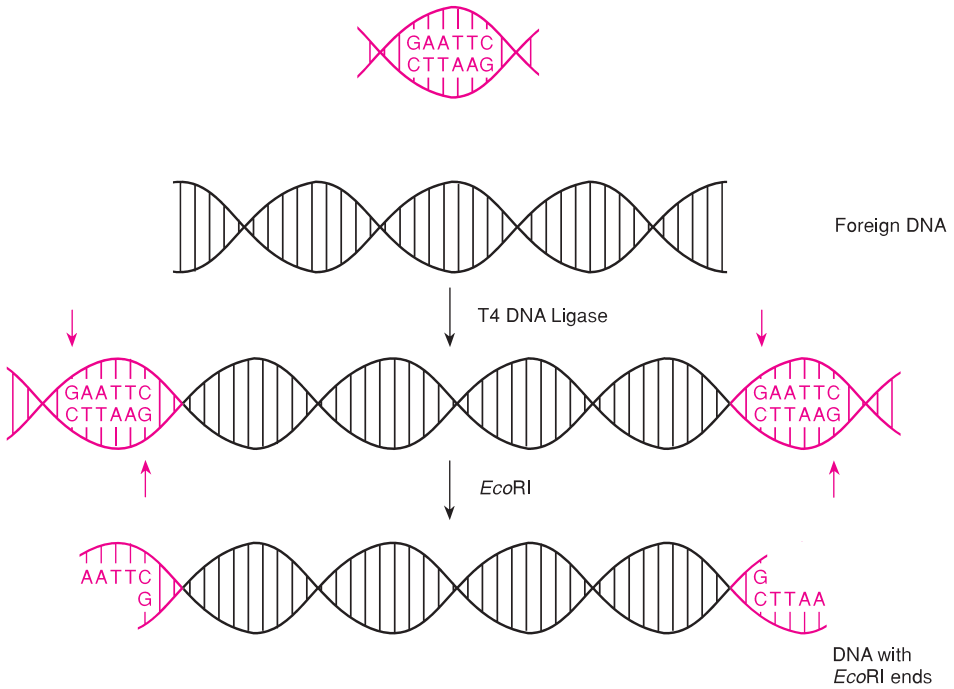


Figure 13.9 Linkers: small segments of DNA with an internal restriction site. Linkers can be added to blunt-ended DNA by T4 DNA ligase. The restriction enzymes create DNA with ends that are compatible with any DNA cut by the same restriction enzyme (in this case, *Eco*RI).

ties of α- and β-globin messenger RNAs. Also, ribosomal RNA and many transfer RNAs are relatively easy to isolate in quantities adequate for cloning.

Double-stranded DNA for cloning is made from the purified RNA with the aid of the enzyme reverse transcriptase, isolated from RNA tumor viruses (see chapter 10). We describe here the conversion of RNA to DNA using a eukaryotic messenger RNA with a 3′ poly-A tail (fig. 13.10). In the first step, a poly-T primer is added, which base-pairs with the poly-A tail of the messenger RNA. This short, double-stranded region is now a primer for polymerase activity—a free 3′-OH exists. The primed RNA is then treated with the enzyme reverse transcriptase, which will polymerize DNA nucleotides using the RNA as a template. The result is a DNA-RNA hybrid molecule (fig. 13.10*c*).

The hybrid is now treated with the enzyme RNaseH, which creates random nicks in the RNA part of the RNA-DNA hybrid. These nicks provide the primer configuration for repair synthesis, the same repair done on Okazaki fragments when RNA primer is removed and replaced by DNA. Thus, the hybrid is treated with DNA polymerase I, which replaces each small RNA segment with DNA, base by base. Finally, the short DNA segments of the second DNA strand are united with DNA ligase (fig. 13.10*f*). The resulting double-stranded DNA is referred to as **complementary DNA (cDNA).** Hence, starting with a piece of single-stranded messenger RNA, we have generated a piece of double-stranded DNA. This piece can now be cloned using the blunt-end methods we have described.

If the RNA is not available in large enough quantities, it is possible to synthesize DNA in vitro if the amino acid sequence of its expressed protein is known. A possible nucleotide sequence can be obtained from the genetic code dictionary (see table 11.4) if the sequence of amino acids is known from the protein product of the gene. This method will probably not re-create the original DNA because of the redundancy in the genetic code. In other words, any one of six different codons could have coded a particular leucine in a protein. Despite an element of guesswork, it is possible to synthesize a piece of DNA that will code for a particular protein. Currently, automated machines that add one base at a time in ten-minute cycles can synthesize DNA sequences of over one hundred bases.

You will notice that the methods we have described, making cDNA or synthetic DNA using the genetic code dictionary, produce DNA missing the gene's promoter and its transcriptional control sequences as well as other untranslated areas of the DNA (introns). If it is desirable to clone an intact gene with its promoter and introns, then cloning can be done by creating random pieces of

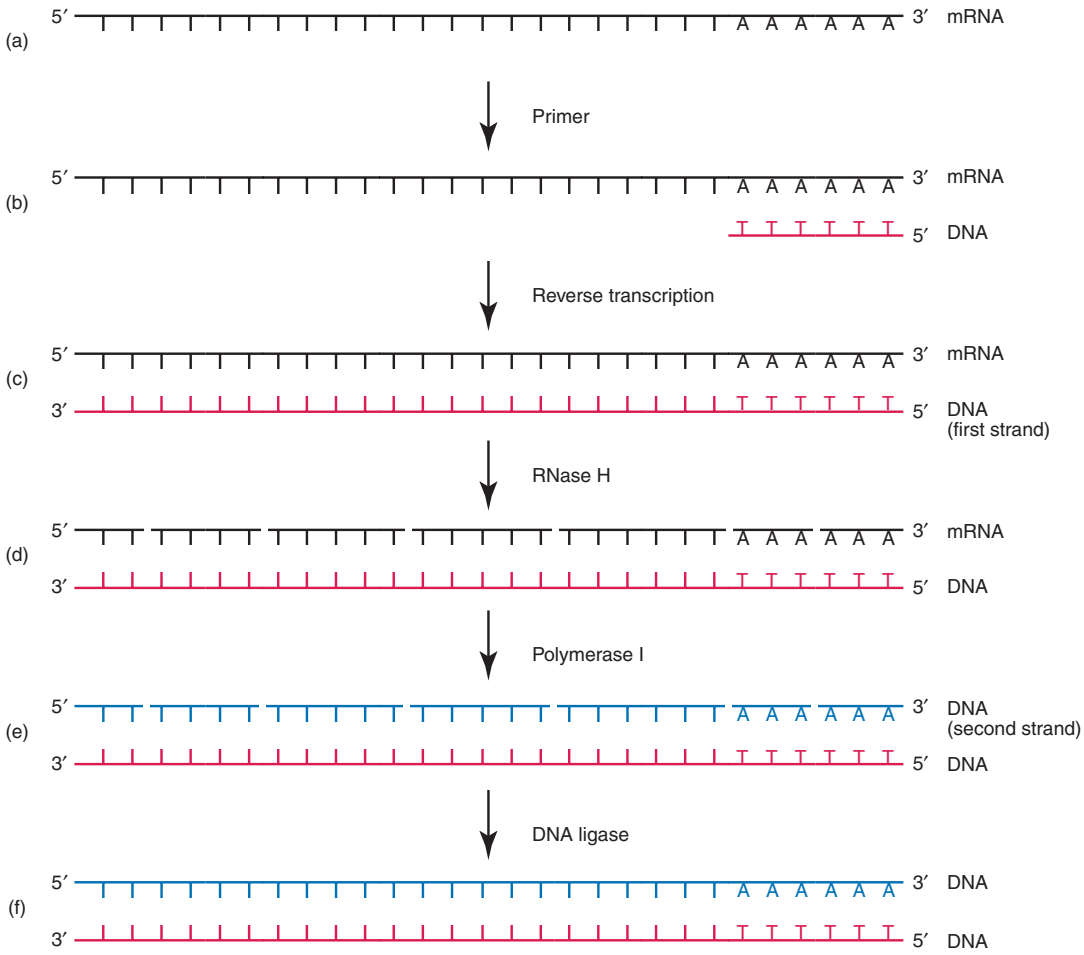


Figure 13.10 (*a*) A messenger RNA, shown in *black,* begins as a single strand. (*b*) A poly-T DNA segment (*red*) is added as primer; it complements the 3′ poly-A tail of the eukaryotic messenger RNA. (*c*) Reverse transcriptase acts on this primed configuration to synthesize a single strand of DNA from the RNA template. (*d*) The RNA is then nicked randomly by RNase H. (*e*) The RNA segments are then replaced by DNA (*blue*) by the action of DNA polymerase I. (*f*) After DNA ligase treatment, the final result is double-stranded complementary DNA (cDNA).

the genome. The gene of interest can be found either before or after cloning it, although it is usually done after cloning.

Creating a Genomic Library

When cDNA or synthetic DNA cannot be used for cloning, the total DNA of an organism can be broken into small pieces to isolate the desired gene or DNA fragment. The desired DNA can be isolated either before or after cloning. This DNA is referred to as genomic DNA to differentiate it from cDNA.

If the original DNA is isolated before cloning, then only that DNA need be cloned. Alternatively, a "shotgun" approach can be used to clone a sample of the entire genome of an organism (in small pieces, of course), creating a **genomic library,** a set of cloned fragments of the original genome of a species (fig. 13.11). In a genomic library, a desired gene can be located after it is already cloned.

Southern Blotting

When DNA segments are generated randomly, usually by endonuclease digestion, a desired gene must be located. As mentioned, we can look for the gene either before or after it is cloned. We consider first the procedure for locating a specific gene in a DNA digest, before the DNA has been cloned.

To locate a specific gene in the midst of a DNA digest, one must have a specific **probe.** Probes are generally nucleic acids with sequences that precisely locate a complementary DNA sequence by hybridization. The probes are labeled so they can be identified later with autoradiography or **chemiluminescent techniques** (techniques in which tags are used that fluoresce under ultraviolet or laser light). Thus, if we wish to locate the gene for β-globin, we could use radioactively labeled β-globin messenger RNA or radioactively labeled cDNA. RNA-DNA or DNA-DNA hybrids would form between the specific gene and the radioactive probe. Autoradiography or chemiluminescence would then locate the radioactive probe.

Let us assume that we wanted to clone the rabbit β-globin gene. First, we would create a restriction digest of rabbit DNA (fig. 13.12). We would then subject this digest to electrophoresis on agarose to separate the various fragments according to size. Agarose is a good medium for separating DNA fragments of a wide variety of sizes.

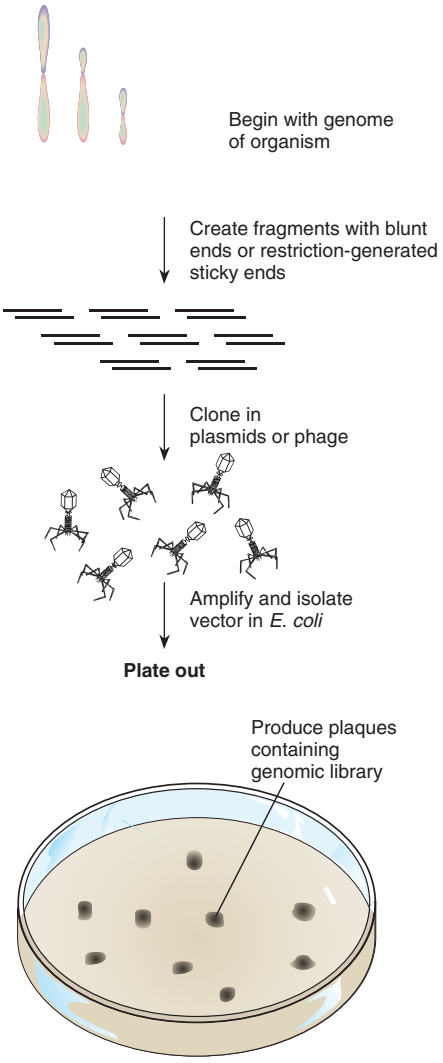


Figure 13.11 Creating a genomic library using the shotgun approach in creating inserts. First, the genome is fragmented. The fragments are then cloned randomly in vectors. The collection of these vectors is referred to as a genomic library.

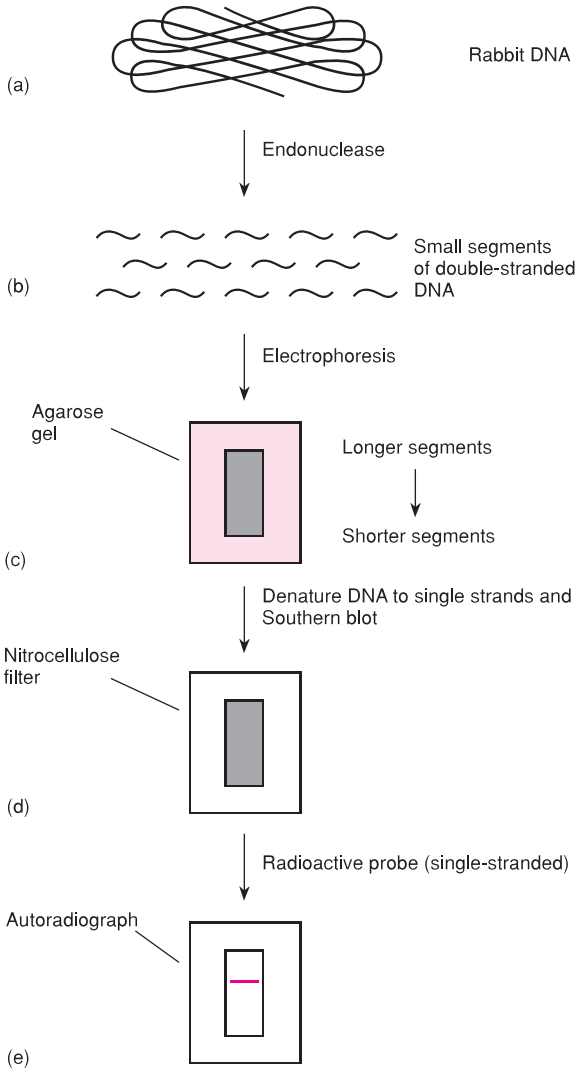


Figure 13.12 Locating the rabbit β-globin gene within a DNA digest using the Southern blotting technique. The rabbit DNA (*a*) is segmented with a restriction endonuclease (*b*) and then electrophoresed on agarose gels (*c*). Southern blotting transfers the DNA to nitrocellulose filters (*d*). Finally, a radioactive probe (β-globin messenger RNA) locates the DNA fragment with the β-globin gene after autoradiography (*e*).

Edwin M. Southern (1938–). (Courtesy of Edwin Southern.)

In a digest of this kind, however, there are usually so many fragments that the result is simply a smear of oligonucleotides, from very small to very large. To proceed further, we have to transfer the electrophoresed fragments to another medium for probing, or the DNA fragments would diffuse out of the agarose gel. Nitrocellulose filters or nylon membranes are excellent for hybridization because the DNA fragments bind to these membranes and will not diffuse out. The transfer procedure, first devised by E. M. Southern, is called **Southern blotting.** In this technique, the double-stranded DNA on the agarose gel is first denatured to single-stranded DNA, usually with NaOH. Then the agarose gel is placed directly against a piece of nitrocellulose filter, and the resulting sandwich is placed agarose-side-down on a wet sponge. Dry filter paper placed against the nitrocellulose side wicks fluid from the sponge, through the gel, and past the nitrocellulose filter, carrying the DNA segments from the agarose to the nitrocellulose (fig. 13.13). NaOH is used as the transfer solution in the tray. The DNA digest fragments are then permanently bound to the nitrocellulose filter by heating. DNA-DNA hybridization takes place on the filter. (A similar technique can be performed on RNA, which is called, tongue-in-cheek, **northern blotting.** Immunological techniques, not involving nucleotide complementarity, can be used to probe for proteins in an analogous technique called **western blotting.**)

A labeled probe can be obtained in several different ways. In this example, the easiest way to obtain a radioactive probe would be to isolate β-globin messenger RNA from rabbit reticulocytes and construct cDNA using the reverse-transcriptase method described. The deoxyribonucleotides used during reverse transcription are then synthesized to contain radioactive phosphorus, ^{32}P. As figure 13.12 shows, after hybridization, a single radioactive band locates a DNA segment with the β-globin gene. Note that the probe, originating from messenger RNA, will lack the introns present in the gene. However, probing is successful as long as there are complementary regions in the two nucleotide strands.

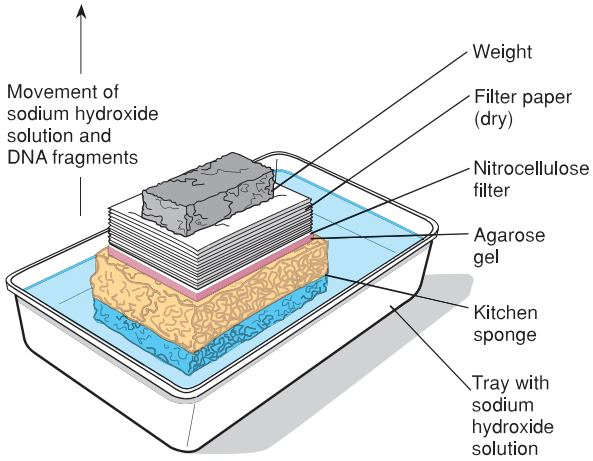


Figure 13.13 Arrangement of gel and filters in the Southern blotting technique. The NaOH buffer is drawn upwards by the dry filter paper, transferring the DNA from the agarose gel to the nitrocellulose filter.

To clone the β-globin gene, a second agarose gel would be run with a sample of the digest used in figure 13.12. That gel, not subject to DNA-DNA hybridization, would have the β-globin segment in the same place. The band, whose location is known from the autoradiograph, could be cut out of the agarose gel to isolate the DNA. We could then clone the DNA by methods discussed earlier in the chapter.

Probing for a Cloned Gene

Dot Blotting

The methods we have described are also useful in locating genes already cloned within plasmids, for example, after a genomic library has been constructed. In this case, electrophoresis and Southern blotting are not needed since we will be probing for a particular sequence of DNA already cloned rather than DNA segments within a digest.

For example, the DNA of a human-mouse hybrid cell line was cloned in order to locate human DNA. In this case, a hybrid cell line had only one human chromosome, chromosome 20. In order to locate DNA from that chromosome, probes were used that were isolated from human chromosomes. The probes were radioactively labeled. Meanwhile, 288 *E. coli* colonies, each containing a hybrid plasmid, were grown and transferred directly to a nitrocellulose filter. In preparation for probing, the cells were lysed and their DNA denatured. The plasmid DNA within the cells of each clone was then hybridized with the radioactive probes. Figure 13.14 is an autoradiograph of the 288 clones. The two dark spots indicate clones car-

rying DNA from human chromosome 20 (those clones "light up" autoradiographically). This technique, hybridization of cloned DNA without an electrophoretic-separation step, is referred to as **dot blotting.**

These techniques can also be carried out without the grid arrangement of colonies by using replica plating as described in chapter 7. Thus, a specific gene can be located after shotgun cloning.

Western Blotting

An entirely different method used to locate particular cloned genes utilizes the actual expression of the cloned genes in the plasmid-containing cells. If a eukaryotic gene is cloned in an *E. coli* plasmid downstream from an active promoter, that gene may be expressed (transcribed and translated into protein). Plasmids that allow the expression of their foreign DNA are termed **expression vectors.** There are many problems with this technique because bacteria normally would not express eukaryotic genes. However, special vectors have been developed in which the cloning site is just downstream from a promoter. The eukaryotic gene thus becomes part of the prokaryotic gene, producing a fusion protein, usually with only a few amino acids from the prokaryote. Of course, the eukaryotic gene must be in the correct orientation and in the correct codon reading frame for appropriate translation, meaning that the success rate of this technique is relatively low.

A particular protein product can be located by western blotting, a method completely analogous to either Southern blotting or dot blotting. In this technique, probing is done with antibodies specific for a particular protein, rather than using a radioactive oligonucleotide probe. A second antibody, specific to the first and labeled with a marker, usually fluorescent, locates the first antibody. For example, assume we are looking for the expression of a particular protein in the clones of figure 13.14. The clones would be transferred to a nitrocellulose membrane, where they would be lysed (e.g., with chloroform vapor). Then an antibody, specific for the particular protein, would be applied. A second antibody, specific for the first antibody and labeled with a fluorescent marker, would be applied to the filters. Fluorescence of the second antibody would locate the presence of the first antibodies and thus indicate which of the clones is expressing the particular gene (fig. 13.15).

Eukaryotic Vectors

The work we have described so far involves introducing chimeric plasmids into bacteria, primarily *E. coli.* However, there are several reasons why we want to extend these techniques to eukaryotic cells (box 13.1). First, a prokaryote like *E. coli* is not capable of fully expressing some eukaryotic genes since it lacks the enzyme systems necessary for some posttranscriptional and posttranslational modifications such as intron removal and some protein modification. Second, we also wish to study the organization and expression of the eukaryotic genome in

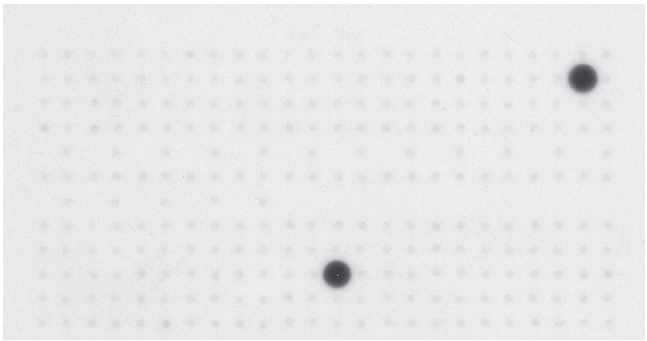


Figure 13.14 Dot blot autoradiograph of 288 clones of DNA from a mouse-human hybrid cell line. After lysing samples of each clone on a nitrocellulose filter, the investigator hybridized the clones with radioactive probes for human-specific sequences. The two *dark spots* indicate clones carrying human DNA. The slight background radiation in most other spots provides the spot pattern needed to orient the investigator. (Source: Courtesy of Nick O. Bukanov.)

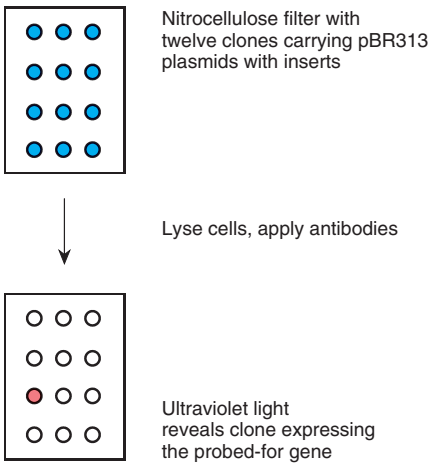
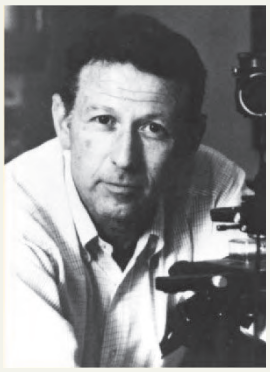


Figure 13.15 Western blot technique is used to locate an expressed protein from among many clones. Clones that may carry the expressed protein are lysed. Tagged antibodies are applied to locate the protein; a second antibody locates the first antibody either through a fluorescent color marker, as shown here, or with autoradiography.

BOX 13.1



Paul Berg (1926–).
(Courtesy of Dr. Paul Berg.)

Ethics and Genetics

The Recombinant DNA Dispute

Paul Berg shared the 1980 Nobel Prize in chemistry for creating the first cloned DNA molecule, a hybrid λ phage that contained the genome of the simian tumor virus, SV40. The fact that he could do this work was worrisome to many people, himself included. The recombinant DNA dispute was underway. Berg voluntarily stopped inserting tumor virus genes into phages that attack the common intestinal virus *E. coli.*

People continue to worry about the dangers of working with recombinant DNA. One immediate and obvious concern is that cancer or toxin genes will "escape" from the laboratory. In other words, recombinant DNA technology could create a bacterium or plasmid that contained toxin or tumor genes. The modified bacterium or plasmid could then accidentally infect people. A 1974 report by the National Academy of Sciences led to a February 1975 meeting, which took place at the Asilomar Conference Center south of San Francisco. Berg convened this meeting, which over one hundred molecular biologists attended. The recommendations of the Asilomar Committee later formed the basis for official guidelines developed by the National Institutes of Health (NIH). In essence, NIH established guidelines of containment.

Containment means erecting physical and biological barriers to the escape of dangerous organisms. The NIH guidelines defined four levels of risk, from minimal to high, and four levels of physical containment for them (called P1 through P4). The most hazardous experiments, dealing with the manipulation of tumor viruses and toxin genes, require extreme care, which included negative-pressure air locks to the laboratory and experiments done in laminar-flow hoods, with filtered or incinerated exhaust air.

Biological containment means developing host cells and manipulated vectors that are incapable of successful reproduction outside the lab, even if they escape. High-risk work was done with host cells or vectors that were modified. For example, a bacterium of the *E. coli* strain EK2 cannot survive in the human gut because it has mutations that do not permit it to synthesize thymine or diaminopimelate. The lack of thymine-

vivo (in the living system), something we can only accomplish by working directly with eukaryotic cells. Finally, we wish to learn how to manipulate the genomes of eukaryotes for medical as well as economic reasons. To these ends, we discuss eukaryotic plasmids and the direct manipulation of eukaryotic genomes in vivo.

Yeast Vectors

Yeasts, small eukaryotes that can be manipulated in the lab, like prokaryotes, have been studied extensively. Baker's yeast, *Saccharomyces cerevisiae,* has a naturally occurring plasmid. In addition, bacterial plasmids have been introduced into yeast. Unfortunately, the cells tend to lose these plasmids. This tendency has been overcome, however, by constructing bacterial plasmids that contain a yeast centromere (CEN) and the origin of yeast DNA replication (ARS for autonomously replicating sequence; fig. 13.16). The yeast then carries the plasmids from one generation to the next. The plasmids can have telomeric sequences inserted, and they can then be made linear by cutting the telomeric sequences with endonuclease. Alternatively, the plasmids can be linearized first, and then have telomeric sequences added to their ends. The plasmids are then called **yeast artificial chromosomes (YACs).** The particular advantage YACs have is that they are capable of accepting very large pieces of inserted DNA. Remember that a cosmid can hold about 50 kb; a YAC can hold as much as 800 kb or more. The ability to clone this much DNA is valuable when working with large eukaryotic genes and in the Human Genome Project (see the section with this title later in the chapter).

Recombinant DNA studies in yeast have increased our knowledge about gene regulation in eukaryotes, about how the centromere works, and about the way in which the tips of eukaryotic linear chromosomes are replicated. In addition, YACs have allowed us to analyze and sequence very large segments of eukaryotic DNA.

synthesizing ability is lethal because the cell cannot replicate its DNA. The diaminopimelate is a cell-wall constituent; without it, the cells burst. These bacteria also carry mutations that make them extremely sensitive to bile salts. Thus, if by accident the cells were to escape, they would pose virtually no threat. The plasmids used for recombinant research were modified so that they could not be transferred from one cell to the next. Again, if containment failed, neither the host cells nor their plasmids would survive.

In 1979, the guidelines were relaxed. Although it was wise to be cautious, it appears that initial fears were unwarranted. Recombinant DNA work now seems to pose little danger: Containment works very well, and engineered bacteria do very poorly under natural conditions. *E. coli* has been living in mammalian guts for millions of years, so it has had numerous opportunities to incorporate mammalian DNA into its genome (intestinal cells are dying and sloughing off into the gut all the time). No "Andromeda strain" has arisen, nor do we foresee one in the future.

Current concern is focused on the acceptability of genetically modified crops (GM crops). As we will discuss later, one fourth of American cropland is planted with genetically modified crops, modified mainly for insect resistance. These modifications have curtailed our use of insecticides. (For cotton and corn, for example, liquid insecticide use dropped by 3.6 million liters and powdered insecticide by 300,000 kilograms in 1999.) However, people are concerned with the effects these modifications might have on natural ecosystems: How many valuable insects will be killed by mistake? Although Third World countries are desperate for these technologies, the United States, European and Asian trading partners are demanding that the crops we export be genetically unmodified. Farmers are also concerned that genetically modified crops have been modified to be sterile (so called "terminator technology") so that farmers would need to buy new seeds each year.

More recently, the recombinant DNA dispute has taken a whole new twist. It now has surfaced as a conflict between academic freedom and industrial secrecy. It seems that recombinant DNA technology is very lucrative. Numerous academic scientists have either begun genetic engineering companies or become affiliated with pharmaceutical companies. However, the philosophies of private enterprise and academia are often in conflict. Academic endeavors are presumably open, with free exchange of information among colleagues, whereas private enterprise entails some degree of secrecy, at least until patents are obtained to protect the investments of the companies. Thus, a basic conflict can arise for scientists trained in gene cloning. The conflict has been prevalent since late 1980, when the first patent for recombinant DNA techniques was awarded to Stanford University and the University of California. When, in April 2000, United States President Bill Clinton and British Prime Minister Tony Blair issued a joint statement asking that human genome data not be patented, the American stock market took a major downturn. This is a tumultuous time for biotechnology.

Animal Vectors

The vehicle most commonly used in higher animals is the DNA tumor virus SV40. (SV, or simian vacuolating virus, was first isolated in monkeys; however, it can transform normal mouse, rabbit, and hamster cells. Unlike the use of the word *transformation* in bacteria, transformation in eukaryotes refers to the changing of a normal cell into a rapidly growing, cancerous one.) SV40 is an icosahedral particle with a small (5,224 base pairs) chromosome, which is a circular, double-stranded DNA molecule.

Like λ vectors, SV40 virions allow foreign DNA to replace part of their DNA. The viruses can then be used in recombinant DNA studies in one of two ways (fig. 13.17). They can replicate and complete their life cycle with the help of nonrecombinant viruses, or they can replicate in the host without making active virus particles by existing as circular plasmids in the cytoplasm or by integrating into the host's chromosomes. SV40 has become a valuable tool in mammalian genomic studies. For example, the rabbit β-globin gene was cloned in SV40, and enhancer sequences (see chapter 10) were discovered in SV40. DNA tumor viruses have also deepened our understanding of transformation in eukaryotes (oncogenesis).

Plant Vectors

The best-studied system for introducing foreign genes into plants is the naturally occurring crown gall tumor system. The soil bacterium *Agrobacterium tumefaciens* causes tumors, known as crown galls, in many dicotyledonous plants (fig. 13.18). In essence, the crown gall is made of transformed plant cells. These cells have been transformed by a plasmid within the bacterium called the *tumor-inducing,* or *Ti,* plasmid. Transformation occurs when a piece of the plasmid called T-DNA (for transferred DNA) is integrated into the chromosome of the plant host. Crown gall cells produce amino acid derivatives, termed *opines,* that the *A. tumefaciens* cells use. By manipulating this

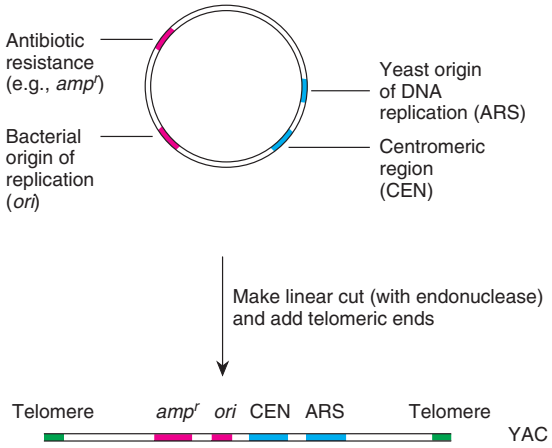


Figure 13.16 *Escherichia coli* plasmid pBR322 modified for use in yeast. This plasmid survives and replicates in both yeast and *E. coli* because it contains the origin of replication for both, as well as a yeast centromeric region (CEN). When it is made linear and telomeres are added, the yeast artificial chromosome (YAC) becomes suitable for cloning large pieces of DNA.

system, geneticists have begun to understand the transformation process in plants as well as to develop a manipulatable system for introducing foreign genes into plants.

The study of genetics in plants has been boosted a great deal by the availability of model organisms similar to *E. coli,* yeast, and fruit flies. Recently, much attention has focused on the meadow weed, *Arabidopsis thaliana* (fig. 13.19). This small plant is ideal for studying plant genetics because its genome is small, approximately 100 million base pairs located in only five chromosomes ($2n = 10$). This is only about five times the genome of yeast or twenty times the genome of *E. coli.* Thus, in terms of genome size, it is quite manageable. *A. thaliana* has joined the ranks of organisms whose genomes have been sequenced. The plants are easy to grow in very large numbers, and each plant produces as many as ten thousand seeds. Hence, this organism compares very favorably with fruit flies and yeast for studying questions of gene control in a eukaryote, in this case a plant.

Expression of Foreign DNA in Eukaryotic Cells

Foreign DNA can be introduced into eukaryotic cells in methods similar to bacterial transformation. However, the process in eukaryotes is called **transfection** because, as we described, the term *transformation* in eukaryotes is used to mean cancerous growth. Eukaryotic organisms that take up foreign DNA are referred to as **transgenic.** Most of the techniques described here transcend taxonomic lines.

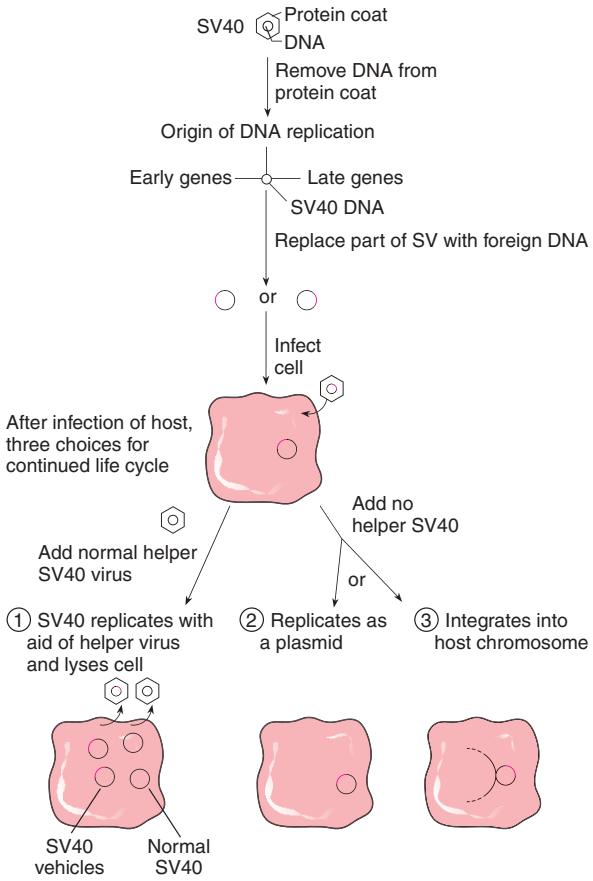


Figure 13.17 SV40 virus can be used as a gene cloning vehicle. Although part of the virus is replaced by inserted DNA during cloning, it can still replicate with the aid of normal helper viruses (nonrecombinant SV40). Without the aid of helper viruses, it can either replicate as a plasmid or integrate into the host chromosome.

Figure 13.18 Crown gall on tobacco plant (*Nicotiana tabacum*) produced by *Agrobacterium tumefaciens* containing Ti plasmids. (Courtesy of Robert Turgeon and B. Gillian Turgeon, Cornell University.)

Animal cells, or plant cells with their walls removed (protoplasts), can take up foreign chromosomes or DNA directly from the environment with a very low efficiency (in the presence of calcium phosphate). Directly injecting the DNA greatly improves the efficiency. For example, transgenic mice are now routinely prepared by injecting DNA into either oocytes or one- or two-celled embryos obtained from female mice after appropriate hormonal treatment (fig. 13.20). After injection of about 2 picoliters (2×10^{-12} liters) of cloned DNA, the cells are reimplanted into the uteruses of receptive female hosts. In about 15% of these injections, the foreign DNA incorporates into the embryo. Transgenic animals are used to study the expression and control of foreign eukaryotic genes. In 1988, a transgenic mouse prone to cancer was the first genetically engineered animal to be patented. This mouse provides an excellent model for studying cancer (see chapter 16). (A controversy arose as to whether engineered higher organisms should be patentable; currently they are.) Mice have already been successfully transfected with a rat growth-hormone gene (fig. 13.21), and transgenic sheep have been produced that express the gene for a human clotting factor. The latest recombinant DNA dispute arises from the cloning of sheep in 1997 (box 13.2).

Transfection can also be mediated by retroviruses (RNA viruses containing the gene for reverse transcriptase). For example, a retroviral vector infected and repaired human white blood cells lacking the enzyme

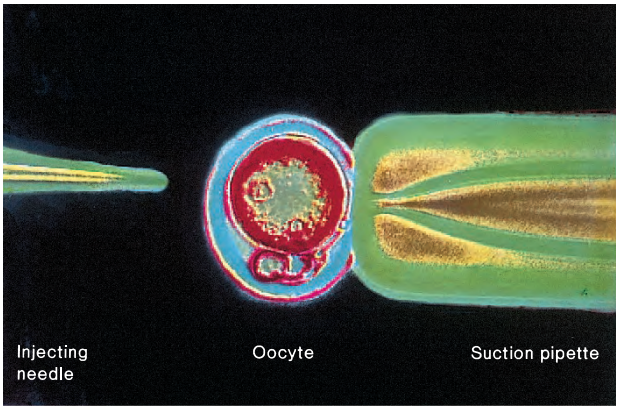


Figure 13.20 Injection of DNA into the nucleus (germinal vesicle) of a mouse oocyte. The oocyte is held by suction from a pipette. (© John Gardon/Phototake.)

Figure 13.19 A dwarf form of the plant *Arabidopsis thaliana.* (Source: *Science,* Vol. 243, March 10, 1989, cover. ©1989 AAAS, Washington, D.C. Photo by DeVere Patton. Courtesy of E. I. DuPont de Nemours and Company.)

Figure 13.21 Mouse littermates. The larger one is a transgenic mouse containing the rat growth-hormone gene. (Source: Richard D. Palmiter, Ralph L. Brinster, et al., "Dramatic growth of mice that develop from eggs microinjected with metallothionein growth hormone fusion genes," *Nature* 300, 16 December 1982, cover. Photograph by Dr. Ralph L. Brinster. Copyright © 1982, Macmillan Magazines Ltd.)

BOX 13.2

Ethics and Genetics

Cloning Dolly

Fiction writers in the past have created stories in which scientists cloned one person to create numerous copies. The themes of these stories have varied from the cloning of Adolph Hitler to the cloning of a very busy man to help him fulfill his day-to-day obligations. The possibility of those scenarios came a bit closer to reality in February 1997 when a group of scientists from the Roslin Institute and PPL Therapeutics, both in Edinburgh, Scotland, reported in *Nature* magazine that they had successfully cloned a sheep from a cell taken from the udder of a six-year-old ewe. The cloned lamb was named Dolly (fig. 1). In the past, genetically identical animal embryos had been created only with amphibian cells, and those created from adult nuclei had never successfully reached adulthood. Cloning in which the nuclei came from fetal cells or cells from cell lines had been successful before in mammals.

To clone an animal, it is necessary to begin with an egg, the only cell known to initiate and support development. In order to clone an individual, using the word *clone* to mean create a genetically identical copy, it is necessary to get an egg without a nucleus and then to transplant a nucleus of known origin. Techniques for nuclear transplantation had been worked out with frogs and toads in the 1950s. The Scottish scientists succeeded in obtaining sheep eggs, enucleating them (removing their nuclei), and then transferring in donor nuclei by fusing the donor cells and the enucleated eggs with an electrical pulse. The electrical pulse also initiated development of the egg. Although only one pregnancy of the twenty-nine initiated was successful, the lamb that was born seems normal in every way; it has since produced offspring.

Others had tried this type of experiment with many types of animals, including mice. They were not successful for numerous reasons. The most likely explanation for the recent success, according to the scientists, is that the donor cells were kept in a nongrowth phase for several days, which may have synchronized them with the oocyte. Thus, the nucleus and the oocyte were at the same stage of the cell cycle and thus compatible. Other reorganizations that had to take place in the donor chromosomes are not really known for certain, but one thing is clear: the nucleus of an adult cell in the sheep has all of the genetic material needed to support normal growth and development of an egg. (The work has since been repeated with goats, cattle, and mice.)

There are numerous ramifications to the success of this work. First, mammal cloning could become a routine procedure. This would allow us to study mammalian development and to replicate genetically identical individuals, particularly transgenic animals that would have particular genomes of value. We can also use these techniques to study aging, since an "old" nucleus is initiating the development of a new organism. Also of interest is the interaction of a particular genome with a particular cytoplasm, since the cytoplasm contains not only the materials needed for early development, but also cell organelles, including mitochondria that have their own genetic material.

Finally, ethical issues must be considered if this technique is successful with human beings. Parents might wish to clone a deceased child or to obtain an immunologically compati-

adenosine deaminase. A retrovirus responsible for a form of leukemia in rodents, the Moloney murine leukemia virus, was engineered so that all the viral genes were removed and replaced with an antibiotic marker (neomycin resistance) and the human adenosine deaminase gene. The virus binds to the cell surface and is taken into the cell, its RNA is converted to DNA by reverse transcription, and the DNA is incorporated into one of the cell's chromosomes. It is not possible for this highly modified virus to attack and damage the cells unless a helper virus is added. Unlike the SV40 viruses in figure 13.17, the modified Moloney viruses cannot initiate a successful infection without the helpers because vital genes have been removed.

Three other recent techniques deliver recombinant DNA to eukaryotic cells: electroporation, liposome-mediated transfer, and "biolistic" transfer. In **electroporation,** exogenous DNA is taken up by cells subjected to a brief exposure of high-voltage electricity. Presumably, this electric field creates transient micropores in the cell membrane, allowing exogenous DNA to enter.

Liposome-mediated transfection is a technique that encapsulates foreign DNA in artificial membrane-bound vesicles called **liposomes.** The liposomes are then used to deliver their DNA to target cells. In one experiment, 50% of mice injected with these DNA-containing liposomes were successfully transfected—they expressed the proteins the transfecting DNA encoded.

Last are techniques developed to deliver foreign DNA into mitochondria and chloroplasts. These have proven difficult targets for genetic engineering because,

ble sibling for a child who needs an organ or bone marrow transplant. Others might oppose cloning based on their religious and moral convictions. In response to these latter considerations, President Bill Clinton urged Congress to ban the cloning of human beings in the United States in 1997 for at least five years.

Figure 1 Dolly is the first cloned sheep produced by the transfer of a nucleus from the cell of an adult sheep. (AP/Wide World Photos.)

among other reasons, they have double-membrane walls that have not proven amenable to delivery of recombinant DNA. Recently, transfection has been successful in both mitochondria and chloroplasts using a **biolistic** (biological ballistic) process, literally shooting recombinant DNA coated on tungsten microprojectiles into these organelles.

Knockout Mice

Normally, a gene used to transfect mice is incorporated randomly in the mouse genome. However, in about one in one thousand experiments, the gene replaces the normal gene by a process similar to meiotic recombination (*homologous recombination;* see chapter 12). With this process in mind, geneticists have been able to select for homologous recombination; by transfecting with defective genes, they have created mice without working copies of a particular gene. The mice produced are called **knockout mice,** and they give geneticists the opportunity to study the phenotype of an animal that lacks a particular gene.

The geneticist first creates a vector with the modified gene in question. In addition, flanking regions to that gene are added so that homologous recombination can occur. Finally, two antibiotic genes are introduced so that selection for successful transfection takes place. Within the flanking regions, the gene for neomycin resistance (neo^r) is inserted; its product inactivates the antibiotic neomycin (fig. 13. 22). Outside of the flanking regions,

the gene for thymidine kinase *(tk)* is inserted. This gene phosphorylates the drug *gancyclovir;* the phosphorylated gancyclovir is a nucleotide analogue that is incorporated during DNA synthesis, killing the cell. Thus, the combination of the *tk* gene and gancyclovir is lethal; without the *tk* gene, gancyclovir is harmless. If cells are exposed to both drugs, neomycin and gancyclovir, normal cells will be killed by neomycin, cells with the *tk* gene will be killed by gancyclovir, and only the cells with the neo^r gene but lacking the *tk* gene will survive. These alternative outcomes allow us to select the cells in which homologous recombination took place (fig. 13.22).

Cells that did not incorporate the vector will die from the effects of neomycin (they are neomycin sensitive). Cells that randomly took up the vector DNA by nonhomologous recombination will contain the neo^r and *tk* genes and will be killed by gancyclovir. However, cells that underwent homologous recombination will contain the neo^r gene but lack the *tk* gene; these cells will therefore survive in the presence of both antibiotics (fig. 13.22). Geneticists can isolate embryonic stem cells from mice, cells that can produce any mouse tissue. The cells are transfected and then grown in tissue culture in the presence of neomycin and gancyclovir, and only cells that undergo homologous recombination will survive. These cells are then injected into early-stage mouse embryos to become part of the developing mouse. The mice that develop will be chimeric; some will have incorporated the transfected cells into the germ line and become heterozygotes for the disabled gene. Finally, when mice like this are mated, one fourth of their offspring will be homozygous for the disabled gene. Thus, knockout mice, have been created through this ingenious technique.

Knockout mice are especially useful for studying development and immunology. For example, if the gene for the Mullerian-inhibiting substance is knocked out, males are infertile because they develop female reproductive organs. This experiment led to insight into the genetic path for sex determination (see chapter 5). Hundreds of knockout experiments are published each year.

Reporter Systems

We conclude this section by discussing two **reporter systems,** systems used to indicate that a transfection experiment was successful. Plants can be transfected with

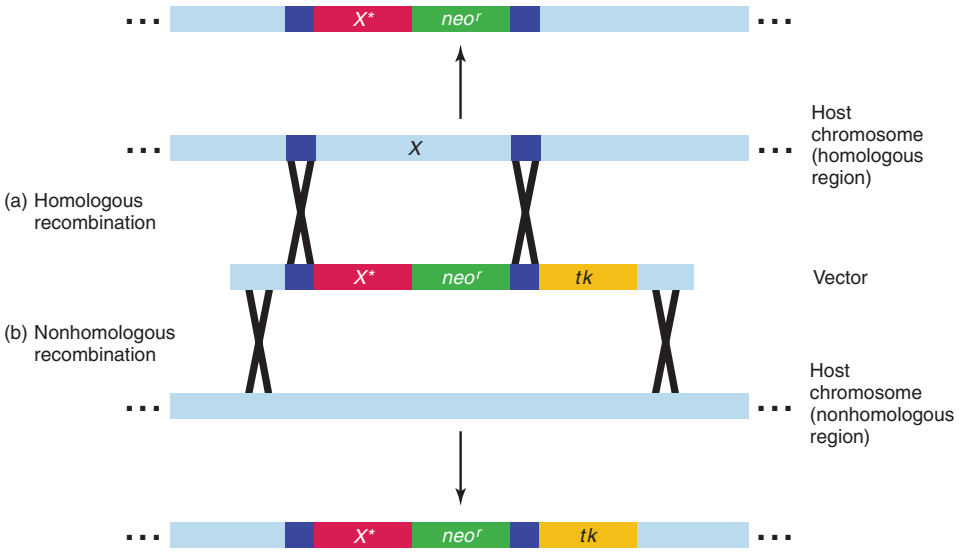


Figure 13.22 Creating a knockout mouse. A vector is created that has a disabled (nonfunctional) form of the gene in question *(X^*; red)*. Next to the gene in question is the neomycin resistance gene *(neo^r; green)*; both genes are surrounded by regions *(blue)* that flank the normal gene on its chromosome. Finally, outside the flanking regions in the vector is the gene for thymidine kinase *(tk; yellow)*. In homologous recombination (*a*), involving crossovers in the homologous flanking regions, the disabled gene and the neomycin resistance gene replace the normal gene on the cell's chromosome. In nonhomologous recombination (*b*), almost the entire vector is incorporated into the host chromosome, including the thymidine kinase gene. Techniques then allow for the selective growth of cells with the rare homologous recombination event. One fourth of the offspring of heterozygous chimeric mice will be knockouts for the gene in question.

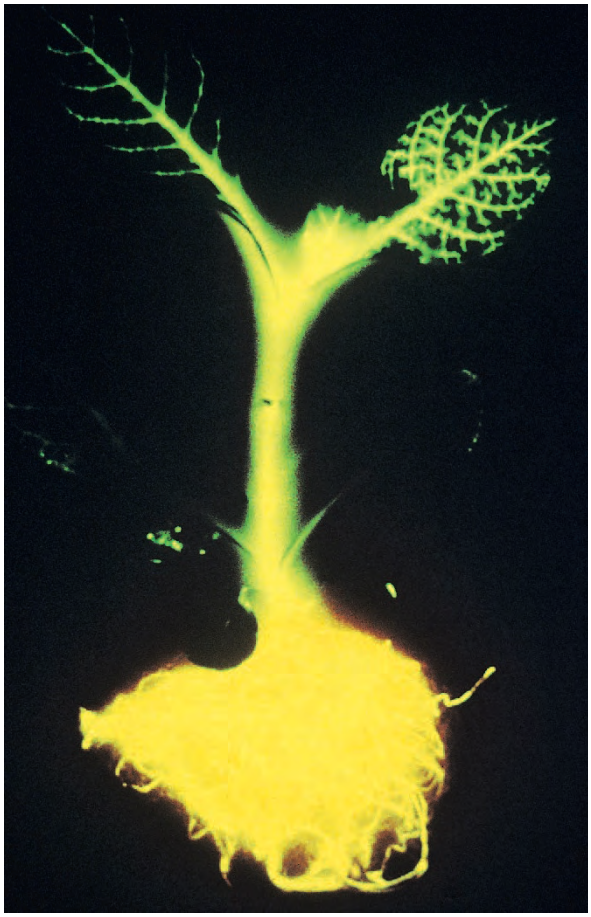


Figure 13.23 Luminescent transgenic tobacco plant containing the firefly luciferase gene. The plant was watered with luciferin, resulting in a firefly glow. (© Science VU/Keith V. Wood/Visuals Unlimited.)

the Ti plasmids of *Agrobacterium tumefaciens,* as alluded to earlier. When a plant is infected with *A. tumefaciens* containing the Ti plasmid, a crown gall tumor is induced when the Ti plasmid transfects the host plant, transferring the T-DNA region. Those cells transfected with the T-DNA are induced to grow as well as to produce opines that the bacteria feed on. Much recent research has concentrated on engineering Ti plasmids to contain other genes that are also transferred to the host plants during infection, creating transgenic plants. One series of experiments has been especially fascinating.

Tobacco plants have been transfected by Ti plasmids containing the luciferase gene from fireflies. The product of this gene catalyzes the ATP-dependent oxidation of luciferin, which emits light. When a transfected plant is watered with luciferin, it glows like a firefly (fig. 13.23). The value of these experiments is not the production of glowing plants, but rather the use of the glow to "report" the action of specific genes. In further experiments, the promoters and enhancers of certain genes were attached to the luciferase gene. As a result, luciferase would only be produced when these promoters were activated; thus, the glowing areas of the plant show where the transfected gene is active.

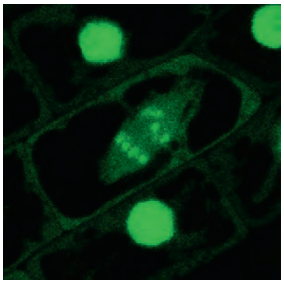


Figure 13.24 Expression of green fluorescent protein in root cells of the plant *Arabidopsis thaliana* under fluorescent light. Chromosomes are visible in a cell undergoing mitosis and chromatin is visible as circles of green in interphase nuclei. The green fluorescent protein gene was fused to the carboxy terminus of the gene for the transcription factor Cry2, controlling genes for the phototrophic response (bending toward light). (Copyright Sean Cutler, Stanford University Plant Biology Department.)

One of the more recent reporter systems developed uses a gene from jellyfish that produces a **green fluorescent protein.** The value of this system is that it "reports" when ultraviolet light shines on it, rather than requiring an addition, as in the luciferase system. The gene for the green fluorescent protein is recombined with a gene in question, and then the transfection is performed. If the gene in question has transferred successfully, carrying the gene for the green fluorescent protein, the fluorescent protein will report it when activated by ultraviolet light (fig. 13.24).

RESTRICTION MAPPING

The number of cuts that a restriction enzyme makes in a segment of double-stranded DNA depends on the size of that DNA, its sequence, and the number of base pairs in the recognition sequence of the particular enzyme. That is, a restriction enzyme with only three base pairs in its recognition sequence will cut more times than one with six base pairs in its sequence, since the probability of a sequence occurring by chance is a function of the length of that sequence. A sequence of three bases occurs more often by chance ($1/4^3 = 1/64$ base pairs) than a sequence of six bases ($1/4^6 = 1/4{,}096$ base pairs). *Hin*dII,

for example, cuts the circular DNA of the tumor virus SV40 into eleven pieces; some restriction enzymes can cut *E. coli* DNA into hundreds of pieces. The product of the action of a restriction enzyme on a DNA sample is called a **restriction digest.**

Using electrophoresis, we can separate the fragments of a restriction digest by size. With techniques to be described later, we can locate the restriction sites on the original gene or piece of DNA. That is, we can construct a map of the restriction recognition sites that will give us the physical distance between sites, in base pairs (fig. 13.25). This **restriction map** is extremely valuable for several reasons. For example, when the radioactive nucleotide tritiated thymidine was added for a very short period of time during the beginning of DNA replication in SV40 viruses, the radioactivity always appeared in only one restriction fragment. This demonstrated that SV40 replication started from a single, unique point; that point was localized to a particular segment of the SV40 chromosome.

In addition, a restriction map often allows researchers to correlate the genetic map and the physical map of a chromosome. Certain physical changes in the DNA, such as deletions, insertions, or nucleotide changes at restriction sites, can be localized on the genetic map. These changes can be seen as changes in size, or in the total absence, of certain restriction fragments when compared with wild-type DNA. This information allows us to see changes in the DNA; it also gives us information about the evolution of species (see chapter 21). The differences in fragment sizes are called **restriction fragment length polymorphisms (RFLPs)** and have proven valuable in pinpointing the exact location of genes and determining the identity or relatedness of individuals. A restriction digest is also useful for isolating short segments of DNA that can be easily sequenced.

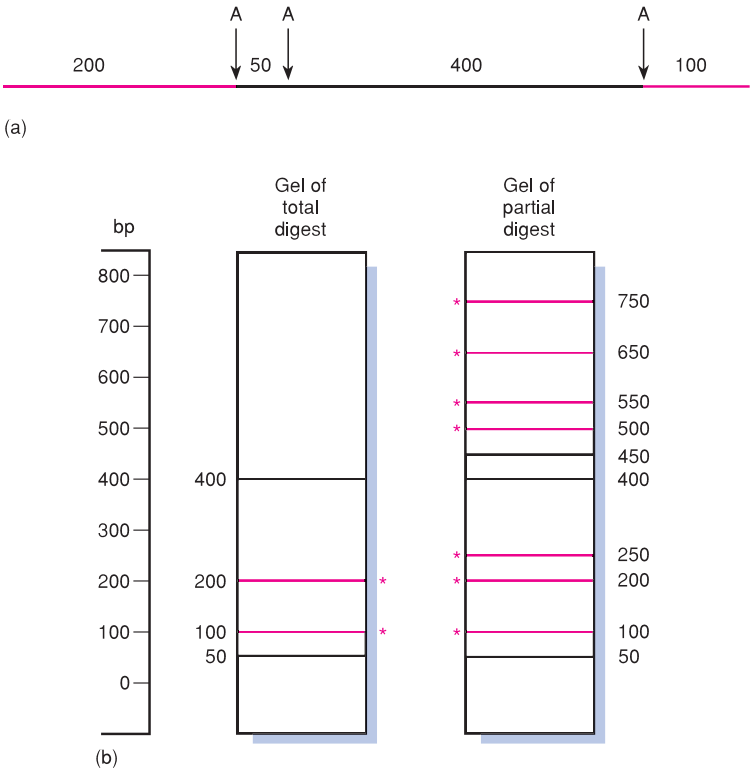


Figure 13.25 Restriction map from electrophoresis of a restriction endonuclease digest. (*a*) Original piece of DNA, showing restriction sites marked by *A*. (*b*) Agarose gels showing bands of total and partial restriction digests. *Asterisks* mark radioactive bands produced by end-labeled segments. At the *left* is the scale of molecular weight markers in base pairs (e.g., 800 bp, 700 bp). The total digest produces fragments that are 400, 200, 100, and 50 bp—and the 200 and 100 bp fragments are end labeled. The partial digest yields six additional bands.

Constructing a Restriction Map

How do we construct a restriction map? Figure 13.25 shows a hypothetical piece of DNA cut by restriction enzyme *A*. Below this map is a diagram of the electrophoresed digest on agarose gels, which are usually used because their porosity allows DNA fragments of relatively large size to move. The restriction enzyme makes three cuts in the DNA, generating four fragments that are 200, 50, 400, and 100 base-pairs long. The banding pattern on the gel at the *left* in figure 13.25*b* is the result of the electrophoresing of that digest. (Note that smaller segments move faster than larger segments.) The sizes of the segments are determined by comparison with standards of known size (not shown, although the scale is indicated on the left). The gel does not reveal the order of these segments on the chromosome. Several methods can be used to determine the exact order of the restriction segments on the original piece of DNA.

Before restriction enzyme digestion, the 5′ ends of the DNA can be labeled radioactively with ^{32}P using the enzyme polynucleotide kinase. Since the enzyme is acting on double-stranded DNA, both ends will be labeled. Upon electrophoresis after digestion of the DNA in figure 13.25, the 200-base-pair and 100-base-pair (bp) bands will be labeled radioactively, indicating that these segments are the termini of that piece of DNA. However, we still don't know the order of the middle pieces.

The order of the other segments can be determined by slowing down the digestion process to produce a **partial digest.** If the reaction is cooled or allowed to proceed for only a short time, not all restriction sites will be cut. Some pieces of DNA will not be cut at all, some will be cut once, some twice, and some cut at all three restriction sites. The result of electrophoresis of this partial digest is seen at the right in figure 13.25*b*. From this gel, we can reconstruct the segment order. This gel contains the four original segments plus six new segments, each containing at least one uncut restriction site.

From the total digest gel, we know that the 200 and 100 bp segments are on the outside because they were labeled radioactively. This means the 50 and 400 bp segments are on the inside. In the partial digest, we find a 250 bp segment but not a 150 bp segment, which tells us that the 50 bp segment lies just inside and next to the 200 bp terminus (fig. 13.26*b*). There is a 500 bp segment but not a 600 bp segment, which tells us that the 400 bp segment lies adjacent to the 100 bp terminus (fig. 13.26*c*). An unlabeled 450 bp segment confirms that the 400 and 50 bp segments are adjacent and internal in the DNA. We thus unequivocally reconstruct the original DNA (compare fig. 13.26*e* with fig. 13.25*a*), creating a map of sites of restriction enzyme recognition regions separated by known lengths of DNA.

Double Digests

In practice, restriction mapping is usually done with several different restriction enzymes. Figure 13.27 is a map of the DNA of figure 13.25, with the recognition sites of a second endonuclease, *B,* included. Using the same methodology just outlined, we can show that the order of the *B* segments is 350, 250, and 150 base pairs arising from two cuts by endonuclease *B.* What we do not know is how to overlay the two maps. Do the *B* segments run left to right or right to left with respect to the *A* segments (fig. 13.27*a* and *b*)? We can determine the unequivocal order by digesting a sample of the original DNA with both enzymes simultaneously, thus producing a **double digest.**

The two orders shown in figure 13.27*a* and *b* are used to make different predictions about the double digest. From the first order (*a*), we predict a 200 bp end segment, radioactively labeled. From the second order (*b*), we predict that the labeled 200 bp segment will be cut back to 150 base pairs: there should not be a labeled 200 bp segment. The double digest shows a labeled 200 bp segment, indicating order (*a*). All other aspects of order (*a*) are consistent with the double digest.

Restriction mapping thus provides us with a physical map of a piece of DNA, showing restriction endonuclease sites separated by known lengths of DNA. This technique

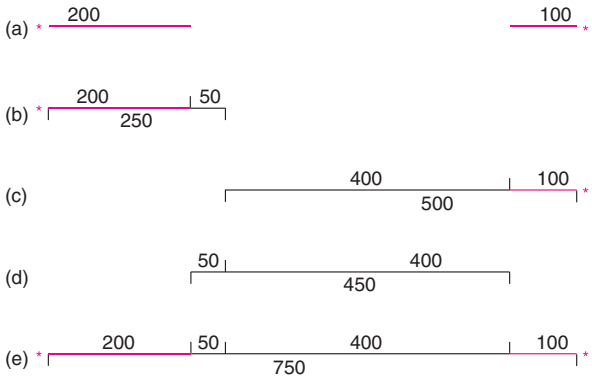


Figure 13.26 Steps in the reconstruction of the DNA from figure 13.25. *Asterisks* show ^{32}P end labels. From the total digest, the 100 and 200 bp segments are established as the end segments (*a*). Since there are also 50 and 400 bp fragments within the DNA (established from the total digest), only certain bands (fragments) are possible from the partial digest, which establishes that the 50 bp fragment is adjacent to the 200 bp fragment and the 400 bp fragment is adjacent to the 100 bp end segment (steps *b* and *c*). The occurrence of an unlabeled 450 bp fragment in the partial digest verifies the existence of the 50 and 400 bp fragments (*d*), yielding the final structure (*e*). All the fragments in the partial digest are consistent with this arrangement.

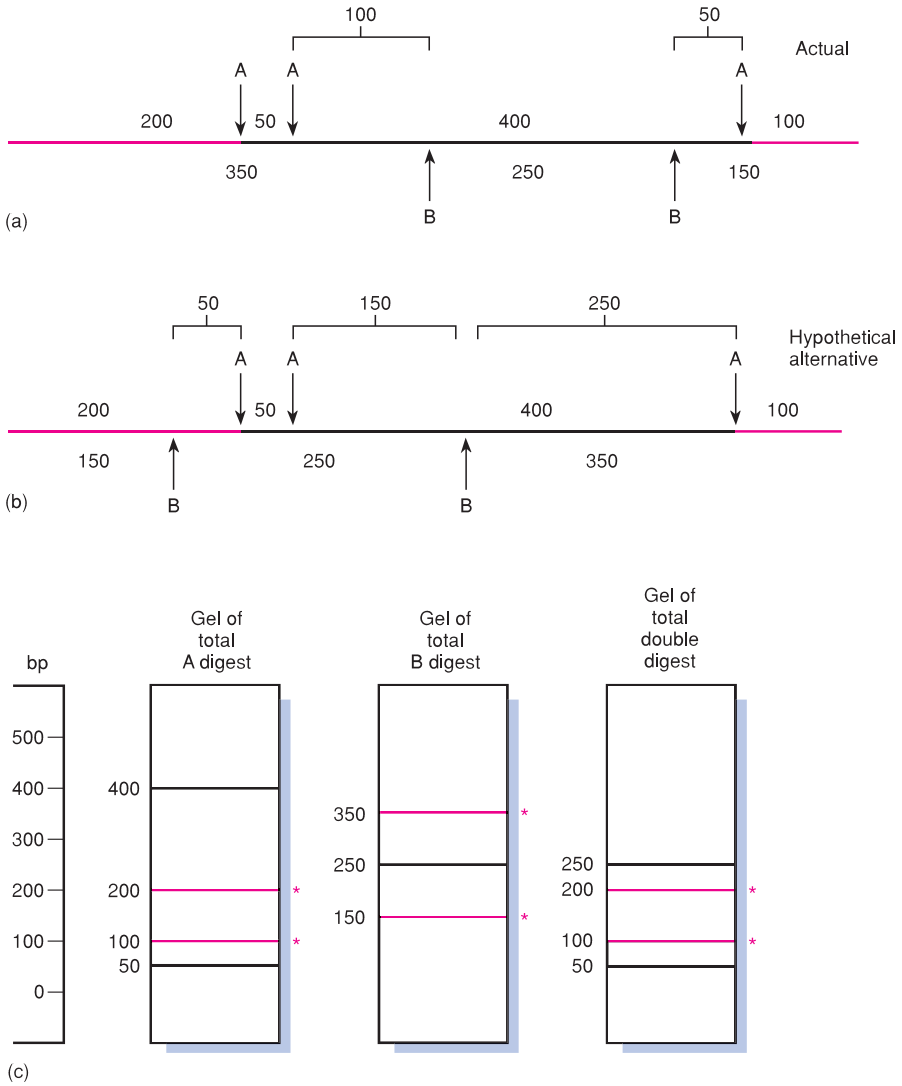


Figure 13.27 Overlay of the recognition sites affected by two different restriction endonucleases (*A* and *B*) on the same piece of DNA. (*a*) Actual arrangement. (*b*) Hypothetical alternative arrangement. (*c*) Electrophoresis of the total restriction digests by A alone, B alone, and both. *Asterisks* indicate radioactive end-labeled bands. Order (*a*) is consistent with all the bands found in all the digests, whereas order (*b*) is not. For example, in order (*b*) an internal (unlabeled) 150 bp fragment is predicted, but this fragment is not found in the total digest.

gives us short DNA segments of known position that we can sequence, as well as a physical map of the DNA that can be compared with the genetic map and can locate mutations and other particular markers.

Restriction Fragment Length Polymorphisms

Restriction fragment length polymorphisms (RFLPs), obtainable from restriction digests, are proving to be very valuable genetic markers in two areas of study: human gene mapping and forensics. In a restriction digest of the whole human genome, there might be thousands of fragments from a single restriction enzyme. Unique probes have been developed for Southern blotting these digests. Genetic variation usually comes in the form of a second allele that, due to a mutation, lacks a restriction site and is therefore part of a larger piece of DNA (fig. 13.28). Some probes have uncovered **hypervariable loci** with many alleles (any one

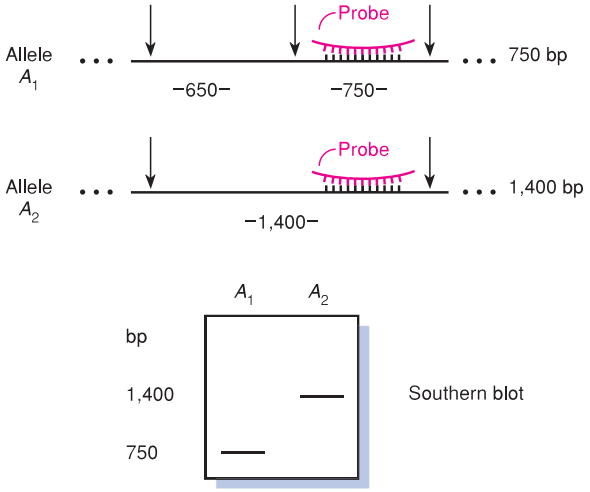


Figure 13.28 Restriction fragment length polymorphism (RFLP) analysis. Allele A_1 is a gene segment, 750 bp long, identified by probe binding. In allele A_2, the restriction site to the *left* of allele A_1 has been changed. The probe thus recognizes a 1,400 bp fragment instead of the 750 bp fragment. During Southern blotting, two different bands show up from the two alleles. *Arrows* indicate restriction sites.

person has, of course, only two of the many possible alleles). A population's genetic variation is generated because these hypervariable loci contain many tandem repeats of short (10 to 60 bp) segments. Due presumably to unequal crossing over (see chapter 8), just one of these loci, called **variable-number-of-tandem-repeats (VNTR) loci,** can generate much variation. As a result, probing for one of these VNTR loci in a population reveals many alleles.

The Southern blots of such digests create a **DNA fingerprint** of extreme value in forensics. DNA extracted from blood or semen samples left by a criminal can be compared with DNA patterns of suspects (fig. 13.29). When a single probe recognizes a number of different loci, each individual will have many bands on a Southern blot, with most people producing unique patterns. In one system, developed by A. Jeffreys, a single probe locates fifty or more variable bands per person. If Jeffreys's probes are

Alec Jeffreys (1950–). (Courtesy of Dr. Alec Jeffreys.)

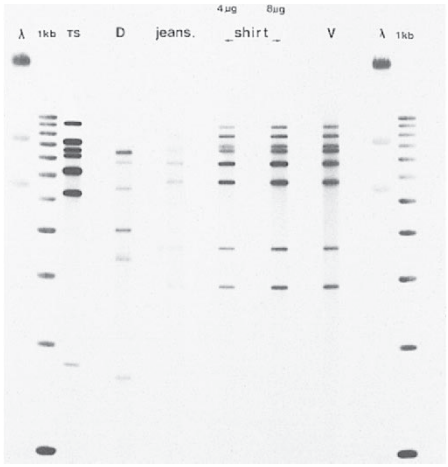


Figure 13.29 Forensic use of DNA fingerprinting. Southern blot of DNA from victim (*V*) and defendant (*D*) in a crime. *Jeans* and *shirt* refer to blood samples taken from the clothing of the defendant. The pattern on the shirt clearly matches the victim's blood, not the defendant's own blood. All of the other lanes of the blot contain controls and size standards. The probability that the blood stains were not from the victim was estimated at one in thirty-three billion, more than the number of people on earth. However, these probabilities are controversial, depending on statistical assumptions about variability within racial and ethnic subpopulations. (Courtesy of Cellmark Diagnostics, Germantown, MD.)

used to compare the patterns, the likelihood that the two patterns would match randomly is infinitesimally small. This technique thus has greater power to identify individuals than using the prints from their fingertips.

POLYMERASE CHAIN REACTION

In the past, in many instances (in museum specimens, dried specimens, crime scene evidence, and fossils), a DNA sample was available, but in such small quantity or so old as to be considered useless for study. That situation changed in 1983 when Kary Mullis, a biochemist working for the Cetus Corporation, devised the technique we now refer to as the **polymerase chain reaction (PCR).** PCR can be used to amplify whatever DNA is present, however small in quantity or poor in quality. The only requirement is that the sequence of nucleotides on either side of the sequence of interest be known. That information is needed to construct primers on either side of the sequence of interest. Once that is done, the sequence between the primers can be amplified.

In the PCR technique, the primers and the ingredients for DNA replication are added to the sample. Then, the mixture is heated (e.g., 95° C for twenty seconds) to denature the DNA. The temperature is then lowered (e.g., 55° C for twenty seconds) so that primers can anneal to their complementary sequences. The temperature is then raised again (e.g., 72° C for twenty seconds) for DNA replication. Then, a new cycle of replication is initiated (fig. 13.30). The various stages in the cycle are controlled by changes in temperature since the temperatures for denaturation, primer annealing, and DNA replication are different. About twenty cycles of PCR produces a million copies of DNA; thirty cycles make a billion copies. The technique is aided by using DNA polymerase from a hot-springs bacterium, *Thermus aquaticus,* that can withstand the denaturing temperatures. Thus, after each cycle of replication, no new components have to be added to the reaction mixture. Rather, the cycling can be continued without interruption in PCR machines (simply programmable water baths that accurately and rapidly change the water temperature that surrounds the reaction mixture). Some machines can process ninety-six samples at a time.

PCR has been used to create DNA fingerprints by amplifying **microsatellite DNA.** These are repeats of very short sequences of DNA dispersed throughout the genome. For example, cytosine-adenine (CA) repeats occur tens of thousands of times in eukaryotes, in repeats of from twenty to sixty base pairs. As in the case of VNTR loci, there is tremendous variability among people in the number of these repeats at a locus, due presumably to crossover errors. Unlike the situation with VNTR loci, however, PCR amplification of one of these loci can be done without restriction cutting, Southern blotting, and probing—PCR gives the results directly upon electrophoresis. All we need are the surrounding primer sequences to any microsatellite locus. PCR is now a routinely used tool in the laboratories of molecular geneticists. They use it to rapidly amplify the DNA regions of interest for research or forensic uses.

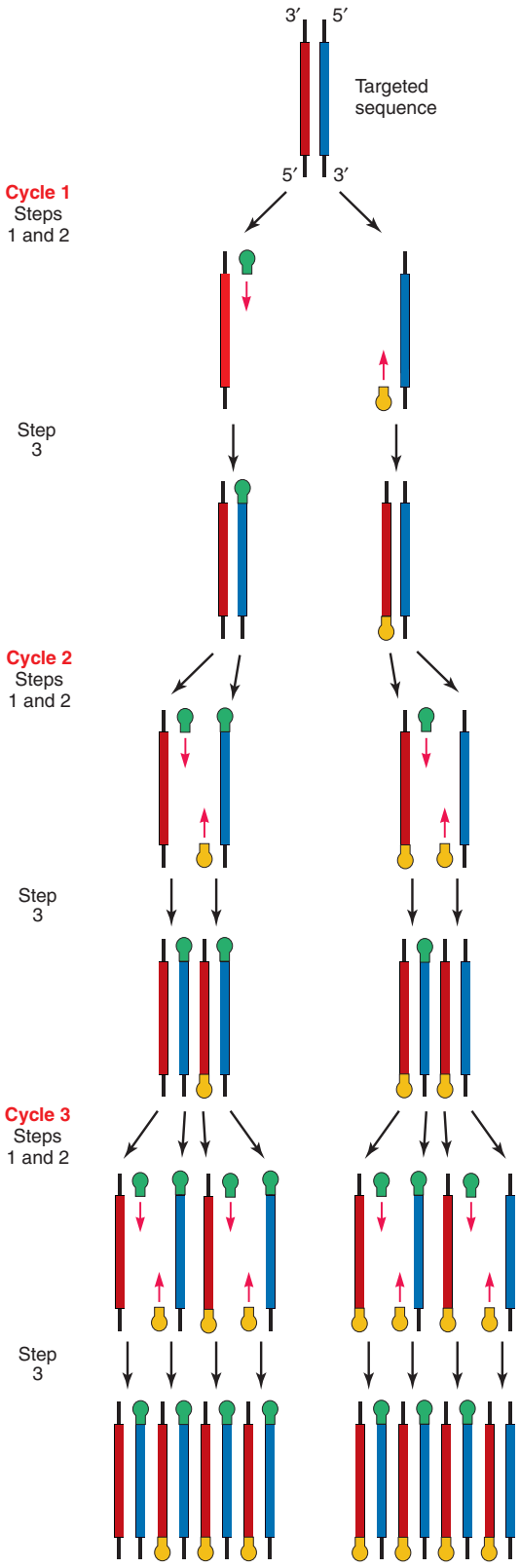


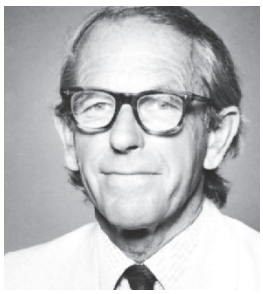
Figure 13.30 Polymerase chain reaction. DNA is denatured, (step 1), primer oligonucleotides that are complementary to end sequences on the two strands anneal (step 2), and DNA replication takes place (step 3). Each step in the cycle is controlled by temperature changes. The targeted sequence is shown as *red* on one stand and *blue* on the other. Primers are shown as either *green* or *yellow* lollipops. A green primer begins the copying of the red strand into a complementary blue strand; a yellow primer begins the copying of a blue strand into a complementary red stand. In three cycles, one double-stranded region of DNA becomes eight. The process requires the addition of primers, deoxynucleotide triphosphates, and DNA polymerase, as well as changing temperature cycles.

We now turn our attention to a major result of recombinant DNA technology, DNA sequencing. Recombinant DNA technology, with its ability to isolate and amplify small, well-defined regions of chromosomes, has allowed the development of DNA sequencing techniques.

DNA SEQUENCING

Paul Berg of Stanford University, Walter Gilbert of Harvard University, and Frederick Sanger of the Medical Research Council in Cambridge, England, shared the 1980 Nobel Prize in chemistry. Berg won for creating the first cloned DNA molecules when he spliced the SV40 genome into phage λ. Gilbert and Sanger were awarded the prize for independently developing methods of sequencing DNA. Gilbert, along with Allan Maxam, developed a method of DNA sequencing called the *chemical method.* It involves chemically breaking down the DNA at specific bases. Sanger, who won a Nobel Prize in 1959 for sequencing the insulin protein, later took part in developing methods for sequencing RNA. His sequencing method, developed with Alan Coulson, involved DNA synthesis and was called the *plus-and-minus method.* The further development of the method by Sanger, Coulson, and S. Nicklen, using specific chain-terminating nucleotides, led to a modification of the plus-and-minus method known as the *dideoxy method.*

Walter Gilbert (1932–). (Photo: Rick Stafford.)

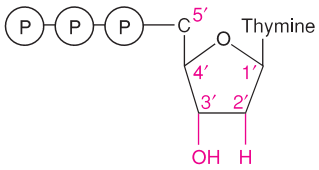


Frederick Sanger (1918–). (Courtesy of Dr. Frederick Sanger.)

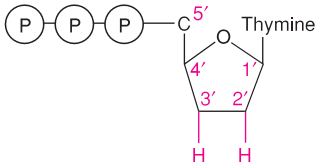
The Dideoxy Method

In the **dideoxy method,** manipulation of DNA synthesis enables DNA sequencing. Remember from chapter 9 that DNA synthesis occurs at a primer configuration, one in which double-stranded DNA ends with a 3′-OH group on one strand. The other strand continues as single-stranded DNA (fig. 13.31, *middle*). The dideoxy method creates a primer configuration of the DNA to be sequenced and enables replication to proceed. A trick, using chain-terminating nucleotides, stops DNA synthesis at known positions. These chain-terminating nucleotides are formed of sugars lacking OH groups at both the 2′ and 3′ carbons (hence the term *dideoxy*). Without a 3′-OH group, a dideoxynucleotide cannot be used for further DNA polymerization (fig. 13.31).

Chain-terminating nucleotides permit synthesis to be stopped at a known base. The sample to be sequenced is elongated separately in four different reaction mixtures, each having all four normal nucleotides but also having a proportion of one of the chain-terminating dideoxy nucleotides. For example, if the pool of thymine-containing triphosphate nucleotides contains a portion of the dideoxythymidine triphosphate molecules, then synthesis of the growing strand is sometimes terminated when adenine (the complement of thymine) appears on the template, creating fragments that end in thymine. Similar



Deoxythymidine triphosphate (dTTP)



Dideoxythymidine triphosphate (ddTTP)

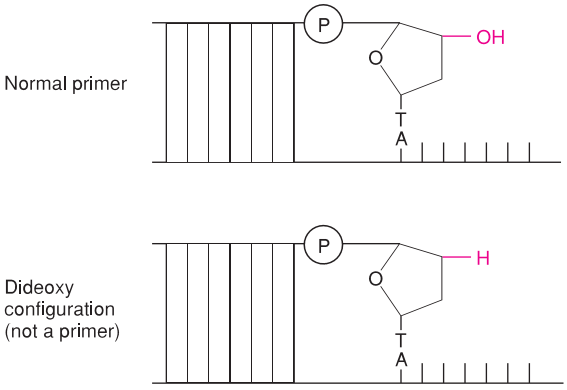


Figure 13.31 Dideoxy nucleotides cause chain termination during DNA replication. The dideoxy primer configuration lacks the 3′-OH group needed for chain lengthening in a normal primer configuration.

reactions are carried out in separate test tubes for each of the other nucleotides, producing fragments that terminate when the respective complementary nucleotide is present. The resulting fragments from each reaction are electrophoresed, generating a pattern on the gel that reveals the sequence of the newly synthesized DNA. Let us go through an example.

In figure 13.32*a,* we show the DNA to be sequenced, a small segment of nine base pairs. To sequence this segment, one must get one strand of this double-stranded segment into the configuration shown in figure 13.32*b*. The DNA to be sequenced must be the template for new DNA synthesis. (We will soon discuss how we obtain the required configuration.) Having created the necessary primer configuration, we take four subsamples of it, each including all four nucleoside triphosphates plus DNA polymerase I. At least one of the nucleoside triphosphates is radioactively labeled, usually with ^{32}P. This label allows us to identify newly synthesized DNA by autoradiography.

To each of the four subsamples, one of the dideoxynucleotides (dd) is added—one subsample gets ddTTP, one gets ddATP, one gets ddCTP, and one gets ddGTP. These dideoxynucleotides are added in addition to the regular deoxynucleotides to increase the probabil-

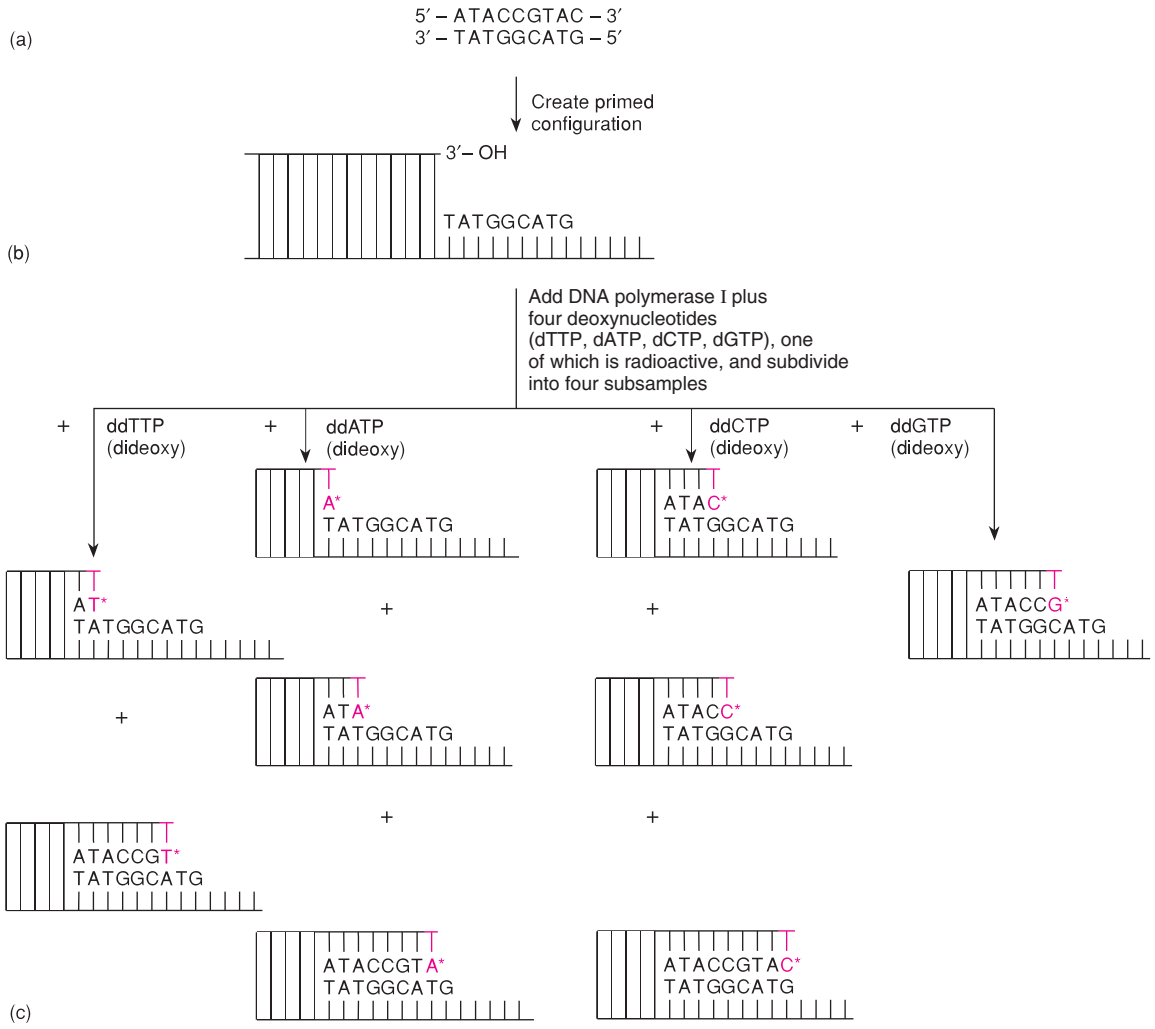


Figure 13.32 Initial steps in the dideoxy method of DNA sequencing. The *asterisks* indicate the dideoxynucleotides. The DNA to be sequenced is placed into a primer configuration (*a*, *b*). Four reaction mixtures are created, each with all four normal nucleotides plus one of the dideoxynucleotides. Thus DNA synthesis in each reaction mixture is stopped a percentage of the time when the complement to the dideoxynucleotide appears in the template (*c*).

ity that chain termination will occur at every appropriate position. If the dideoxynucleotide were added in place of the deoxynucleotide, then the chain would be terminated the first time the complement of that base appeared in the template strand. By mixing the dideoxynucleotides and the deoxynucleotides, we are assured that termination will occur in every appropriate position.

In figure 13.32*c,* we see that the template has two adenines. Therefore, in the ddTTP reaction mixture, adenine's complement (thymine) is needed twice. There are thus two possible points for ddTTP to incorporate, two possible chain terminations, and therefore two fragments that could end in dideoxythymidine, of two and seven bases, respectively. Similarly, there are three possible fragments ending in adenine, of one, three, and eight bases; three ending in cytosine, of four, five, and nine bases; and one ending in guanine, of six bases (fig. 13.32*c* and fig. 13.33, *top*).

After DNA synthesis is completed, the old primer is removed, leaving only newly synthesized DNA fragments (fig. 13.33). Newly replicated segments of various lengths

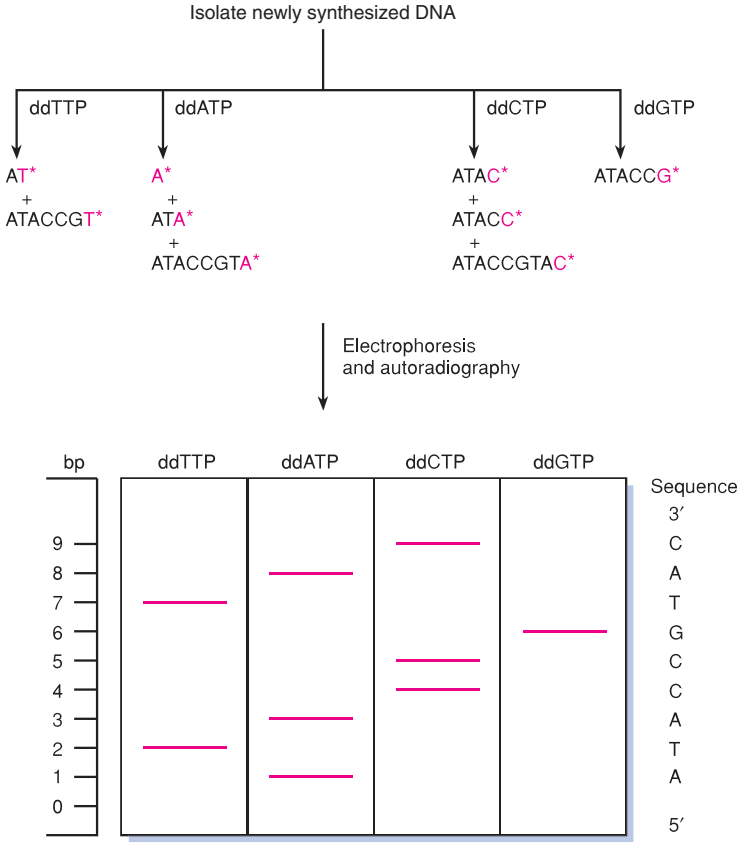


Figure 13.33 Electrophoresis of segments produced by the dideoxy method of DNA sequencing. This method allows direct reading of the sequence. The *asterisks* indicate the dideoxynucleotides. The newly synthesized reaction products seen in figure 13.32 are isolated by removal of the primer and template. Each reaction mixture (e.g., ddTTP is the mixture containing dideoxythymidine triphosphates) produces specific products of specific lengths that can be determined by electrophoresis. In the case of the ddTTP mixture, two fragments ending in thymine are possible; one is two bases long, the other seven bases long. Thus, the complement of thymine, adenine, appears in positions 2 and 7 of the original piece of DNA. However, either the original strand or its complement (the new synthesis) yields the original sequence since DNA is a double helix; the sequence in one strand is always defined by the complementary sequence in the other strand.

from each reaction mixture are placed in separate slots and then electrophoresed on polyacrylamide gels to determine the lengths of the segments. Since only newly synthesized DNA segments are radioactive, autoradiography lets us keep track of newly synthesized DNA. As you can see from the autoradiograph of the gel in figure 13.33, each subsample produces segments that begin at the primer configuration (beginning of synthesis) and end with the chain-terminating dideoxy base. By starting at the bottom and reading up, back and forth across the gel, we can directly determine the exact sequence of the DNA segment. Because they have the appearance of stepladders in each lane (fig. 13.34), the gels are usually referred to as **stepladder gels** or **ladder gels.**

This technique (in the form of the original plus-and-minus method) was first used to sequence the genome of the DNA phage ϕX174 (box 13.3). That phage was used because it lent itself to the sequencing method. It has single-stranded DNA within the phage coat, yet its DNA becomes double-stranded once it enters the bacterium. Creating a primer configuration was thus relatively easy. The double-stranded circle from within the host could be treated with a restriction endonuclease to produce double-stranded fragments (fig. 13.35). These fragments could then be denatured. From this mixture, a particular fragment could be isolated by electrophoresis. The isolated strand would reanneal to the single-stranded DNA taken from the phage heads, forming a primer for new growth. The same restriction endonuclease would free the new growth after it had taken place. Thus, the dideoxy method was relatively easy to apply to the 5,387-base chromosome of ϕX174.

Creating a General-Purpose Primer

To make the dideoxy method efficient, researchers created a general primer for routine sequencing work by recombinant DNA engineering of an *E. coli* vector, the single-stranded DNA phage M13. This phage is similar to ϕX174 in that both are packaged as single-stranded DNA, and both are replicated to double helices within the host. Therefore, the double-stranded form within the host, called the replicating form, can be engineered by standard methods, and the single-stranded form can be used for sequencing. The system works as follows.

By very clever engineering, J. Messing and his colleagues created cloning sites for a variety of restriction enzymes in a bacterial gene (*lacZ*) that had been inserted into M13 (fig. 13.36). The gene is for the β-galactosidase enzyme that normally breaks down lactose. It also breaks down an artificial substrate of the enzyme, X-gal, which is normally colorless. When cleaved by β-galactosidase, X-gal becomes blue. Thus, in the presence of the functional *lacZ* gene, M13 plaques are blue. If the gene is disrupted by a cloned insert, X-gal does not break down,

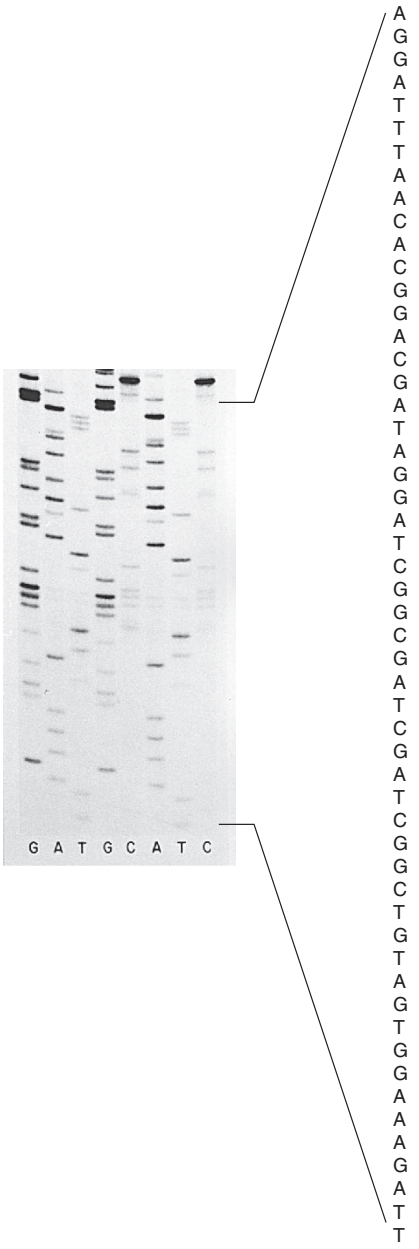


Figure 13.34 Autoradiograph of a dideoxy sequencing gel. The letters *G, A, T,* and *C* along the *bottom* refer to the ddGTP, ddATP, ddTTP, and ddCTP reaction mixtures, respectively. Lanes are repeated for easier identification of the bands. The sequencing is also verified by sequencing the complementary strand and checking for agreement. (Courtesy of Richard J. Roberts.)

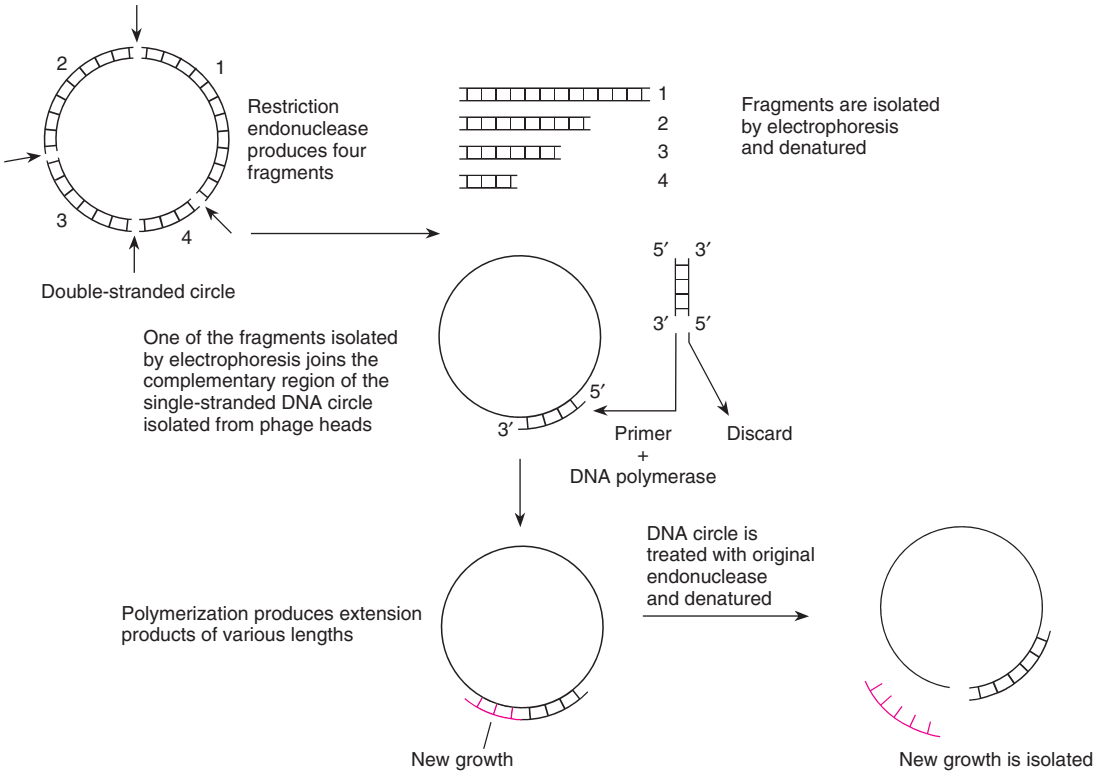


Figure 13.35 The genome of phage ϕX174 lent itself to the dideoxy method (originally, the plus-and-minus method) of DNA sequencing. Because the phage occurs in both the single- and double-stranded forms, it can be manipulated for sequencing. The double-stranded form is fragmented with an endonuclease. One fragment is isolated by electrophoresis and hybridized to the single-stranded form, creating a primer for new DNA synthesis and thus for dideoxy sequencing. Newly synthesized DNA can be isolated by treating it with the same restriction enzyme, which will create the same cut made originally. The newly isolated pieces can then be electrophoresed as in figure 13.34.

and hence the plaques are colorless. (M13 doesn't form true plaques because it doesn't lyse the *E. coli* cells. It does form turbid sites due to reduced bacterial growth.)

An oligonucleotide primer can be synthesized that is complementary to a region of the phage DNA upstream from the cloning sites. Single-stranded phage DNA containing a cloned insert is isolated and hybridized with the synthetic oligonucleotide. This operation creates the primer configuration for dideoxy sequencing of the cloned DNA. Virtually any clonable segment of DNA can be sequenced using this very general method. Theoretically, that segment could be any size.

Stepladder gels, however, are effective only up to about four hundred base pairs. To sequence larger regions requires sequencing overlapping segments and reconstituting the sequence by the overlap pattern, similar to the methods we described for amino acid sequencing (chapter 11, box 11.1). Overlapping segments of DNA are usually obtained by using two or more restriction enzymes.

The most recent innovation in DNA sequencing involves using four fluorescent dyes, each fluorescing at a different wavelength (505, 512, 519, and 526 nm); each of the four dideoxy nucleotides has a different dye attached. After the newly synthesized fragments are isolated, the products from all four reactions are run together in the same lane of a polyacrylamide gel. The gel is then scanned with an argon laser that excites the dye molecules. An instrument records the color of the peaks, reading the sequence directly and automatically (fig. 13.37). This method greatly simplifies sequencing since it is automated. It also alleviates the necessity for radioactive tags.

BOX 13.3

Experimental Methods

Genes Within Genes

Complete sequencing of a DNA genome using Sanger and Coulson's plus-and-minus method (the forerunner to the dideoxy method) was first accomplished with ϕX174, a virus that contains a single-stranded DNA circle of 5,387 bases within its protein capsule. Once injected into the host, the DNA is replicated to form a double helix that then proceeds in normal viral fashion to replicate itself, manufacture its own coat proteins, lyse the cell, and escape. This virus has nine genes. The virion is a small, twenty-faced polyhedron with a small spike at each of its twelve vertices. This spike attaches ϕX174 to *E. coli.* The coat accounts for one protein and the spike accounts for two. Thus, three of the virus's nine genes manufacture coat proteins. Figure 1 illustrates the location of the genes in ϕX174, obtained through standard mapping methods.

From the information obtained from the sequencing of MS2, an RNA virus, geneticists believed that there should always be a nontranslated sequence between genes, presumably for the purpose of controlling expression of each gene. However, careful perusal of the nucleotide sequence of ϕX174 provided several surprises. First, the ends of three genes overlapped the beginnings of the next genes (*A-C, C-D,* and *D-J*); in the first two cases, the initiation codon is entirely within the end of the previous gene, but read in a different frame of reference. In the sequence ATGA, the ATG is the initiation of the next gene, whereas the TGA is the termination of the previous gene. In the *D-J* interface, one A is shared: TAATG (UAAUG in ribose nucleotides; fig. 2). It is the number 3 base of the termination codon and the number 1 base of the initiation codon. The surprises did not end there.

At first, with the sequence of nucleotides spread out in front of them, the researchers could not find the *B* and the *E* genes; they appeared to be missing. Upon careful analysis, however, the scientists found that the *B* gene was entirely within the *A* gene and the *E* gene was entirely within the *D* gene (fig. 3). Their finding went against theory. We were led to believe, from logical arguments, that genes cannot substantially overlap. There would be too much of a constraint on function: The functional sequence of one gene would also have to be a functional sequence in the other. Similarly, there would be an evolutionary constraint involved. The genes would have to evolve together. But here we have two cases in which genes do overlap. How could overlapping genes come about?

There are a large number of thymine bases in the ϕX174 genome. In the *D* gene particularly, many of the codons end with thymine. The imbedded *E* gene is read on a shifted frame with *D* so that the terminal bases of *D*'s codons are the middle bases of *E*'s. A look at the genetic code (see table 11.4) shows that the codons with U in the middle (*E*'s codons) are mainly for hydrophobic amino acids. Thus, *E* is a protein with detergent properties. In fact, it is the protein responsible for the dissolution of the outer cell wall of the host bacterium, a process that a detergent can accomplish in vitro. The properties of the *E* gene, then, are more the properties of its individual amino acids rather than their exact sequence.

In the *A-B* case, there is an indication that the two genes were once autonomous. This indication is based on the patterns of the codons; *A*'s codons tend to end in thymine before the overlap, but thereafter, in the region of overlap, *B*'s codons end in thymine, whereas *A*'s codons do not. Presumably, a mutational event tagged the *B* material onto the end of the earlier, shorter *A* gene and improved its enzymatic ability. We can only speculate, however.

The amazing arrangement of this viral DNA is one of extreme economy. The protein package is small, yet a minimum of nine genes had to be packed into it. We have seen this kind of economy before in the codon usage of mitochondrial DNA (see chapter 11).

As more sequencing has taken place, geneticists have discovered other novel overlap situations. For example, in one case, two genes were transcribed from opposite strands of the same region of DNA from a rat. On one strand, the gonadotropin-releasing hormone gene (*GnRH*) is located. On the other is a gene (*RH*) that produces a protein expressed in the heart.

Gene overlap is known to occur in bacteria as well. In *E. coli,* the promoter for the *ampC* gene (coding for the enzyme β-lactamase) begins within the last ten codons for the *frdC* gene, which codes for a subunit of the enzyme fumarate reductase. There is evidence that in this arrangement the *frdC* terminator can have some regulatory control of *ampC* transcription. (See chapter 14 for a discussion of regulatory processes in prokaryotes.)

With DNA sequence data, including the complete sequences of other chromosomes such as those of SV40 and mitochondria, we have accumulated much information about gene arrangements. Overlap to one degree or another has been found in small viruses (ϕX174, SV40), large viruses (λ), mitochondrial chromosomes, bacterial DNA, and even eukaryotes, in which several cases are now known in which genes are located within introns of other genes. In one of the few examples known, three genes are located in an intron of the neurofibromatosis gene, a gene that causes a disfiguring neurological disease. Although relatively uncommon, overlap and embedding of genes may have some regulatory role in transcription in addition to minimizing the length of the chromosome.

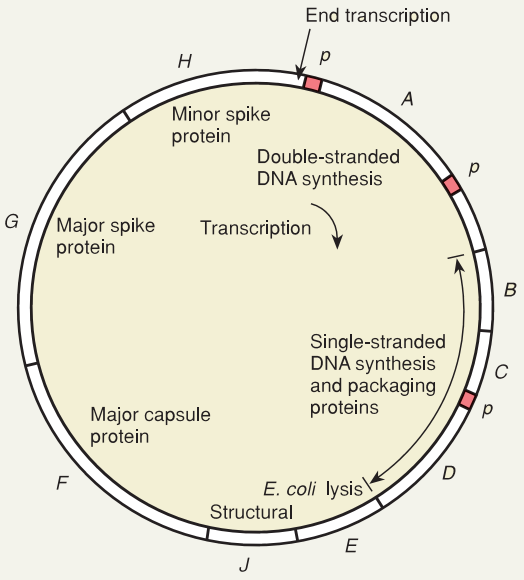


Figure 1 Presumed location of the nine genes of phage ϕX174 on its circular chromosome. Transcription begins at three different places, each marked *p*, for promoter. The function of each gene appears within the circle.

Figure 2 Sequence, shown as ribose nucleotides, where genes *E* and *D* end and gene *J* begins. Each is out of register with the other two. The A of AUG for gene *J,* for example, is the second A of the UAA terminator of gene *D*.

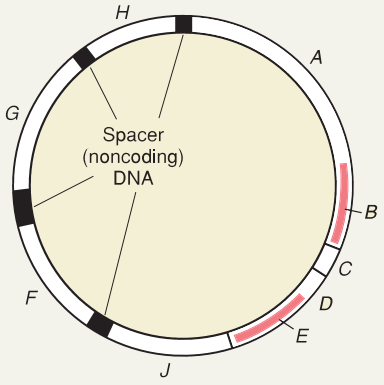


Figure 3 The actual map of the nine genes of phage ϕX174. Note that *B* is entirely within *A* and *E* is entirely within *D*.

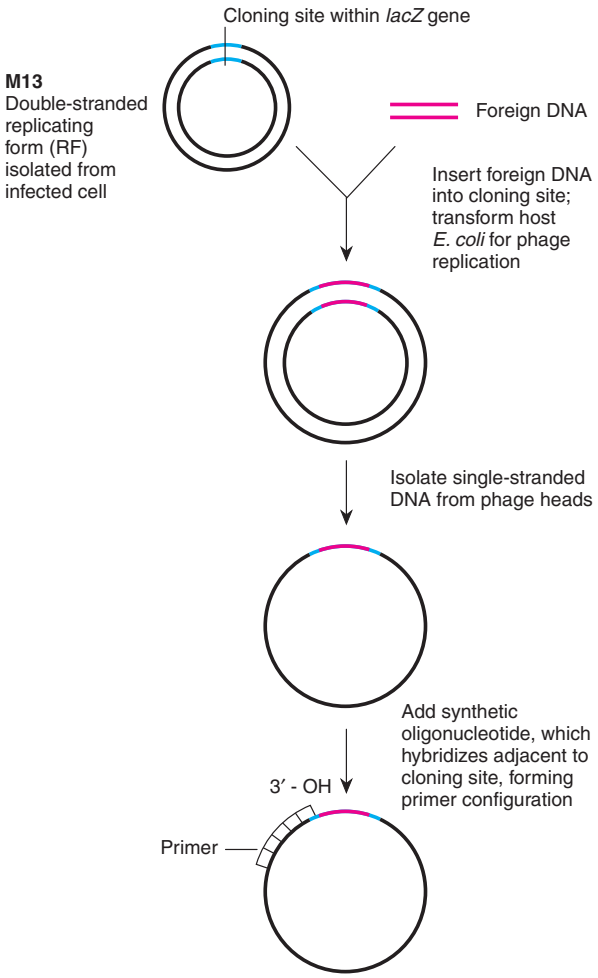


Figure 13.36 Phage M13, a useful vector for sequencing a piece of cloned DNA by the dideoxy method since it exists in both single- and double-stranded forms. In addition, it contains restriction sites within a copy of the *lacZ* gene (*blue*). This allows for the selection of clones with inserted pieces of foreign DNA. An artificial oligonucleotide, which hybridizes adjacent to the cloning site, provides the primer configuration needed for new synthesis.

Either the dideoxy or the chemical method of sequencing (not discussed) allows us to read the sequence of hundreds of nucleotides on a single gel. Whole viral, prokaryotic, and eukaryotic genomes, and numerous regions of interest in prokaryotes, eukaryotes, and viruses have been sequenced. As W. Gilbert said in his Nobel Prize acceptance speech in 1981, "When we work out the structure of DNA molecules, we examine the fundamental level that underlies all processes in living cells."

MAPPING AND SEQUENCING THE HUMAN GENOME

Locating a Gene of Interest

Genes of importance can be searched for directly. A breast cancer gene provides a good example. Other genes that have been found this way include the genes for cystic fibrosis and Huntington disease. The concept of finding a gene is relatively simple; the methodology is tedious. Searching for many genes, including medically important genes such as one for breast cancer, means looking for a gene only by its symptoms; that is, we don't know the protein product of the gene or its location. Searching begins by looking at pedigrees of families segregating the disease and then trying to correlate the occurrence of the disease with a particular RFLP or microsatellite marker. When this is done, the gene has been localized to a particular region of a particular chromosome. Then, with a genomic library, chromosome walking (see the next section) is done until a gene in the neighborhood of the marker is found that could be the target gene. With the gene in hand, its sequence and protein product can be determined, a first step in medical treatment.

Chromosome Walking

Despite the limited size of any one inserted piece of foreign DNA, it is possible to learn about longer stretches of DNA by using a technique of overlapping clones called **chromosome walking.** Let us say that a particular gene (in region A) is located in clone 1, as discovered through probing. The cloned insert can be removed, using the same restriction enzyme initially used to insert it in the vector, and broken into small pieces that are used as probes themselves. The idea is to locate another clone with an inserted region that overlaps the first one (fig. 13.38). The second clone is now treated the same way—with segments used to probe for yet another overlap farther down the chromosome. In this way, relatively long segments of a chromosome can be available for study in overlapping clones.

One obvious use of chromosome walking is to discover what genes lie next to each other on eukaryotic chromosomes. The technique is very tedious and is halted at certain areas not amenable to walking, such as repeated sequences found in the DNA of eukaryotes (see chapter 15). Once an overlapping probe contains a commonly repeated sequence, it hybridizes to many clones that do not contain adjacent segments. This "cross-referencing" lessens the value of the technique. Currently, newer techniques (termed **chromosome jumping**), designed to bypass regions not amenable to walking, are being developed. These techniques depend on the ability to locate

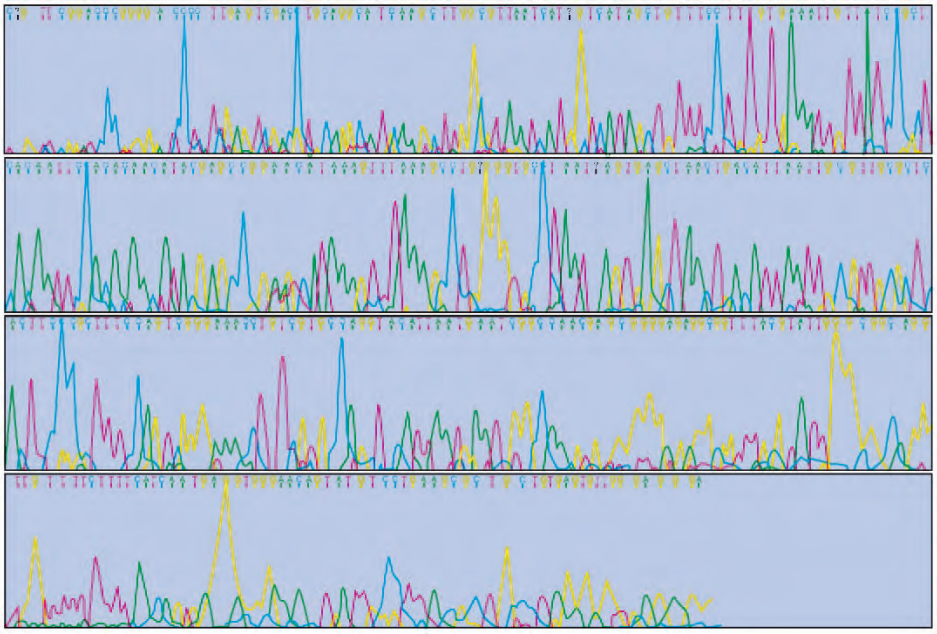
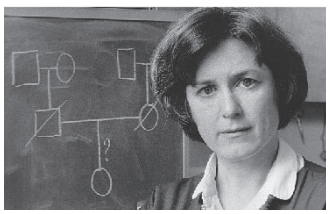


Figure 13.37 Processed data from automated DNA analysis using fluorescent dyes. The DNA is sequenced by attaching a different fluorescent dye to each dideoxy base. Thus, the dideoxy bases can be identified by their fluorescent color in a laser light rather than by which lane they occupy in a gel. Only one lane need therefore be run. In this diagram, guanine is *yellow,* cytosine is *blue,* adenine is *green,* and thymine is *red.* The sequence is read *left to right, top to bottom.* (From L. Johnston-Dow, et al., *BioTechniques,* 5:754–65, 1987, copyright © 1987. Eaton Publishing, Natick, MA. Reprinted with permission.)

the two ends of a segment without having to walk through the middle. Ends of a segment can be located if the region has been inverted or if a large region is cloned and the middle part later removed, leaving just the ends. A probe of the ends allows the investigator to locate clones with first one end and then the other, effectively jumping over the intervening region.

The Breast Cancer Gene

The initial location of the breast cancer gene *BRCA1* was determined by M. King in 1990 using a marker (*D17S74*) on the long arm of chromosome 17 (fig. 13.39); it was the 183rd marker that King had tried (fig. 13.40). The breast cancer gene *BRCA1* was particularly difficult to locate because it accounts for only about 5% of all breast cancers. However, it accounts for a much higher percentage of inherited, early onset breast cancers, those in women under fifty years of age. One woman in two hundred inherits this gene, and among those women, 80 to 90% risk developing the disease. The actual locating and cloning of this gene was done in 1994 by a team led by M. Skolnick. The gene codes for a protein of 1,863 amino acids; it seems to act as a tumor suppressor protein (see chapter 16). Its mechanism of action is as a transcription factor associated with RNA polymerase II (see chapter 10).



Mary-Claire King (1946–). (Courtesy of Office of Public Information, Berkeley Campus, University of California. Photograph © Jane Scherr.)

The Human Genome Project

The Standard Method

In chapter 6, we developed a human chromosome map. Generally, a locus was located on a particular chromosome by tissue culture techniques (somatic-cell hybridization). Loci could be pinpointed further using aberrant chromosomes, such as those with deletions. If a locus was present when the intact human chromosome was present but absent if the deletion chromosome was present, the gene could be localized to the deleted region. In addition, probes for specific genes can show us roughly where that gene is located (fig. 13.41).

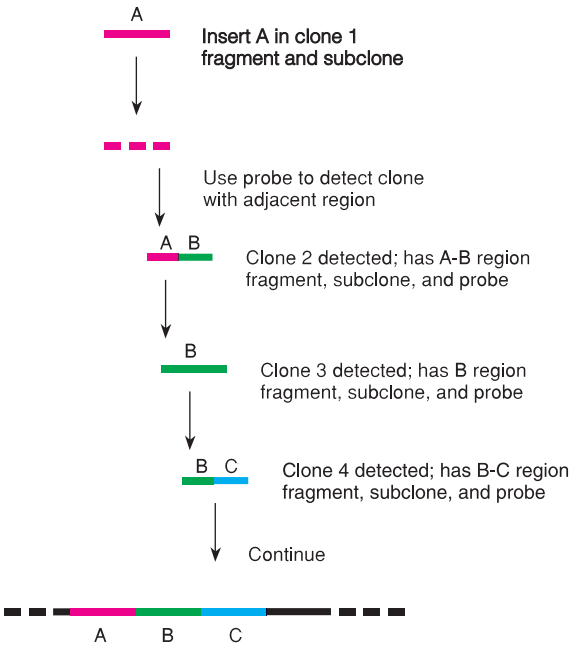


Figure 13.38 Chromosome walking technique. This technique allows one to study long chromosomal regions by locating overlapping cloned inserts. We begin with a specific cloned piece of DNA, referred to as insert A. This piece is fragmented to create probes for other clones in a genomic library that contains regions that overlap A (the next region down is referred to as B). The A-B clone is itself then fragmented to create probes to repeat the process, moving down the chromosome.

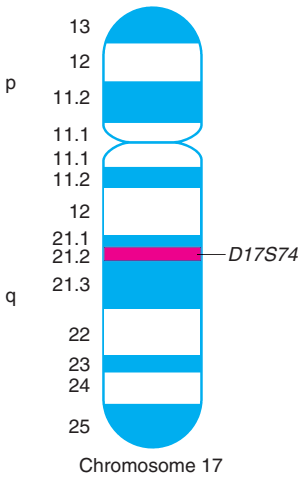


Figure 13.39 A Giemsa-banded chromosome 17, showing the numbering of the regions and the location in region 21q in which marker *D17S74* is located. The terminology of the marker is that of section 74 of chromosome 17. This marker correlated to the position of the *BRCA1* gene.

Two general methods were developed for mapping the human genome, the *standard method,* supported in large part by federal funding, and the **whole-genome shotgun method** used by the Celera Genomics Company. In the standard method, the project is reduced to finding a segment of the genome and locating where it belongs. The segment is then sequenced. By the overlap of sequenced pieces, the whole genome is pieced together. Mapping is done chromosome by chromosome since individual chromosomes can be isolated in large numbers by the methods of flow cytometry, described in chapter 15. In the initial stages of the Human Genome Project, when the primary task was mapping, yeast artificial chromosomes (YACs) were the primary cloning agent. However, as the emphasis of the project shifted

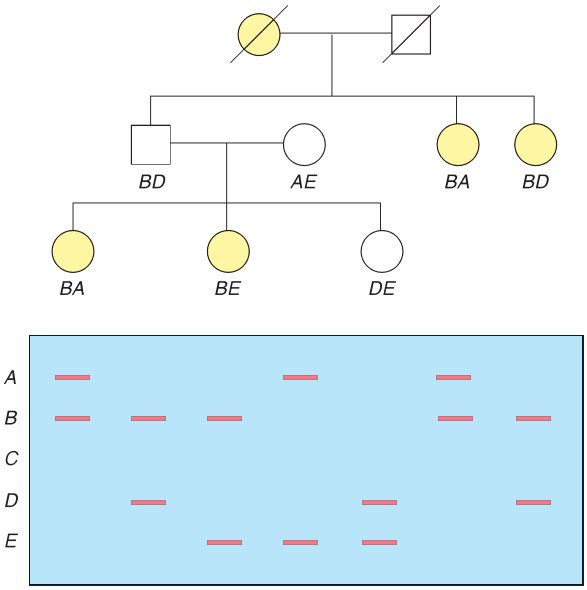


Figure 13.40 A pedigree of a family of individuals in which early onset breast cancer is segregating. At the *bottom* of the figure is a gel of the various bands produced, showing the alleles of *D17S74,* marked *A–E* in decreasing size of fragment probed. The individuals in the pedigree are shown directly over their lanes in the gel. The original parents were dead (*diagonal line*) and thus were not typed. The mother, two of her daughters, and two of her granddaughters were diagnosed with breast cancer in ages ranging from twenty-three to forty-five years of age (*yellow*). Note that in every case of breast cancer, the woman has the *B* allele of marker *D17S74.* It is this correlation that localized the breast cancer gene to that region of the chromosome. *D17S74* was the 183rd marker M. King and her colleagues studied; the other markers showed no correlation with breast cancer. (Reprinted with permission from J. M. Hall et al., "Linkage of Early-Onset Familial Breast Cancer to Chromosome 17q21," *Science,* 250:1684–89, 1990. Copyright © 1990 American Association for the Advancement of Science.)

Figure 13.41 The physical location of a gene or marker can be found by probing chromosomes with a complementary DNA sequence that has a specific fluorescent compound bound to it. When activated, the probe is seen (*bright yellow spots*) in a laser scanning confocal microscope. The chromosomes are counterstained with propidium iodide, which makes them fluoresce *red.* In this case, the probe has located a sequence on human chromosome 11. (© Peter Menzel/Photographed at Yale University Medical School.)

to sequencing, **bacterial artificial chromosomes (BACs)** were used. The bacterial artificial chromosomes are derivatives of the fertility factor (F factor, see chapter 7). They have properties of stability and homogeneity that make them more compatible with automated sequence techniques.

To begin sequencing, each individual chromosome is broken up into overlapping segments of about 150,000 bp in a BAC library. Each BAC is then digested into smaller pieces that are cloned in cosmids or P1 phages digested into smaller pieces for sequencing.

Before we define the techniques further, we should mention that we are not dealing with just one map of the genome, but several different kinds of maps. Although the ultimate goal was the complete DNA sequence of the genome, yielding the exact location of every gene, we needed to go through several stages to get there—remember, we are trying to keep track of 3.3 billion bases. We are familiar with the genetic linkage map of chromosomes described in chapter 6. These maps are called **classical linkage maps;** they define distances in recombination frequencies. A **modern linkage map** is one that uses RFLP markers along its length instead of genes. There is also a **physical map,** in which distances are in physical units of base pairs. These maps can be of microsatellite markers or of **sequence-tagged sites (STSs).** Sequence-tagged sites are DNA lengths of 100-500 base pairs that are unique in the genome. They are created by polymerase chain reaction amplification of primers obtained by sequencing segments of the genome. The primers are then tested to be sure the sequence is unique. About 50% of attempts yield sequence-tagged sites.

The physical map can also be marked off in differences among individuals that amount to changes in single base pairs. These differences are called **single-nucleotide polymorphisms (SNPs**—pronounced "snips"). These are located about every one thousand bases along the human genome. These single-nucleotide polymorphisms are expected to be especially useful in keeping track of differences among individuals in genes responsible for diseases.

RFLPs, microsatellite markers, STSs, and SNPs allow us to keep track of BACs and cloned pieces in cosmids and P1 phages. However, as we locate various DNA pieces, we will be building up continuous regions of a chromosome by overlapping these pieces. These overlapping, contiguous clones are referred to as **contigs.** This process is repeated chromosome by chromosome. In other words, we are creating a library of overlapping clones that cover the complete length of each chromosome. In essence, we are putting together a linear jigsaw puzzle. Contigs are created by comparing the segments that clones have in common, if any (fig. 13.42). From shared segments, we can infer which parts of the clones overlap. Through this process, contigs of parts of the chromosome can be built up (fig. 13.43). Later, contigs comprising part of a chromosome can be ordered by taking an end clone of a completed contig and using it as a probe to begin chromosome walking to find an end clone of a nearby contig (fig. 13.44).

For example, let's begin with a BAC from chromosome 7 of 150,000 base pairs long with three sequence-tagged sites located along its length. We can determine neighboring BACs by shared sequence-tagged sites. The BAC is then digested and cloned into cosmids. The

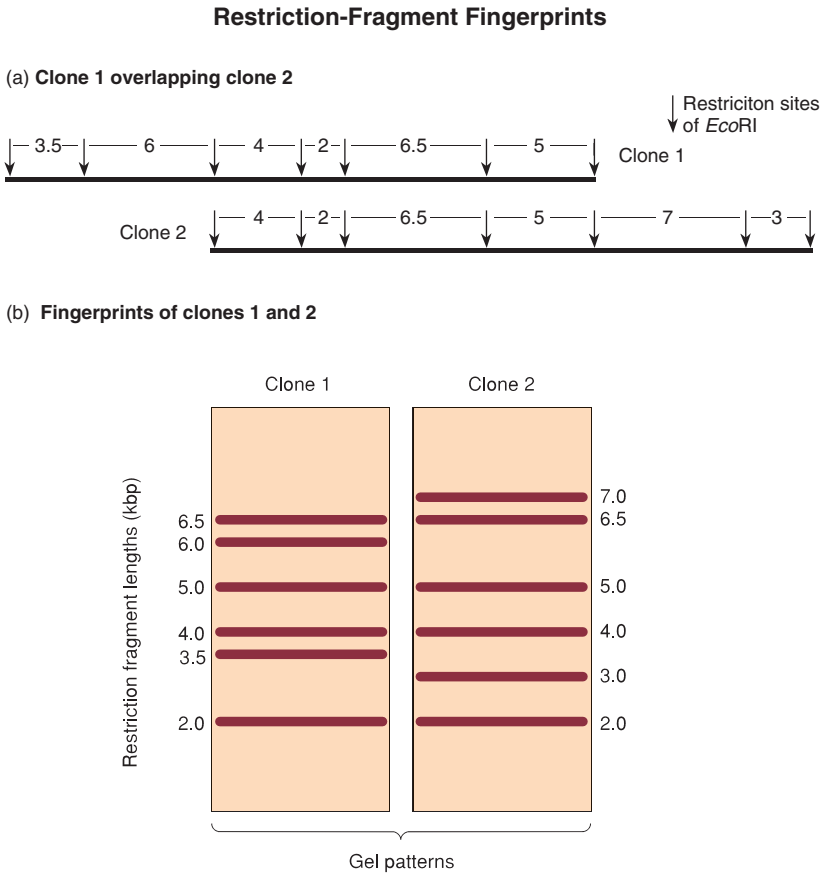


Figure 13.42 To create contigs, researchers must find overlapping clones and determine their region of overlap. In part (*a*), we have two overlapping pieces of DNA, found by chromosome walking. The pieces are digested with *Eco*RI and electrophoresed, producing the blots in part (*b*). From these gels, we see that fragments of 2.0, 4.0, 5.0, and 6.5 kb pairs are in common, indicating that they are in the region of overlap in both clones. We have thus isolated the overlap region and the unique end regions of both clones (compare *c* with *a*). Restriction maps can then be made of each segment, ordering the pieces. (Reprinted courtesy of *Los Alamos Science,* Volume Number 20, a publication of Los Alamos National Library, Los Alamos, NM.)

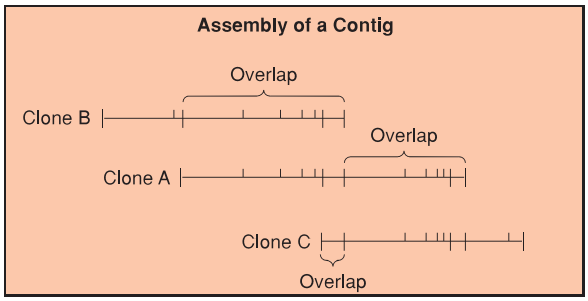


Figure 13.43 A contig is further built up by assembling pairwise overlapping clones into longer sequences. Here we see that clone A overlaps clone B to the *left* and clone C to the *right*. In this case, there is one fragment common to all three clones. By comparing clones in this manner, we can march down the chromosome, creating a larger and larger contig. (Reprinted courtesy of *Los Alamos Science,* Volume Number 20, a publication of Los Alamos National Library, Los Alamos, NM.)

overlap of cosmids can be determined by sequence-tagged sites, RFLPs, SNPs, or microsatellites in common. The cosmids are then digested and sequenced. From the sequences we work back, finding overlap and thereby constructing a contig of that BAC. The same process is carried out on neighboring BACs, extending the contig eventually to cover the entire chromosome.

At the initiation of the Human Genome Project, various goals were set. A modern linkage map of microsatellite markers of the human genome was targeted to be complete when markers were spaced about 0.7 centimorgans (about 700,000 base pairs) apart. That goal was reached in 1996 with 2,335 microsatellite markers located on the genome. The physical map of sequence-tagged sites would be considered complete with markers every 100,000 bases, the equivalent of 30,000 sequence-tagged sites in the genome cloned in BACs. That goal was reached in 1997. The sequence of the complete genome was targeted for 2001 and announced in 2000.

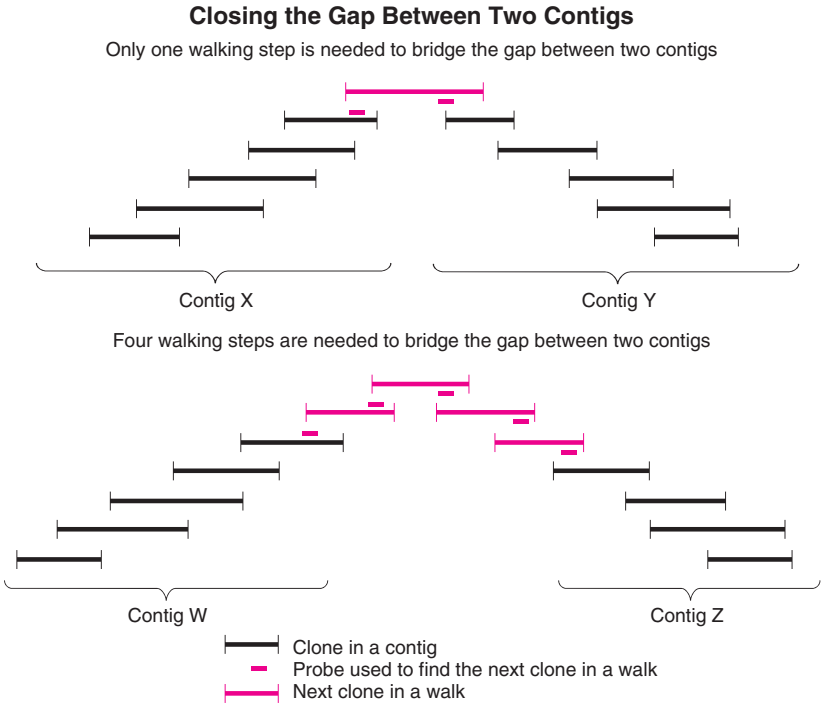


Figure 13.44 When contigs of large parts of a chromosome are built up, they need to be connected. We can do this directly if there is an overlap at the end of one contig and the beginning of the next. Barring that, we must do chromosome walking to find clones that bridge the gap between two contigs. At the *top* of the figure, in typical chromosome walking technique, the DNA of an end clone is fragmented and used to probe for an overlap. In this case, one clone is found that overlaps two contigs and thus joins them into one long contig. In the *bottom* portion of the figure, the walk requires finding four overlapping clones that bridge the gap between the two contigs. In both cases, the process is successful, joining two contigs into one longer one. (Reprinted courtesy of *Los Alamos Science,* Volume Number 20, a publication of Los Alamos National Library, Los Alamos, NM.)

One of the reasons that goals were optimistic is that methods of mass production have been developed as the project has moved along. These methods include the automation of sequencing and cloning and the development of some new technology. For example, scientists at Affymetrix, Inc., have developed the equivalent of a DNA probe computer chip. Thousands of known DNA sequences are synthesized on a glass substrate. The DNA to be probed is introduced to this chip, where hybridization will take place. Using fluorescent technology, successful probing can be determined using a laser confocal scanning system (fig. 13.45). These chips allow extremely rapid analysis of DNA sequences. Several other manufacturers have developed similar technologies.

As mentioned at the beginning of this chapter, J. C. Venter of Celera Genomics was a co-announcer of the completion of the sequencing of the entire human genome. Venter and his colleagues used a whole-genome shotgun method in which the entire human genome was broken into small segments, cloned, and sequenced. The Celera group will then piece together the genome with a massive computing effort. Previously, it had been thought that this method could not work on a genome as large as the human genome. Venter and his colleagues, however, had sequenced the *Drosophila* genome (180 million base pairs) by March of 2000 by this method. Venter and his colleagues had also sequenced the first true organism, the bacterium *Haemophilus influenzae* (1.8 million base pairs) in July of 1995. Since that time, the yeast *Saccharomyces cerevisiae* (12 million base pairs) was sequenced in 1996, and a significant genetic model organism, the nematode worm, *Caenorhabditis elegans* (97 million base pairs; see chapter 16), was sequenced in 1998. Since 1995, numerous other bacteria and eukaryotes have had their genomes sequenced.

Bioinformatics and Proteomics

These incredible accomplishments in genomics have given rise to two newly named sciences, **bioinformatics** and **proteomics.** Bioinformatics is the science of mining the data from the DNA sequences obtained from sequencing. Mining refers to the storage, retrieval, and analysis of the data. Proteomics is the study of the **proteome,** from *protein*s of the gen*ome,* and refers to the study of the complete set of proteins from a particular

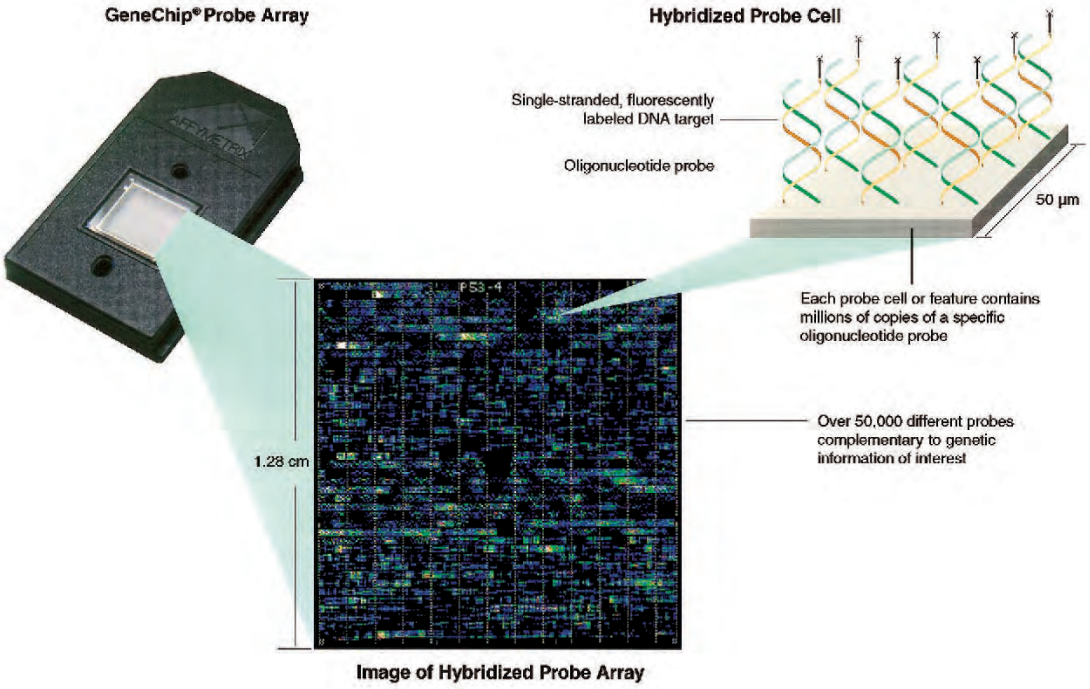


Figure 13.45 The Gene Chip® DNA probe array is a glass wafer containing from sixty-five thousand to one million or more different DNA sequences. The chips are created by photolithographic techniques, similar to those used in computer chip manufacture. The DNA being probed must have fluorescent molecules attached to permit rapid screening. (Courtesy of Affymetrix, Inc.)

genome. It is the protein analogue to genomics. It is estimated that there are from 50,000 to 2 million different proteins in biological systems, although the number of distinct shapes—motifs—may only be about five thousand. The role of proteomics is to characterize all proteins and determine their structures and shapes. As classical genetics worked from phenotype to genotype, modern molecular genetics is working from genotype (genomics) to phenotype (proteomics).

Ethics

In addition to the expected scientific and medical information that we will gain from sequencing of the human genome are ethical problems that the project will create. We will shortly have the ability to test people for various genes that we cannot test for now, such as genes for latent diseases like coronary artery disease and cancer. Can insurance companies then demand to test individuals for a whole battery of genes and decide afterward whether that person is insurable or what that person's insurance rates should be? Will many persons find themselves uninsurable because they have genes that might predispose them to cancer? Will individuals find themselves unemployable because of similar problems? What should doctors do about diagnosing a genetic disease (such as Huntington disease) that has no cure? Should they tell the patient? Another ethical issue is the extent to which genetic intervention should be used to change the course of a person's life. With the knowledge of the sequence and location of our genes, and the technology to transfer genes into people, will transgenic people become the norm (see box 13.2)? Should we not only cure diseases this way but tailor a person to some ideal? Will genetic intervention into our basic genetic blueprint be routine? To address these questions, an ethics panel has been set up as part of the Human Genome Project.

PRACTICAL BENEFITS FROM GENE CLONING

Throughout this chapter, we have mentioned applications of genetic engineering. Here we summarize some of the accomplishments and future directions in the medical, agricultural, and industrial arenas.

Medicine

In medicine, genetic engineering has had remarkable successes in some areas. On the one hand, basic knowledge about how genes work (and don't work) has advanced tremendously. On the other hand, recombinant DNA methodology has made available large quantities of substances previously in short supply. These include insulin, interferon (an antiviral agent), growth hormone, growth factors, blood-clotting factors, and vaccines for diseases such as hepatitis B, herpes, and rabies. Advances in AIDS and cancer research are discussed in chapter 16. Genetic engineering is making it possible to manufacture antibodies to diagnose and treat diseases. The sequencing of the human genome will further aid medicine by identifying the genes for various diseases, a first step in discovering cures. So far, several genes of great importance have been located, cloned, and sequenced. We also pointed out the use of restriction fragment length polymorphisms and the polymerase chain reaction as techniques of tremendous power in identifying individuals for forensic purposes.

On another front, transgenic mice and cloned sheep have shown that genetic engineering can be applied to higher organisms (fig. 13.46). The use of this technology to treat human diseases, however, is only just beginning. In July 1990, the National Institutes of Health approved gene therapy treatments on people: A child was infused with cells to replace a gene for the enzyme adenosine deaminase, an enzyme whose absence results in a dysfunctional immune system. Although the latter treatment was successful, it had been augmented by other treatments, rendering the conclusions equivocal. Mice and dogs have had hemophilia B corrected by infusion of a genetically modified adenovirus. AIDS, hemophilia, cystic fibrosis, and diabetes are other diseases that should be amenable to gene therapy in the near future.

Figure 13.46 The sow shown is transgenic, producing large quantities of human protein C in her milk. The protein controls blood clotting and is normally found only in trace quantities in human blood. (Courtesy of William H. Velander, Virginia Tech.)

Agriculture

Currently, in the United States, approximately one quarter of farmland is planted with crops that are genetically modified. Most are resistant to certain insect pests because they contain genes from *Bacillus thuringiensis* (often referred to as Bt). These genes are for insecticidal proteins called δ endotoxins. For example, the proteins Cry1A and Cry1C from *Bacillus thuringiensis* protect the plants against larval forms of lepidopterans such as the European corn borer. Cry3A protects against coleopterans such as the Colorado potato beetle. In excess of fifty genetically altered crop plants have been approved for planting, including those protected against insect pests, frost, and premature ripening. Rice is being modified so that its vitamin A potential is maintained even after the husks are removed, a procedure done to allow for storage since the husks become rancid. That change alone will improve the health of millions of people throughout the world. Box 13.1 discussed some of the ethical concerns surrounding genetically modified crop plants.

Industry

Industrial applications of biotechnology include engineering bacteria to break down toxic wastes, modifying yeast to use cellulose to produce glucose and alcohol for fuel, using algae in mariculture (the cultivation of marine organisms in their natural environments) to produce both food and other useful substances, and developing better food processing methods and waste conversion. As an example, baker's yeast (*Saccharomyces cerevisiae*) has been modified with a plasmid that contains two cellulase genes, an endoglucanase and an exoglucanase, that convert cellulose to glucose. The yeast can then convert glucose to ethyl alcohol. These yeasts are now capable of digesting wood (cellulose) and converting it directly to alcohol. The potential exists to harvest the alcohol the yeast produces as a fuel to replace fossil fuels that are in dwindling supply and are polluting the planet.

As you can see, there is no one direction that biotechnology is taking. Many advances are being made that will probably affect every person's life in a beneficial way. Cautious optimism is certainly in order.

SUMMARY

STUDY OBJECTIVE 1: To look at the techniques of gene cloning 359–377

Recombinant DNA techniques revolve around the cloning of foreign DNA in a plasmid or phage. Cloned DNA can be amplified, expressed, and sequenced. Gene cloning techniques came about with the discovery of restriction endonucleases. Type II restriction endonucleases cleave DNA at palindromic regions, which have twofold symmetry.

Recombinant vectors can be constructed several different ways. Foreign and vector DNA can be made compatible by treating each with the same restriction endonuclease—each will then have the same sticky ends. If that does not work, T4 DNA ligase can join blunt ends. In a variation of this method, linkers containing restriction sites are added to vector and foreign DNA. These linkers are then treated with a restriction endonuclease that gives the DNA sticky ends.

DNA to be cloned can be synthesized from an RNA template (cDNA) or isolated by various techniques. If messenger RNA is available, it can be converted into a clonable complementary DNA with the enzyme reverse transcriptase. If DNA is to be isolated directly, it must be identified among all the other DNA fragments created. Locating a desirable piece of DNA is done with probes, complementary nucleic acids labeled with radioactivity or chemiluminescence. Southern blotting, a transfer technique, is used first, followed by DNA-DNA or DNA-RNA hybridization and autoradiography. If the DNA is cloned first, as in the creation of a genomic library, probes can be created or expression of the cloned gene can be determined.

Eukaryotic vectors have been developed, including yeast plasmids, tumor virus vehicles in animals, and crown gall tumor plasmids in plants. Eukaryotes can be transfected by foreign DNA and express it in transgenic organisms. DNA can be injected, shot in on projectiles, electroporated, or introduced by viruses, plasmids, or liposomes. Knockout mice, lacking a specific gene, can be created.

STUDY OBJECTIVE 2: To examine the techniques of creating restriction maps 377–383

Restriction digests can be separated by electrophoresis, then used to construct a restriction map. This is a map of the DNA showing the location of restriction enzyme recognition sites. The genetic maps, generated by mating analysis, can then be superimposed on the restriction maps, locating regions of interest on the physical map. Restriction fragment length polymorphisms (RFLPs) provide a tool for locating genes through linkage analysis and are also valuable in forensic science. The polymerase chain reaction (PCR) is a technique used to rapidly amplify particular segments of DNA.

STUDY OBJECTIVE 3: To study the methods of DNA sequencing 383–390

DNA is usually sequenced by one of two methods. The dideoxy method developed by Sanger and his colleagues requires the synthesis of DNA in the presence of chain-terminating (dideoxy) nucleotides. Electrophoresis followed by autoradiography allows the sequence of nucleotides synthesized to be determined directly. Fluorescent labeling allows computerized sequence determinations. The phage ϕX174 was sequenced in its entirety through the forerunner of this technique, the plus-and-minus method. Gilbert and Maxam's chemical method also is used widely.

STUDY OBJECTIVE 4: To look at the goals and methods of the Human Genome Project 390–397

The Human Genome Project is a massive, international effort to map and sequence all 3.3 billion bases of the human genome. Initial success was announced in the spring of 2000. Modern linkage maps are being created of restriction sites, microsatellite markers, sequence-tagged sites, and single-nucleotide polymorphisms. They are being coordinated with physical maps created with overlapping contiguous clones of chromosomes. These techniques currently allow us to find genes of interest. The project also includes the sequencing of the genomes of other relevant organisms.

STUDY OBJECTIVE 5: To look at the practical benefits and human issues of genetic engineering 397–398

Genetic engineering is moving forward on a number of fronts. Medical, agricultural, and industrial applications are becoming widespread.

SOLVED PROBLEMS

PROBLEM 1: A piece of eukaryotic DNA is obtained by using a restriction endonuclease that leaves blunt ends (*Hae*III). How could we get this piece of DNA into a *Bam*HI site in plasmid pBR322, and how would we know when the foreign DNA has been cloned?

Answer: Since the two pieces of DNA (the eukaryotic piece and the plasmid) have different ends, they must be made compatible before cloning. The simplest way would be to attach blunt-ended linkers to the foreign DNA with phage T4 DNA ligase (see fig. 13.9). The linkers, of course, would have a *Bam*HI site within. After the linkers are attached to the foreign DNA, it would be treated with the *Bam*HI restriction enzyme, giving the foreign DNA *Bam*HI ends. The plasmid is then also treated with the restriction enzyme and the two (the foreign DNA and the cut plasmid) are now mixed together in the presence of *E. coli* DNA ligase, which seals up the plasmids, with or without cloned inserts (see fig. 13.6). Since they have compatible ends, some of the time, a piece of foreign DNA is inserted into a plasmid. The plasmids are then taken up by *E. coli* cells that are grown overnight in an incubator. The bacterial colonies are then replica-plated on media with the antibiotics ampicillin or tetracycline. Colonies that are resistant to ampicillin but sensitive to tetracycline are assumed to be bacteria containing plasmids with cloned inserts (see fig. 13.8).

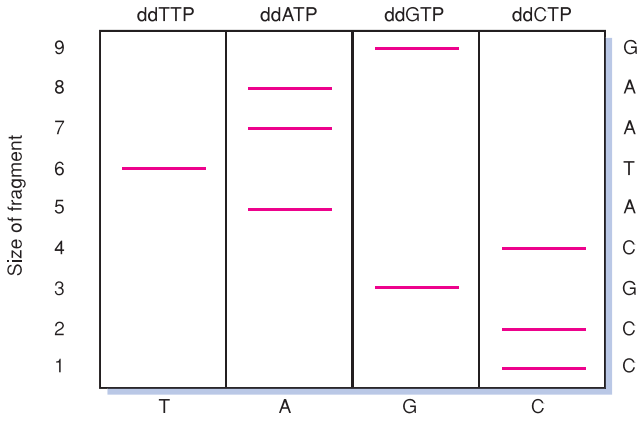
PROBLEM 2: How does a reporter system work?

Answer: A reporter system is a genetically manipulated system that displays a particular phenotype or reaction when a desired event has taken place. In this chapter, we discussed the firefly luciferase reporter system in which the desired result (transcription of a particular promoter) causes a transgenic tobacco plant to glow. Let us say that we are studying the control of transcription of a particular eukaryotic gene. We could attach the promoter of that gene to the firefly luciferase gene in a Ti plasmid by cloning techniques. The plasmid could then transfect tobacco plants, and we could continue our experiment to determine whether the promoter under study is active under various conditions. We would know whether it was active by watering the plants with luciferin. If the plant glows, then the luciferase gene product is present, which means that the promoter under question is active. In other words, the glowing of the plant "reports" the action of the promoter under question; the promoter is active because it allowed the transcription of the luciferase gene. We also discussed the green fluorescent protein reporter system.

PROBLEM 3: A piece of DNA has the sequence 3′-GGCG-TATTC-5′. It is sequenced using the dideoxy method. How many bands are found on the ladder gel? How many bands and of what size are found for each reaction mixture?

Answer: Since the piece of DNA is nine bases long, the total number of bands in all four lanes of a sequencing gel add up to nine (see fig. 13.33). By each reaction mixture, we mean the four reaction mixtures each with one of the dideoxynucleotides. In the reaction mixture with ddTTP,

chain termination occurs at the adenine in the piece of DNA; that is, a DNA segment was synthesized that is six bases long (see the following figure). In the reaction mixture with ddATP, chain termination occurs opposite each of the thymines, producing DNA segments of five, seven, and eight nucleotides. In the reaction mixture with ddGTP, chain termination occurs opposite the cytosines in positions three and nine. And, in the reaction mixture with ddCTP, chain termination occurs after synthesis of segments one, two, and four bases long. Note that the gel gives us the sequence of the complement strand of the original piece of single-stranded DNA.



PROBLEM 4: A linear DNA molecule 1,000 base pairs long is digested with the following restriction enzymes, producing the following results:

*Eco*RI	400 bp, 600 bp
Bgl II	250 bp, 750 bp
*Eco*RI + *Bgl* II	250, 350, 400 bp

Determine the restriction map.

Answer: Each enzyme alone produces two fragments, so the molecule has one site for each enzyme. Since we get different-sized fragments with each enzyme, the sites must be located asymmetrically along the DNA. Draw these sites:

*Eco*RI *Bgl*II

400 ↓ 600 and 250 ↓ 750

The *Eco*RI fragment that lacks a *Bgl*II site should appear in the double digest. If *Bgl*II cuts within the 400 base-pair fragment, we would expect to see 150, 250, and 600 base-pair fragments. We don't see this, so the *Bgl*II site is not within the 400 base-pair *Eco*RI fragment. Thus, the map looks like this:

*Eco*RI *Bgl*II

400 ↓ 350 ↓ 250

EXERCISES AND PROBLEMS*

GENOMIC TOOLS

1. What specific properties of type II endonucleases make them useful in gene cloning?
2. The following is a double helix of DNA. What, if any, are potential restriction enzyme recognition sequences?

   5′-TAGAATTCGACGGATCCGGGGCATGCAGATCA-3′
   3′-ATCTTAAGCTGCCTAGGCCCCGTACGTCTAGT-5′

3. Assuming a random arrangement of nucleotides on a piece of DNA, what is the probability that a restriction endonuclease whose recognition site consists of four bases (a four-cutter) will cut the DNA? What is the probability for a six-cutter? an eight-cutter?
4. Under what circumstances is a restriction endonuclease unsuitable for cloning a piece of foreign DNA?
5. What methods exist to create sticky ends or create ends for joining two incompatible pieces of DNA? When is each method favored?
6. Diagram a possible heteroduplex between two phage λ vectors, one with and one without a cloned insert, created by DNA-DNA hybridization.
7. What are the differences among plasmid, cosmid, expression vector, and YAC? Under what circumstances is each useful?
8. What are the steps by which messenger RNA can be converted into cDNA? How would we obtain radioactive cDNA? radioactive messenger RNA?
9. What is chromosome walking? When is it used?
10. How would we isolate a human alanine transfer RNA gene for cloning? How would we locate a clone with a human alanine transfer RNA gene in a genomic library?
11. What are the differences among Southern, western, northern, and dot blotting?

*Answers selected exercises and problems are on page A-14.

12. How would you develop a probe for a gene whose messenger RNA could not be isolated? How could an expression vector be used to isolate a cloned gene?

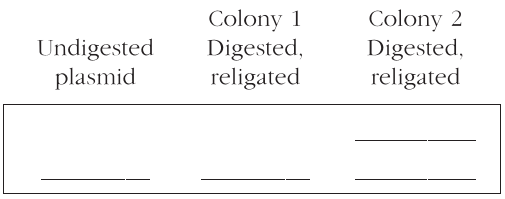
13. How are *E. coli* plasmids manipulated to survive in yeast? How can virus genomes, such as SV40 and phage λ, survive as functioning vectors when parts of their genomes are replaced by cloned DNA?

14. What methods are used to get foreign DNA into eukaryotic cells? What is transfection? What is a transgenic mouse?

15. Exonuclease III is an enzyme that sequentially removes bases from the 3′ end of double-stranded DNA. The following two molecules, each 100 bp long, are digested with exonuclease III. Molecule 1 is completely digested; molecule 2 is only partially digested. Explain these results.

Molecule 1:	CGTTCAG...
	GCAAGTC...
Molecule 2:	AAAAAAAAAA...
	TTTTTTTTTT...

16. A plasmid that contains an *Eco*RI site within a gene for ampicillin resistance is cut with *Eco*RI, and then religated. This plasmid is used to transform *E. coli* cells, and the plasmid is reisolated from the ampicillin-resistant colonies. The reisolated plasmids from two different colonies are electrophoresed, and the results appear in the following figure.



How do you account for the two bands in colony 2?

17. Most human genes contain one or more introns. Since bacteria cannot excise introns from nuclear messenger RNA (snRNPs are needed), how can bacteria be used to make large quantities of a human protein?

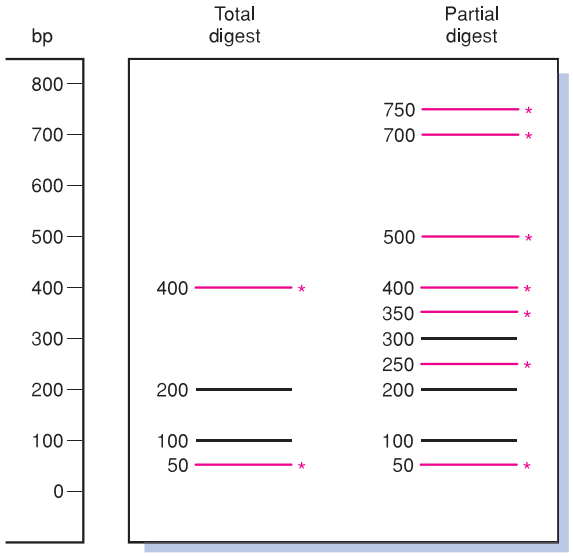
RESTRICTION MAPPING

18. How are DNA fingerprints useful in forensic cases? Could they be used in paternity exclusion?

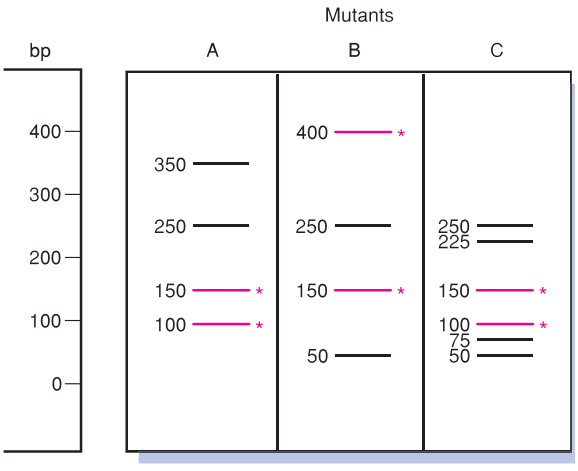
19. The following segment of DNA is cut four times by the restriction endonuclease *Eco*RI at the places shown. Diagram the gel banding that would result from electrophoresis of the total and partial digests. Note the end-labeled segments and regions where several segments form bands at the same place on the gel.

100 ↓ 300 ↓50↓ 250 ↓ 150

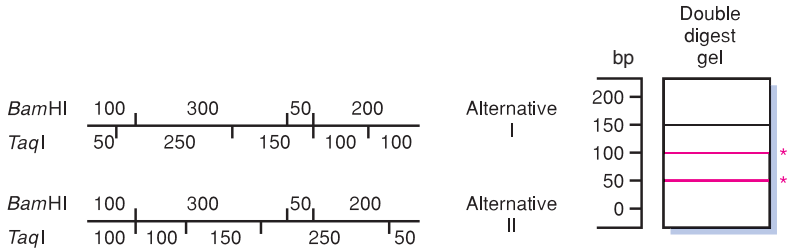
20. The following figure shows a gel of a total and partial digest of a DNA segment treated with *Hin*dII. End-labeled segments are noted by asterisks. Draw the restriction map of the original segment.



21. Several mutants of the DNA segment shown in problem 19 were isolated. They gave the following gel patterns when the total digests were electrophoresed. Asterisks denote the end-labeled segments. Can you determine the nature of the mutations?



22. Restriction maps of a segment of DNA were worked out separately for *Bam*HI and *Taq*I. Two overlays of the maps are possible. The double-digest gel is shown in the following figure (asterisks denote end labels). Which overlay is correct?



23. A linear DNA molecule 1,000 bp long gives the following size fragments when treated with these restriction enzymes. Derive a restriction map.

*Eco*RI:	300 bp, 700 bp
*Bam*HI:	150 bp, 200 bp, 250 bp, 400 bp
*Eco*RI + *Bam*HI:	50 bp, 100 bp, 200 bp, 250 bp, 400 bp

24. A linear DNA molecule cut with *Eco*RI yields fragments of 3 kb, 4.2 kb, and 5 kb. What are the possible restriction maps?

25. You have double-stranded DNA that you radioactively label at the 5′ ends. Digestion of this molecule with either *Eco*RI or *Bam*HI yields the following fragments. The numbers are in kilobases (kb), and an asterisk indicates the fragments that are labeled.

*Eco*RI: 2.8, 4.6, 6.2*, 7.4, 8.0*

*Bam*HI: 6.0*, 10.0*, 13.0

If unlabeled DNA is digested with both enzymes simultaneously, the following fragments appear: 1.0, 2.0, 2.8, 3.6, 6.0, 6.2, 7.4. What is the restriction map for the two enzymes?

26. A 12 kb DNA molecule cut with *Eco*RI yields one 12 kb fragment. When the original molecule is cut with *Bam*HI, three fragments of 2 kb, 4.5 kb, and 5.5 kb are produced. When the fragment from *Eco*RI is treated with *Bam*HI, four fragments of 2 kb, 2.5 kb, 3.0 kb, and 4.5 kb are produced. Draw a restriction map.

27. A plasmid 3 kb in length contains a gene for ampicillin resistance and a gene for tetracycline resistance. The plasmid has a single site for each of the following enzymes: *Eco*RI, *Bgl*II, *Hin*dIII, *Pst*I, and *Sal*I. If DNA is cloned into the *Eco*RI site, resistance to either antibiotic is not affected. DNA cloned into the *Bgl*II, *Hin*dIII, or *Sal*I sites abolishes tetracycline resistance, and DNA inserted into the *Pst*I site eliminates ampicillin resistance. If the plasmid is digested completely with enzyme mixes, the following fragments result:

Mixture	Fragment Size (kb)
*Eco*RI + *Pst*I	0.7, 2.3
*Eco*RI + *Bgl*II	0.3, 2.7
*Eco*RI + *Hin*dIII	0.08, 2.92
*Eco*RI + *Sal*I	0.85, 2.15
*Eco*RI + *Bgl*II + *Pst*I	0.3, 0.7, 2.00

Draw a restriction map of the plasmid, and indicate the locations of the resistance genes and the sites of enzymatic cleavage.

28. A gene has the following *Eco*RI restriction map (in kilobases):

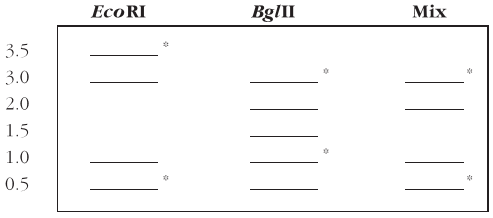
1.0 ↓ 0.7 ↓ 2.0

Draw the gel pattern expected from

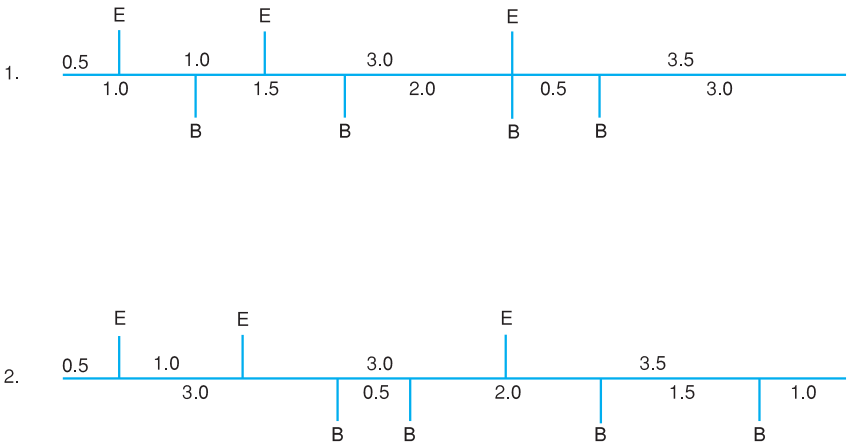
a. a mutant that has lost the site between the 1.0 and 0.7 kb fragments.

b. a mutant that has a new site within the 2.0 kb fragment.

29. A DNA fragment 8 kb in size is labeled with ^{32}P at the 5′ ends. It is then digested with *Eco*RI, *Bgl*II, or a mixture of both enzymes. The size of the fragments and the labeled fragments (*) appear as follows. Sizes are in kilobases.

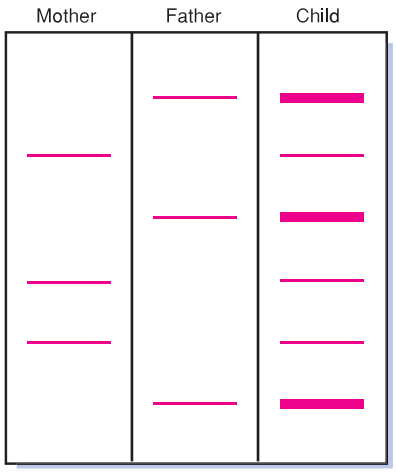


Which of the following two maps is consistent with the results?



30. You now take an unlabeled molecule from problem 29, digest it with *Hin*dIII, and get two fragments, 5.5 and 2.5 kb in size. If *Hin*dIII does not cut within the 3.5 kb *Eco*RI fragment, what size fragments do you expect in a double digest of *Hin*dIII and *Eco*RI?

31. Two normal individuals have a child with Down syndrome. RFLP analysis with a probe from chromosome 21 is performed on all three individuals, and the results of the gels appear as follows. Based on these results, what can you conclude about the origins of the number 21 chromosomes?



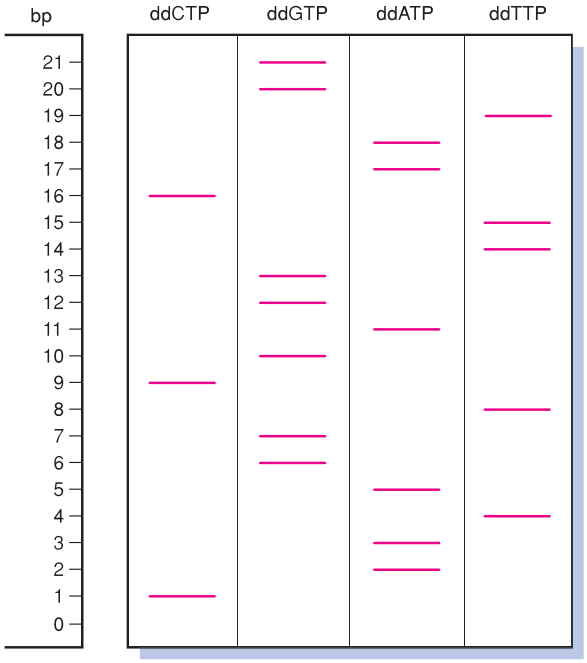
POLYMERASE CHAIN REACTION

32. What is PCR? When is it used?

DNA SEQUENCING

33. What are the steps in the dideoxy method of DNA sequencing? How has the technique been improved with fluorescent dyes?

34. The following diagram is of a dideoxy sequencing gel. What is the sequence of the DNA under study?



35. How can a particular piece of DNA be manipulated to be in the appropriate configuration for dideoxy sequencing?

36. Provide, if possible, DNA sequences that can mark the termination of one gene and the initiation of another, given that the genes overlap in one, two, three, four, five, six, or seven bases.

37. Draw the expected gel pattern derived from the dideoxy sequencing method for a template strand with the following sequence:

5′-CAGCGAATGCGGAA-3′

38. A DNA strand with the sequence 3′-GACTATTCCGAAAC-5′ is sequenced by the dideoxy method. If the reaction mixture contains all four radioactive deoxynucleotide triphosphates plus dideoxythymidine, what size labeled bands do you expect to see on the gel?

MAPPING AND SEQUENCING THE HUMAN GENOME

39. What is hypervariable DNA? a RFLP? a VNTR locus? microsatellite DNA? a sequence-tagged site? (*See also* RESTRICTION MAPPING)

PRACTICAL BENEFITS FROM GENE CLONING

40. Describe some areas of practical benefit from genetic engineering. Why might some people be concerned about its widespread use?

CRITICAL THINKING QUESTIONS

1. In the past, geneticists have used several different methods to splice pieces of DNA that do not have compatible "sticky ends." We mentioned blunt-end ligation and the addition of linkers containing specific restriction sites. Given that nucleotides can be added to the 3′ ends of double-stranded DNA with the enzyme deoxynucleoside terminal transferase, can you see another way to create compatible ends on foreign and vehicle DNA?

2. The motion picture *Jurassic Park* was based on the premise that DNA of dinosaurs could be extracted from the blood-meals of mosquitoes preserved in amber and inserted into the genome of a frog, which would then produce living dinosaurs. Is this premise reasonable?

Suggested Readings for chapter 13 are on page B-11.