DNA Manipulation and Its Applications

**Chapter Outline**

- Sequencing DNA
- Methods for Amplification of Select Segments of DNA
- Amplification by the Polymerase Chain Reaction
- DNA Cloning
  - Restriction Enzymes Are Used to Cut DNA into Well-Defined Fragments
  - Plasmids Are Used to Clone Small Pieces of DNA
  - Bacteriophage λ Vectors Are Useful for Cloning Larger DNA Segments
  - Cosmids Are Used to Clone the Largest Segments of DNA
  - Shuttle Vectors Can Be Cloned into Cells of Different Species
- Constructing a “Library”
  - A Genomic DNA Library Contains Clones with Different Genomic Fragments
  - A cDNA Library Contains Clones Reflecting the mRNA Sequences
  - Numerous Approaches Can Be Used to Pick the Correct Clone from a Library
  - Cloning in Systems Other than Escherichia coli
- Site-Directed Mutagenesis Permits the Restructuring of Existing Genes
- Recombinant DNA Techniques Were Used to Characterize the Globin Gene Family

**DNA Sequence Differences Were Used to Derive Defective Hemoglobin Genes**

**The β-Globin cDNA Probe Was Used to Characterize the Normal β-Globin Gene**

**Chromosome Walking Permitted Identification of the Regions around the Adult β-Globin Genes**

**Walking and Jumping Were Both Used to Map Cystic Fibrosis Gene**

**DNA from different sources can be knitted to produce new genetic alignments.**

A broad array of techniques has made it possible to investigate the fine structure of DNA as well as to redesign existing genes. The impact of these techniques can be overestimated. It is possible to map and isolate gene complexes of eukaryotes—features that could never have been accomplished by classical genetic methods. Genes can be designed to achieve a variety of practical goals. New genes may be inserted into the same species in which the original unmodified genes came or into other species.
In this chapter we first consider DNA sequencing. Then we explore the different approaches for amplifying and isolating specific genes or gene segments. Following this, we examine the methods currently available for restructuring existing DNA sequences. Finally, we look at some of the major advances in our understanding of gene structure and location that have resulted from the new technology. This part of the discussion focuses on two examples: The mapping of the human globin gene family and the mapping of the gene responsible for the genetically inherited disease, cystic fibrosis.

Sequencing DNA

If we are going to manipulate a segment of DNA, it is most useful to know something about its primary structure. The most complete information comes from total sequence analysis.

Initial efforts at sequencing nucleic acids were confined to RNA molecules that could be readily isolated in pure form. The first sequence to be determined was that for tyrosine tRNA from yeast. From roughly 100 pounds of yeast Robert Holley was able to isolate enough of the tyrosine tRNA to carry out a sequence analysis. This historically significant effort required a number of enzymes and chromatographic techniques. We shall not elaborate on this accomplishment because sequencing of RNA is no longer done directly. In fact, the sequencing of RNA has been replaced by the sequencing of "cDNA," which results from the reverse transcription of RNA into DNA.

Two quite different methods have been developed for sequencing DNA. One method, developed by Walter Gilbert and Alan Maxam and involving cleavage of preexisting DNA, uses a chemical approach. A second method, involving premature termination of newly synthesized DNA, uses an enzymatic approach that was developed by Fred Sanger.

For pure sequencing, Sanger's is the method of choice. It employs chain-terminating deoxyribonucleoside triphosphates to produce a continuous series of fragments in reactions catalyzed by DNA polymerase. Deoxyribonucleoside triphosphates (ddXTPs) resemble deoxyribonucleoside triphosphates except that they lack a 3'-OH group. They can add to a growing chain during polymerization, but they cannot be added onto, and as a result they act as chain terminators.

The DNA being sequenced is mixed with a suitable primer, radioactive dXTPs, DNA polymerase I (PolI), and a small amount of one ddXTP. The primers determine where DNA synthesis starts, and the ddXTP determines the base type where elongation stops. The products of four separate reaction mixtures, each differing only by the ddXTP it contains, all fragments with an reaction 2, using ddCTP, contains all C terminals. After the newly synthesized oligonucleotides are denatured from the template by denaturation they are separated by electrophoresis for a limited time on polyacrylamide gels. The positions of the fragments on the gel are determined by autoradiography. The sequence is read directly from the autoradiogram, starting with the first band (smallest) band at the bottom of the gel and moving upward, appearing from reaction 4, would be T; and 800 residues can be read from a single gel.

Methods for Amplification of Specific Segments of DNA

Cellular genomes are very large; even E. coli has more than a million base pairs, and eukaryotic genomes frequently contain a billion or more base pairs. Because of their large size, it is impractical to fractionate the cellular genome to obtain enough of a particular DNA segment for other investigations. Two methods of amplifying specific segments of DNA have been developed. The first method, polymerase chain reaction (PCR), uses DNA PolII using DNA primers that bind to the region of interest. This method involves repeated synthesis and is appropriately named the polymerase chain reaction (PCR) method. The second method involves DNA amplified in vivo. We discuss both of these methods and the reverse transcription method they are both useful in many ways.

Amplification by the Polymerase Reaction

PCR entails enzymatic amplification of specific DNA sequences using two oligonucleotide primers that flank the DNA segment to be amplified. The primers must complement opposite strands of DNA, and their 3' ends anneal in order to form an elongation template: each other.

The PCR procedure has three steps, which are repeated many times in a cyclical manner:

1. Denaturation of the original double-stranded DNA at high temperature
2. Annealing of the oligonucleotide primers to their complementary template at low temperature (37°C)
Figure 27.1

The Sanger dideoxynucleoside method of sequencing DNA. (a) A suitable template is chosen, and the primer is chosen so that DNA synthesis begins at the point of interest. The primer is radioactively labeled. In addition to the template–primer complex the reaction mixture contains all four radioactive deoxyribonucleoside triphosphates and small amounts of a single dideoxynucleoside triphosphate. The dideoxy compound serves as a chain terminator. (b) After synthesis in the presence of DNA polymerase I, the products of the reaction mixture are separated by gel electrophoresis and analyzed by autoradiography. The gel is run under denaturing conditions in warm urea so that single-stranded fragments are separated strictly according to size. For a given dideoxy compound a series of fragments terminating with that particular base should give a set of bands on the gel. The interpretation of the gel pattern is given in Figure 27.1. The smallest labeled fragment moves the fastest and appears closest to the bottom of the gel. (d) A typical sequencing film. The sequence begins CAATAAACGG. (Courtesy of Gibco-BRL, Life Tech Inc., Gaithersburg, Md.)

These steps are illustrated in figure 27.2. Each set of three steps comprises a cycle. The extension products of one primer provide a template for the other primer in a subsequent cycle so that each successive cycle essentially doubles the amount of DNA. This results in the exponential accumulation of the specific target fragment by approximately $2^n$, where $n$ is the number of cycles. The specific target fragment is also referred to as the "short product" and is defined as the region between the 5' ends of the extension primers. Each primer is physically incorporated into one strand of the short product.

Other products are also synthesized during the amplification of cycles, such as the "long product" of length, which is derived from the template molecule. However, the amount of long product only increases exponentially during each cycle of the amplification process; the quantity of original template remains constant.

At the end of the PCR process the short product is overwhelmingly abundant compared with the long product, so that its purification is not required for most purposes.
Figure 27.2
Steps in the polymerase chain reaction (PCR). The DNA to be amplified is denatured and annealed with two oligonucleotides that flank the region of interest. These oligonucleotides (or primers) are extended. Extension continues to the ends of the DNA strands. The products are again denatured and annealed to primers for a second round of extension. This process of denaturation, annealing, and primer extension is repeated many times. The primary product of the reaction is duplex DNA, bounded by the sequences of the primers. (From J. L. Marx, Multiplying genes by leaps and bounds, Science 240:1408–1410, June 10, 1988. Copyright 1988 by the AAAS. Reprinted by permission.)
Figure 27.3

Cleavage map of the SV40 genome. The zero point of the map is the unique EcoRI site. For clarity, the circular genome is shown opened at the R1 site, and the cleavage sites (and resulting fragments) for each restriction enzyme are indicated on a separate line.

DNA Cloning

The second method for DNA amplification is more complicated than PCR, but it has several advantages. DNA to be amplified by cloning is linked to a plasmid or a virus that can be replicated indefinitely in the appropriate host cell. After amplification the DNA of interest can be cut from the plasmid or virus and reisolated by gel electrophoresis. Cloning is not only useful for amplifying a segment of DNA, it can be adapted to the isolation of a DNA segment of interest from a large mixture such as is obtained from the isolation of the entire genome.

Restriction Enzymes Are Used to Cut DNA into Well-Defined Fragments

Most of the enzymes that are absolutely essential for cloning were discussed in the previous chapter. The most important enzymes that have not been discussed yet are the restriction enzymes. Systematic cleavages of duplex DNA at specific sites requires restriction enzymes. Each species of bacteria harbors a unique restriction enzyme, and hundreds of restriction enzymes with different specificities have been isolated, giving researchers a great deal of choice as to how and where DNA is cut. Some of the most commonly used restriction enzymes and their recognition sites are indicated in table 27.1. Most of these enzymes recognize a sequence of either four or six contiguous base pairs. The recognition sites are situated so that a blunt-ended or staggered DNA results from the cleavage reaction. As a rule recognition sites are located on an axis of symmetry so freshly cleaved segments have identical structures at ends.

A viral genome cleaved exhaustively with a particular restriction enzyme usually yields several fragments; the restriction enzyme cleavage sites for the 5,300 bp of the SV40 virus genome are shown in figure 27.3. These fragments obtained after cleavage can be separated according to size by gel electrophoresis. Nondenaturing conditions are used so that the duplex strands stay together. The size of a fragment is, the slower it migrates on the gel. After electrophoresis for a time sufficient to separate the fragments, the gel is stained with a fluorescent dye such as ethidium bromide and viewed under long-wavelength ultraviolet light (long-wavelength UV is used because it does not excite the DNA). Individual fragments may be extracted from the gel for sequencing, PCR amplification, or cloning (described later on).

The problem of determining how a set of restriction fragments are normally connected is resolved by determining the sequences by a second set of fragments cut by a different restriction enzyme. The overlapping information obtained from the two sets of fragments permits a determination of the complete sequence of the intact genon.
Table 27.1
Recognition Sequences and Cutting Sites of Selected Restriction Enzymes

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Recognition Sequences</th>
<th>Enzyme</th>
<th>Recognition Sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>AluI</td>
<td>↓ AGCT↑ TCGA</td>
<td>HpaII</td>
<td>↓ CCGG↑ GGGC</td>
</tr>
<tr>
<td></td>
<td>↓ GGAATCC↑ CCTAGGG</td>
<td>Kpall</td>
<td>↓ GGTACCC↑ CCATGGG</td>
</tr>
<tr>
<td></td>
<td>↓ AGATCT↑ TCTAGA</td>
<td>Mbol</td>
<td>↓ GATC↑ CTAG</td>
</tr>
<tr>
<td></td>
<td>↓ ATCGAT↑ TAGCTA</td>
<td>PstI</td>
<td>↓ CTGCAG↑ GACGTC</td>
</tr>
<tr>
<td></td>
<td>↓ GAATTC↑ CTTAAG</td>
<td>Pvul</td>
<td>↓ CGATCG↑ GCTAGC</td>
</tr>
<tr>
<td></td>
<td>↓ GCCCG↑ CCGG</td>
<td>SalI</td>
<td>↓ GTGCAC↑ CAGCTG</td>
</tr>
<tr>
<td></td>
<td>↓ GTpGPuAC↑ CApGPuTG</td>
<td>SmaI</td>
<td>↓ CCCGGGG↑ GGGGCCC</td>
</tr>
<tr>
<td></td>
<td>↓ AAGCTT↑ TTCGAA</td>
<td>Xmal</td>
<td>↓ CCCGGGG↑ GGGGCCC</td>
</tr>
</tbody>
</table>

strategy of sequencing overlapping fragments is identical to that used in primary structure determination of proteins (see chapter 3).

Plasmids Are Used to Clone Small Pieces of DNA
In a simple procedure for "DNA cloning," an autonomously replicating plasmid and insert DNA are cut with a restriction enzyme and then the pieces are annealed and covalently joined by the action of DNA ligase. The resulting recombinant molecules are then transfected into E. coli, where they replicate. When plasmid vectors are used, a population of permeabilized cells is bathed in the plasmid DNA containing the inserted DNA. Because only a small number of cells become transfected by this procedure, a way to select cells that carry the desired hybrid plasmids is necessary.

A particularly useful plasmid vector for selective transfected cells called pBR322 is itself a hybrid plasmid (fig. 27.4). This plasmid contains two genes, amp' and which confer resistance to penicillin and tetracycline, respectively. PstI restriction fragments of foreign DNA be inserted into the unique PstI restriction site on pBR322 (see fig. 27.4). This is done by digesting pBR322 with PstI and mixing with the restriction fragments to be cloned at temperatures to permit annealing to take place between two DNAs, and finally ligating the annealed fragments with DNA ligase. The product contains some of the original pBR322 and some pBR322 with inserted foreign DNA. When this mixture is used in transfection, most cells are transfected, some are transfected with pBR322, and are transfected with the desired hybrid plasmid. The types of cells may be readily distinguished by their...
**Figure 27.4**
Structure of the pBR322 plasmid (a) and construction of a hybrid plasmid containing the pBR322 vector and a segment of foreign DNA (b). For pBR322 the unique sites for various restriction enzymes are indicated. Also indicated are the locations of the tetracycline (tet') and the ampicillin (amp') resistance genes and the origin for DNA replication. The hybrid plasmid is constructed by treating the plasmid and the foreign DNA with the PstI restriction enzyme and mixing the two DNAs together in the presence of DNA ligase.

![Diagram of plasmid structure and construction](image)

(a)

(b)

resistant properties. Normal cells do not grow in the presence of tetracycline or penicillin. Transfected cells with the DNA inserted in the plasmid are tetracycline-resistant but penicillin-sensitive, because the insert has disrupted the amp' gene. Cells containing the desired plasmids can be distinguished from those containing pBR322 by “replica plating” (fig. 27.5). The first step when using this is to spread a large population of treated bacteria on an agarose plate containing growth medium. With seemingly homogenous “lawn” of cells developing on the surface of the agarose. Actually, the lawn results in the growth of many microcolonies to the point of complete coverage. At this point a piece of velvet is lightly pressed to the surface of the plate, and this impression is transferred to fresh plates, which contain normal medium or medium supplemented with antibiotics. The results on the replica plates after overnight growth are indicated. The plate in normal medium again gives rise to a lawn of cells as virtually all transferred grow into colonies. The plates containing antibiotics give rise to a few colonies, each of which is derived from a cell that carries the plasmid-conferred drug resistance(s).
**Figure 27.6**

Electrophoretogram of restriction enzyme digests of pBR322 and pBR322 with a DNA insert at the PstI site. The insert is assumed to have no internal BamHI restriction sites. In channels A and B the pBR322 is predigested with PstI and BamHI, respectively. The resulting DNA migrates with the same mobility because the plasmid has one site for each of these enzymes and therefore has the same molecular weight. In C and D the hybrid plasmid containing a DNA insert is treated with BamHI and PstI, respectively. In C the hybrid plasmid has been linearized by one cut at the BamHI site in the tel gene. It runs more slowly than the pBR322 because it is larger than the DNA insert. In D the plasmid cuts at two PstI sites located between the pBR322 sequences and the insert sequences. Consequently, one segment migrates at the rate of a linearized pBR322 plasmid. The other segment, also linearized, migrates at a rate characteristic of the size of the DNA insert. The electrophoresis is run from left to right; fragments are stained with ethidium bromide and photographed with UV light.

**Figure 27.7**

The nutrient agar plate contains a continuous lawn of *E. coli* bacteria except for circular clearings that represent phage plaques. Each plaque was originally derived from a single phage infecting a single *E. coli* bacterium. After infection the phage multiplies, ultimately producing about 100 mature viruses. The phages also produce an enzyme that causes the harboring cell to lyse. When this happens the phages are released, and each of them infects a neighboring cell and goes through the same infectious cycle. The process continues. Each cycle takes about 30 min. Eventually a visible clearing can be seen on the plate.

**Bacteriophage λ Vectors Are Useful for Cloning Larger DNA Segments**

Bacteriophage λ possesses a number of advantages as a cloning vector. DNA fragments as large as 24 kb can be propagated using such vectors. The primary pool can be amplified by limited phage growth as plaques in the entire collection of phage clones (recognized as plaques) can be stored for long periods in a small volume.
Because it does not accommodate molecules of DNA that are much longer than the viral genome, the use of λ as a vector for cloning substantial DNA fragments requires the removal of a significant portion of the viral DNA beforehand. Fortunately, the central third of the genome contains genes that are not essential for phage production and can therefore be deleted.

Bacteriophage λ vectors that accommodate foreign DNA fragments generated by a variety of restriction endonucleases have been constructed. The recombinant DNA molecules that incorporate some of these vectors can be introduced directly into E. coli by transfection. Alternatively, recombinant DNA molecules can be packaged into phage particles and subsequently infected into suitable host cells.

Cosmids Are Used to Clone the Largest Segments of DNA

Although plasmids and bacteriophage λ are both highly useful vectors, the size of the DNA fragments that can be cloned in them is limited. With plasmids, the larger the fragment of foreign DNA inserted, the lower the efficiency of ligation and transfection, making the cloning of DNA fragments larger than 15 kb experimentally difficult. In λ vectors, the length of the nonessential region of λ DNA limits fragment size to 24 kb or less. Also, the original λ vectors do not allow propagation of viable bacterial cells that carry the inserted DNA fragment; the insert is propagated as part of a virus that lyases the cell.

Cosmids were developed as vectors for cloning large DNA fragments. The first part of their name, “cos,” comes from the fact that cosmids contain the cohesive ends, or cos sites of normal λ. These ends are essential for packaging the DNA into λ phage heads. The last part of their name, “mid,” indicates that cosmids carry a plasmid origin of replication like the one found in the pBR322 plasmid. Such cosmids can be used for cloning in the same way as any other plasmid vector. However, because cosmids also contain the cos sites, cosm id DNA along with an inserted DNA fragment can be packaged as a λ phage. The result after packaging is a defective but nevertheless infectious phage particle. Once the cosmid and the inserted DNA fragment are introduced by infection into a λ-sensitive cell, the plasmid replicates. Since cosmids lack the entire bacteriophage genome except for the region adjacent to the cos sites, these vectors can propagate exogenously derived DNA fragments of up to 40–50 kb in length.

Shuttle Vectors Can Be Cloned into One of Different Species

Vectors that include replication systems derived from more than one host species are known as shuttle vectors; vectors commonly include a replication system ablation in E. coli and one that works in a second host, which may be bacterial or eukaryotic. Initial cloning application of the DNA segment to be studied is often carried out in E. coli because it is easier to make large quantities of DNA. The recombinant DNA molecule consisting of the “bifunctional vector” plus the cloned DNA, is then introduced into the second host; its purpose is usually to measure the expression of the gene carried by the vector. Shuttle vectors that can re both E. coli and yeast are the most common.

Constructing a “Library”

Cloning can involve a single vector-linked fragment or a collection of independently isolated linked DNA fragments derived from a single organism. Such a collection is termed a “library” and is the source of well-defined sequences from a genome. Each clone of a library harbors a particular fragment from the desired organism. Within the entire sequence may be repeated, but other sequences are missing. The ideal library, which can only be assembled, represents all of the sequences with the smallest number of clones.

A library from the same cell or organism compared in two ways. The genome may be fragmented to the appropriate vector to produce a vector library. An alternative approach is to construct a library in which the DNA fragments to be cloned are by reverse transcription from the cellular RNA. These libraries have advantages and disadvantages, specific purpose, one library is usually preferred over the other.

The vast majority of DNAs within a library are characterized. As a rule, the task of finding all genes or sequences within a library greatly exceeds that of constructing the library.

A Genomic DNA Library Contains One with Different Genomic Fragments

A major concern in constructing a genomic DNA library is to maximize the probability that all segments of the genome are represented. If the genomic DNA is prepared...
Table 27.2

Theoretical Number of Clones Required to Fully Represent the Entire Genome of Various Organisms

<table>
<thead>
<tr>
<th>Size of Cloned DNA Fragment (bp)</th>
<th>Genome Size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$2 \times 10^6$</td>
</tr>
<tr>
<td></td>
<td>(e.g., bacteria)</td>
</tr>
<tr>
<td>$5 \times 10^3$</td>
<td>400</td>
</tr>
<tr>
<td>$10 \times 10^3$</td>
<td>200</td>
</tr>
<tr>
<td>$20 \times 10^3$</td>
<td>100</td>
</tr>
<tr>
<td>$40 \times 10^3$</td>
<td>50</td>
</tr>
</tbody>
</table>

with a restriction enzyme, an added concern is that the enzyme cleaves genes of interest at one or more sites. To increase the likelihood of isolating desired genes in one piece, different restriction enzymes can be used on different parallel preparations. But even if the genes of interest are not cut by the enzyme(s) chosen, the DNA fragments produced may be inconveniently small to work with. An enzyme that recognizes a sequence of six bases (a six-cutter) gives an average fragment size of 4,096 bp\(^b\), which is a reasonable size for making a plasmid library but much smaller than the size desirable for cloning in $\lambda$ or a cosmid vector. Therefore, when large, randomly generated fragments are desired, the method of choice usually entails making an incomplete digest with a four-cutter restriction enzyme, which produces overlapping ends that can be readily cloned into the chosen vector as described earlier. The extent of digestion is controlled so that cleavage occurs at only some of the restriction enzyme recognition sites and the average size of the fragments produced is in the desired range. The conditions used thus depend on whether the product is going to be cloned in a plasmid, a $\lambda$ phage, or a cosmid vector. Table 27.2 gives the minimum number of clones (i.e., the size of the library) required to fully represent the entire genome in a genomic DNA library, as a function of the average size of the cloned fragments and the size of the genome. Since DNA fragments in a population are cloned on a random basis, the chance of finding a given single-copy gene in a library of the indicated size is 50%. A clone bank should be 3 to 10 times the minimum size to give a high probability that a particular segment is represented.

\(^b\) $1/4,096$ = \(1/4,096\)

**A cDNA Library Contains Clones Reflecting the mRNA Sequences**

A cDNA library consists of a collection of certain DNA copies of the cellular or organismic RNA obtained from a differentiated monoculture, then the library varies in composition according to the type of cell used as the RNA source and to the state of the cell. This variation is a reflection of the abundances of particular mRNAs made by different types. If a cDNA species corresponding to a particular product is desired, it is often possible to select variants suspected to synthesize a large amount of the cDNA. mRNAs or mRNA-related protein. Thus, pituitary cells be used if cDNA encoding growth hormone is required, whereas liver cells can be used if a serum albumin is the goal. The mRNAs present in low amounts in a library may require the screening of a larger library than the present in medium or high abundance.

Once the crude mRNA fraction has been the chosen cells or tissue, it is converted to molecules with the help of reverse transcriptase. Duplex DNA does not have "sticky ends" for insertion into a vector. For this purpose DNA linkers are added as ends. Linkers are synthetic single-stranded oligonucleotides (6, 8, 10, or 12 bases in length) that can form symmetrical, blunt-ended, double-stranded structures containing the recognition sequence for a restriction enzyme. Figure 27.8 shows an example (CCTGGCAGG) containing a Psrl restriction enzyme recognition site, which self-associates to produce a double-stranded structure that adds to the double-stranded DNA and stabilizes the presence of T4 ligase. The resulting prod
Figure 27.8

Insertion of cDNA into pBR322 plasmid by the linker method. The strategy here is to open up the plasmid with a restriction enzyme that makes staggered cuts and to attach linkers that contain the same recognition site to the cDNA. After the linkers are attached to the cDNA, the duplex is treated with the same restriction enzyme (PstI) to expose the overhangs. The two DNAs are mixed together and ligated. After transfection, cells containing the hybrid plasmids are recognized by tetracycline resistance and ampicillin sensitivity. Identification of the insert is discussed in the text.

Numerous Approaches Can Be Used to Pick the Correct Clone from a Library

A library can contain thousands or even tens of thousands of different kinds of clones (see table 27.2), making it a challenging endeavor to isolate a clone with the DNA of interest. Most currently used procedures for screening large numbers of colonies for plasmids or phage that contain specific DNA inserts are variants of the colony hybridization method developed by Grunstein and Hogness. This procedure makes use of a specific radioactive probe that contains some sequences complementary to those in the DNA of interest. The colonies to be screened are first grown on agar petri plates (fig. 27.9). A replica of each plate is made on another agar plate, which is stored for reference. A replica is also made on a nitrocellulose filter. The colonies for the filter are lysed and the contents denatured seriously by treatment with sodium hydroxide. After the denatured DNA is fixed on the filter at each site colony was located. The DNA on the filter is hybridized with a radioactively labeled nucleic acid probe complementary to the specific DNA sequence to be selected. Presence of hybridized probe at sites occupied by colonies from colonies that include the DNA fragment of interest is detected by autoradiography. The colony whose probe hybridizes with the nucleic acid probe can then be selected from the reference plate, which contains a viable colony at a corresponding location.

Cloning in Systems Other than Escherichia coli

Despite the success and broad applications of E. coli cloning systems, instances occur in which gene products need to be made in this bacterium. Either they are not synth...
Figure 27.9

Colony hybridization procedure used to identify bacterial clones harboring a plasmid containing a specific DNA. Step 1: Replica-plate the colonies containing plasmids onto nitrocellulose paper. Step 2: Lys cells with NaOH and fix denatured DNA to paper. Step 3: Hybridize to \(^{32}\)P-labeled DNA carrying the desired sequence and autoradiograph the product. Locations of desired DNA should be emphasized in the autoradiograph. Clones carrying desired plasmids (circled) may then be isolated from a corresponding agar replica plate carrying untreated colonies.

Site-Directed Mutagenesis Form Restructuring of Existing Genes

By combining different procedures of molecular biology, it is now possible to make discrete changes within genes. This technique, called site-directed mutagenesis, is one of the most important in modern genetics and molecular biology. The first site-directed mutagenesis study was carried out by David Shortle and Daniel Nathans with the help of the mutagen sodium bisulfite, which converts C residues to T, so that they become converted to A residues.

Directed mutagenesis as it is practiced today involves the chemical synthesis of a deoxyoligonucleotide that contains discrete changes in its sequence from the wild-type sequence observed in the genome under investigation. This change may be single-base or multibase; they may involve changes, base deletions, or base additions.

Many variations of site-directed mutagenesis are possible. One can start out with a circular, single-stranded DNA molecule and anneal it to a synthetic primer DNA carrying changes (fig. 27.10). This primer can be extended on a DNA polymerase, and the resulting product can be transfected. Finally, clones of cells containing the plasmid with changes are screened using standard methods using cloning vectors. For this reason, the PCR product should contain restriction sites at suitable for cloning. Thus, PCR amplification need not be thought of as alternatives for purposes but as complementing each other to give a variety of approaches.

Recombinant DNA Techniques Used to Characterize the Globin Gene Family

The human globin family is a paradigm of differential gene activity during development. The molecular basis of genetic disorders in sickle cell anemia. Hemoglobin is a tetramer containing two \(\alpha\) and two \(\beta\) subunits.
**Figure 27.10**

Scheme for oligonucleotide-directed mutagenesis of double-stranded circular plasmid DNA. Supercoiled plasmid circles are nicked in one strand and rendered partially single-stranded by treatment with exonuclease. The gapped circles are hybridized with a homologous oligodeoxynucleotide carrying, by design, some mismatches. *In vitro* DNA synthesis, primed in part by the oligodeoxynucleotide, heteroduplex plasmid circles. (Source: After G. Dalbadie-McFar Cohen, A. D. Riggs, C. Morin, K. Itakura, and J. H. Richards, Olig directed mutagenesis as a general and powerful method for studies function, *Proc. Natl. Acad. Sci. USA* 79:6408–6412, 1982.)

![Figure 27.10 Diagram](image)

**Figure 27.11**

Illustration of a general method of mutagenesis using PCR. Primers are represented as short lines with arrowheads pointing in the 3′ direction. The bump in primers 2 and 3 and their products represent a mismatched base, a deliberate alteration in base sequence from that present in the starting DNA. Of the four major products resulting from step 3 only D is extendable by DNA polymerase.

1. PCR using primers 1 and 3 yields:

2. PCR using primers 2 and 4 yields:

3. Mix products from steps 1 and 2 above, denature, and anneal. This should result in a mixture of four products, A, B, C, and D.

4. Extend D with DNA polymerase and amplify with primers 1 and 4 to give final mutated DNA segment.

**DNA Sequence Differences Were Used to Detect Defective Hemoglobin Genes**

All of the hemoglobin genes are represented or more alleles within the human popula
gens are said to be polymorphic. This polymorphic quently shows up in readily detectable phenotypes differences occur in vital areas of the polypeptid Polymorphisms show up in the DNA even more fi
coding and the noncoding regions of a gene, and t differences can be detected even when no visible eff
parent in the organism. Because of their frequency of detection by recombinant DNA methods, DNA j phisms have become extremely useful in map human genome.
**Figure 27.12**
Changes in types of hemoglobin observed in early development. A single switch in gene expression is observed for α-like chains. Two switches in gene expression are observed for β-like chains. The corresponding tetrameric hemoglobin molecules observed at different stages in development are also indicated.

<table>
<thead>
<tr>
<th>Early embryo</th>
<th>8-week gestation</th>
<th>8-month gestation</th>
</tr>
</thead>
<tbody>
<tr>
<td>α-like chains</td>
<td>( \zeta )</td>
<td>( \alpha )</td>
</tr>
<tr>
<td>β-like chains</td>
<td>( \varepsilon )</td>
<td>( \gamma (\gamma_G \text{ and } \gamma_A) )</td>
</tr>
<tr>
<td>Tetramers</td>
<td>( \zeta \varepsilon \gamma_2 )</td>
<td>( \alpha_2 \varepsilon \gamma_2 \text{ or } \zeta \gamma_2 )</td>
</tr>
</tbody>
</table>

**Figure 27.13**
The chromosomal localization and genomic organization of the human globin genes. The \( \alpha \)- and \( \beta \)-globin gene complexes are positioned on chromosomes 16 and 11, respectively. For each complex, the arrangement of genes on the chromosome is depicted above, and the general structure of the major gene is shown below, together with the location of the intervening sequences, or introns (IVS), and codon numbers. Coding regions are shown by solid boxes and IVS regions by open boxes. Genes with the \( \phi \) symbol in front are called pseudogenes because they are sequence-related but not expressed.
**Figure 27.14**
The steps involved in assaying by Southern blotting. The DNA to be analyzed is digested with a restriction enzyme (1). The resulting fragments are electrophoresed on an agarose gel (2). The DNA fragments on the gel are transferred to a cellulose nitrate sheet by placing the cellulose nitrate sheet next to the gel and passing solvent through the gel into the sheet. Flow of the solvent is maintained by blotting the far side of the cellulose nitrate sheet with paper towels. The DNA, first denatured with alkali, flows with the solvent but gets stuck in the sheet (3). The sheet is hybridized to radioactively labeled DNA containing the gene sequence of interest (4). The hybridized sheet is autoradiographed to determine the location of the labeled restriction fragment on the gel (5).

Whereas DNA polymorphisms should be recognizable by sequence differences, it is usually more convenient to detect these polymorphisms by differences in the size of DNA fragments obtained with restriction enzymes. Differences observed in this way are called restriction fragment length polymorphisms (RFLPs).

Kan and Dozy were the first to discover an allele-linked DNA polymorphism in the globin genes. With it they predicted which fetuses carried normal and which carried abnormal sickle-cell genes for beta-globin. To analyze the DNA for these differences they used a technique called Southern blotting. The steps involved in Southern blotting are illustrated in figure 27.14. First the genomic DNA from the test subject is digested with a restriction enzyme to yield specific DNA fragments. These fragments are separated according to size by agarose gel electrophoresis. Next the DNA is denatured and transferred from the agarose cellulose nitrate sheet. The DNA firmly bound to the nitrate sheet is hybridized with a radioactively labeled DNA which carries some of the sequences of interest. The labeled DNA, which hybridizes to specific regions of the nitrate sheet, is detected by autoradiography. By comparing the retained labeled DNAs from DNA of different individuals one can see if the labeled DNAs move with the same or a different rate. If they move differently, there must be a restriction enzyme difference between the individuals. Detection of an RFLP means usually depends on the restriction enzyme used for initial digestion. Some enzymes show a difference do not.

To apply this technique to their hemoglobin genes, Kan and Dozy first had to prepare a DNA probe that hybridized to specific sequences in the beta-globin gene. This was
**Figure 27.15**

Inheritance pattern of an RFLP associated with sickle-cell disease. Humans carry two alleles for the same gene, and each offspring inherits one allele from each of its parents in an entirely random fashion. Normal individuals are homozygous for normal $Hb$ alleles; individuals with sickle-cell trait are heterozygous, with the one normal $Hb$ allele and one $Hb^s$ allele; and individuals with sickle-cell disease are homozygous for the $Hb^s$ allele. At the top (a) we see a three-generation pedigree analysis for a family that carries both the normal and the sickle-cell gene for $\beta$-globin. Males are represented by squares and females by circles. A purple circle or square indicates an individual who is homozygous normal. A half-filled circle or square (red/purple) indicates a heterozygous individual with sickle-cell trait. A filled circle or square (red) indicates a homozygous individual with sickle-cell disease. In (a) both sets of grandparents produce a heterozygous individual with sickle-cell trait. Because one of the grandparents is homozygous normal and the other is heterozygous, there is a 50% chance that the grandparent mating will give rise to a heterozygous offspring as shown and also a 50% chance that the offspring will have normal offspring (not shown). The two heterozygous parents have an increased chance of having abnormal offspring because in this mating each parent carries one abnormal gene $Hb^s$ allele. There is a 25% chance of a homozygous sickle-cell allele offspring, a 50% chance of an abnormal offspring with sickle-cell trait, and a 25% chance of a normal offspring. Below the pedigree chart is the electrophoretic pattern of a $HpaI$ digest probed with globin cDNA by the Southern blotting technique (b). At the bottom we see an interpretation of the normal and abnormal DNAs (c).

(a) Pedigree

(b) Gel patterns of $HpaI$ restricted DNA probed with $\beta$-globin sequences

(c) Origin of 13-kb and 7.6-kb DNA fragments

Southern blotting, Kan and Dozy showed that the $\beta$-globin gene was contained within a 7.6-kb $HpaI$ restriction fragment, whereas the $\beta^S$-globin gene of sickle-cell anemia, $Hb^s$, was contained within a 13-kb fragment (27.15). Further analysis showed that the RFLP was from a $HpaI$ restriction site 5 kb to the 3' side of...
globin gene, that was present in the normal case and absent in \( Hb^c \). Subsequent analysis showed sequence differences within the coding regions. One may wonder why the RFLP outside the coding region was so commonly associated with the abnormal gene. A possible explanation for this is that in the distant past a mutation occurred which resulted in the 7.6-kb type and a few 13-kb types before introduction of the sickle-cell gene mutation. After the \( Hb^c \) mutation was introduced into the 13-kb type, it became greatly expanded because of the selective advantages of this gene in heterozygotes. In this connection it should be noted that heterozygotes carrying one normal gene and one sickle-cell gene fare far better when infected by malaria. As a result in areas where malaria is prevalent the heterozygote has a selective advantage over the normal homozygote.

Knowledge of this and other polymorphisms has been used for pre- or postnatal diagnosis of the sickle-cell gene. Such information can be of great practical value in genetic counseling. Incidentally, it is now possible to diagnose sickle-cell disease (which occurs in individuals that are homozygous for the sickle-cell gene) with greater certainty because the point mutation leading to the defect in the coding region itself produces a recognizable RFLP.

**The \( \beta \)-Globin cDNA Probe Was Used to Characterize the Normal \( \beta \)-Globin Gene**

Detailed mapping with DNA probes was first successfully executed on the human \( \beta \)-globin gene. All members of a human genomic library that annealed to the radioactive cDNA probe for the \( \beta \)-globin gene were isolated, and each of these was sequenced. This analysis resulted in a complete description of the \( \beta \)-globin gene (see fig. 27.13). The \( \beta \)-globin gene is appreciably longer than the \( \beta \)-globin mRNA; in addition to containing regions that are present in the final mRNA the gene contains two “intervening” regions not represented in the mRNA sequences. We have more to say about the significance of these intervening sequences (IVS, also called introns) in the next chapter.

**Chromosome Walking Permitted Identification and Isolation of the Regions around the Adult \( \beta \)-Globin Genes**

The original cDNA probe carrying the \( \beta \)-globin mRNA sequences could only detect members in the genomic library that contained sequences homologous to those present in the probe. To explore the region flanking the \( \beta \)-globin gene the genomic library was probed further. For this purpose members of the genomic library that hybridized with the original cDNA were themselves converted into radioactive probes and these were used to locate additional members of the library that contained sequences flanking the \( \beta \)-globin gene. A cyclic repetition of this process resulted in a gradual expansion of sequences containing the \( \beta \)-globin gene. Using the library in this manner to extend the mapped region is known as chromosome walking (fig. 27.16). A parallel approach was used to extend the map around the adjacent \( \alpha \)-globin gene (see fig. 27.13). It can be seen that in both cases several genes occur in a cluster for each of the protein genes. Most of these can be correlated with the genes that are expressed at different times during development. In addition, genes, called pseudogenes, occur that have strong sequence similarities to known genes but are not expressed. It is not clear if these pseudogenes have a function or simply represent evolutionary “junk,” which has not been removed.

**Walking and Jumping Were Both Used Map the Cystic Fibrosis Gene**

By combining the linkage information obtained with RFLP mapping with other DNA manipulation techniques, it has been possible to locate genes causing some genetic disorders even when these genes are only known from their inheritance patterns. The list of serious disorders that can be linked to single genetic loci is growing, including Huntington’s disease, Duchenne’s muscular dystrophy, polycystic kidney disease, cystic fibrosis, chronic ulcerous disease, peripheral neurofibromatosis, familial polyposis coli, and juvenile endocrine neoplasia. One of the most spectacular achievements has been the determination of the gene causing cystic fibrosis (CF).

Cystic fibrosis is the most common serious genetic disorder in Caucasian populations. The major clinical symptoms include chronic pulmonary disease, pancreatic insufficiency, and an increase in the concentration of sweat electrolytes. Bearers of this disease are readily identified by a test done in childhood. The frequency of the disease is 1 in 2,000, from which it may be calculated that the carrier frequency is about 1 in 20 (the frequency of the heterozygote for a rare allele is twice the square root of the frequency of the homozygote). By classical genetic analysis the CF locus has been assigned to the long arm of chromosome 7 near the \( met \) locus. A map region containing the \( met \) locus and the CF gene is shown in figure 27.17.

In many genetic disorders, cystic fibrosis included, analysis begins before the responsible gene and its product are known. For example, it was initially hypothesized that the CF gene may be one of the genes that code for a receptor protein. Assays were developed to look for the presence of these proteins in cell membranes. The results were negative. However, subsequent work has shown that the CF gene codes for a transmembrane protein that is defective in CF lung cells.
**Figure 27.16**

The linkage map of the human β-globin gene locus as shown by the structural analysis of overlapping λ genomic clones. Both \( \lambda H\beta G1 \) and \( \lambda H\beta G3 \) clones contained the entire β-globin gene. Other clones detected by "walking" led to the discovery of other β-globin-like genes. These included four genes that are expressed and two pseudogenes that are not expressed. The genomic segments of clones isolated are shown together with the cleavage sites for enzyme EcoRI. The numbers on the top line indicate the size fragments in kilobase pairs.

\[
5' \rightarrow 3'
\]

\[
\begin{array}{cccccccccccc}
6.7 & 3.7 & 4.15 & 18.5 & 1 & 0.8 & 1.6 & 2.7 & 0.6 & 2.3 & 7.2 & 3.1 & 2.25 & 5.2 & 3.5 & 3.2 & 4.7
\end{array}
\]

\[\text{kb} \]

\[
\begin{array}{cccccccccccc}
60 & 50 & 40 & 30 & 20 & 10 & 0
\end{array}
\]

\[\lambda H\beta G1 \]

\[\lambda H\gamma G2 \]

\[\lambda H\gamma G5 \]

\[\lambda H\delta G4 \]

\[\lambda H\beta G1 \]

\[\lambda H\beta C \]

**Figure 27.17**

Map of restriction fragment length polymorphisms (RFLPs) closely linked to the cystic fibrosis (CF) gene. The inverted triangle near the right-hand end indicates the location of the \( \Delta F_{508} \) mutation characteristic of most persons with cystic fibrosis disease. (Source: Adapted from B.-S. Kerem, J. M. Rommens, J. A. Buchanan, D. Marki T. K. Cox, A. Chakravarti, M. Buchwald, and L.-C. Tsui, Identification cystic fibrosis gene: Genetic analysis, Science 245:1075, 1989.)

<table>
<thead>
<tr>
<th>Locus</th>
<th>MET</th>
<th>D7S122</th>
<th>D7S23</th>
<th>D7S399</th>
<th>CF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probe</td>
<td>metD</td>
<td>metH</td>
<td>E6</td>
<td>H2.3A</td>
<td>E4.1</td>
</tr>
<tr>
<td>RFLP</td>
<td>BanI</td>
<td>MspI</td>
<td>TaqI</td>
<td>(XV2c)</td>
<td>JG2E1</td>
</tr>
<tr>
<td></td>
<td>TaqI</td>
<td>HinfI</td>
<td>HindII</td>
<td>GtgII (HhaI)</td>
<td>(E.9)</td>
</tr>
<tr>
<td>kb</td>
<td>9</td>
<td>12.5</td>
<td>3.3</td>
<td>500</td>
<td>30</td>
</tr>
</tbody>
</table>

Product are known. Thus, one cannot locate the gene directly, as in the case of the β-globin gene and then determine its approximate location in subsequent analysis. For genes such as those responsible for cystic fibrosis the approximate location is determined by conventional genetic analysis, and then one attempts to close in on the gene by a process called "reverse genetics." Conventional genetic mapping by recombination frequency is not practical with human genes below map distances of 1 centimorgan because of the small number of test recombinants that are ordinarily available for observation. Unfortu na centimorgan on the human genome is equivalent

\(^{c}\) Genetic loci 1 cM apart recombine 1% of the time at meiosis.
Jumping and walking to find the cystic fibrosis gene. Following each jump, the locus defined by the jump was used as a starting point for a chromosome walk. Each DNA segment so found was hybridized to a sweat gland cDNA library until a match was found. It seemed likely that the sweat gland cDNA library would have a good representation of the cystic fibrosis gene transcript because the disease involves the sweat glands.

Predicted structure of the CFTR protein. Cylinders represent membrane-spanning helical segments. The cytoplasmically NBFs are shown as blue spheres with slots to indicate the entry of nucleotides. R represents the large polar domain, w linked to two halves of the protein molecule. Charged amines are shown as small circles with the charge sign. Net charge internal and external loop domains joining the membrane cylinders regions of the NBFs are contained in open squares. Potentially phosphorylation by protein kinases A or C (PKA or PKC); glycosylation (N-linked CHO) are indicated. (K = Lys; R = His; D = Asp; E = Glu.) (From J. R. Riordan et al., Identification of cystic fibrosis gene, Science 245:1066, Sep. 8, 1989. Copyright 19 AAAS. Reprinted by permission.)

Usually a jumping library and a walking library prepared from the same genome and used in co with each other. The region around each locus can be scrutinized by the walking library as shown in figure 27.18.

This complex analysis would be in vain if or have some criterion for knowing when the goal was reached. This is why from the physiological nature of the condition beneficial. In a brilliant strategy a cDNA library was prepared...
the mRNA fraction of sweat gland tissue. Recall that in cystic fibrosis sweat glands malfunction; therefore, the mRNA for the CF gene might be well represented in the mRNA fraction of the sweat gland cells.

While the walking and jumping process was in progress each new segment mapping in the general region of interest was tested against the sweat gland cDNA library. Finally, a member of the walking library was found that annealed with a member of the sweat gland cDNA library. Was this match fortuitous or did it mean the CF gene had been found? To answer this question the cDNA discovered in this way was used to probe the genomic library. By this means a gene was mapped that extended over a region of about 250 kb with 23 introns. A unique transcript, approximately 6,500 nucleotides in length was detected in extracts of sweat gland tissue which matched the transcript size expected from this gene. The protein predicted from a sequence analysis of this transcript consists of two similar motifs, each with (1) a domain having properties consis- with membrane association and (2) a domain believed involved in ATP binding (fig. 27.19). Finally it was derved that many CF patients carry a three base deletion in transcript which should result in the loss of a phenylalanine residue from the protein. This defect correlates with the notion that CF patients have a faulty membrane protein leads to the secretion problems characteristic of the disease. The fact that the abnormal gene is located as close as can tell by classical genetics to the CF locus adds additional support to the notion that the CF gene has been found.

Finding the disease gene does not, of course, that a cure is in the offing. However, the characterization of the disease gene will be a tremendous aid in diagnosing carriers and fetuses that are homozygous for the disease. It also should be a help in focusing approaches finding a cure for the disease.

---

Summary

1. Sequencing DNA uses chemical methods to cleave specific bases in preexisting DNA or carries out the synthesis under conditions where synthesis is interrupted at specific bases.

2. A specific segment of DNA can be synthesized in vitro by the polymerase chain reaction. Short segments of DNA bordering the segment of interest are added to a mixture containing the segment of interest, a DNA polymerase, and the deoxyribonucleotide substrates. The DNAs are first denatured, then annealed, and then synthesized. This cycle is repeated 20 or more times by raising the temperature to stop synthesis and lowering the temperature for annealing and synthesis. The outcome is a mixture in which the vast majority of the DNA is newly synthesized DNA bounded by the sequences of the added primers.

3. Another procedure for amplification cuts DNA containing the segment of interest into small pieces with a restriction enzyme. The cut pieces are incorporated into a plasmid or virus "vector" to be amplified in a suitable host. After growth, the mixture is plated to produce a mixture of bacterial or viral clones. The clone or clones of interest are identified often by hybridization of the clones after replica plating with a radioactive probe, followed by autoradiography to find the clone of interest.

4. Most cloning has been done in E. coli. Yeast is the most used eukaryotic host. Cloning is also possible a number of plant and animal cells.

5. Mapping with recombinant DNA probes was first applied to the human globin genes. Starting probes were obtained by isolating the globin messenger from reticulocytes and converting it into a cDNA, which was used to scan a human genomic library for cDNA probes complementary to mRNA hybridizing sequences. Once detected and purified, these cross-hybridizing members carrying globin messenger sequences were themselves converted to radioactive probes and used to further scan the genomic library for nearby sequences. By repeating this cycle several times, a process known as chromosome walking, revealed a region around the adult hemoglobin gene that contained several closely related genes associated with hemoglobin.

6. Frequently, alleles of the same gene can be distinguished by restriction site differences in the genes themselves or in nearby locations. Alleles identify this way are said to show restriction fragment length polymorphism. The allele responsible for sickle-cell disease was identified in this way.

7. The cystic fibrosis gene has been mapped by chromosome walking and jumping, a newer approach in which the relevant probes contain segments of the
genome that are normally located about 500 kbp from one another. A cDNA library was made from normal sweat gland tissue, chosen because of the disease’s association with abnormal release of sweat salt suggested that the sweat gland would contain an abundance of the messenger associated with the gene. By hybridizing the genomic DNA probes with the cDNA sweat gland library, a segment of genome was identified as a candidate for the cystic fibrosis gene. The gene was characterized in detail and found to be a complex transmembrane protein that carries a specific amino acid change in over half of the patients with cystic fibrosis. This correlation is overwh the gene responsible for cystic fibrosis has been mapped and characterized.

Selected Readings


Watson, J. D., M. Gilman, J. Witkowski, and M. Zol, combinator DNA, 2d ed. Scientific American Box York: W. H. Freeman Company, 1992. This text contains a general elementary text on the subject of recombinant DNA. It contains many exciting chapters on applications and is extremely well referenced.

Problems

1. Read the rest of the sequence in the autoradiogram in figure 27.1d as far as possible.
2. What are the major advantages of the polymerase chain reaction (PCR) method for amplifying defined segments of DNA as opposed to the use of conventional cloning methods? How might the PCR method be used to test for infection with the AIDS virus and how would this be an improvement over the body test currently used? (The current ELISA is an indirect test for the presence of antibody to the HIV proteins.)
3. Calculate the frequency of occurrence of restriction sites for Pst I and HindIII in the DNA from mophile (80% G + C) and from E. coli (52%
4. You just isolated a novel recombinant clone and purified the desired insert (a 10,000 bp linear duplex DNA) from the vector. Now you wish to map the recognition sequences for restriction endonucleases A and B. You cleave the DNA with these enzymes and fractionate the digestion products according to size by agarose gel electrophoresis. Comparison of the pattern of DNA fragments with marker DNAs of known sizes yields the following results:
(a) Digestion with A alone gives two fragments, of lengths 3,000 and 7,000 bp.
(b) Digestion with B alone generates three fragments, of lengths 500, 1,000, and 8,500 bp.
(c) Digestion with A and B together gives four fragments, of lengths 500, 1,000, 2,000 and 6,500 bp.
Draw a restriction map of the insert, showing the relative positions of the cleavage sites with respect to one another.

5. Draw the ends of a DNA fragment digested with the restriction endonuclease BamHI. How do these ends differ from those generated by MboI? If MboI and BamHI ends were to be ligated together, would the resulting junction be cleavable by BamHI or MboI?

6. Describe a procedure for cloning a DNA fragment into the BamHI site of pBR322.

7. How large a genomic library should you construct in order to detect and isolate a 15-kb gene out of a genome containing $3 \times 10^9$ bp?

8. If you were interested in isolating a cDNA for human serum albumin, why would you use a cDNA library established from mRNA isolated from liver? If you wanted to isolate the gene for albumin, why would you use a genomic library established from any human tissue?

9. Which of the E. coli vectors on the left (a, b, c) would be used to achieve the cloning objectives on the right (1-5)?
(a) Plasmid (1) Genomic library
(b) Cosmid (2) DNA sequencing
(c) Lambda (3) cDNA library
   (4) Small inserts
   (5) Genomic walking

10. Site-directed mutagenesis is one of the most powerful tools available to the biochemist. What are some of the applications of this technique? How can the PCR method be used to do site-directed mutagenesis and what is the advantage of this method?

11. The Southern blot technique is often used to compare genes from different organisms. For example, one could use the human globin gene probe described in the text to determine the extent of homology between globin genes from different primates. How could one reduce the stringency of the hybridization conditions (step 4 of fig. 27.14) to permit such a “heterologous hybridization”?

12. An unusual feature of the sickle-cell variant of the β-globin gene is that it directly alters a cleavage site for restriction endonuclease MspII. MspII recognizes the sequence CCTGAGG, which is mutated to CCTGTGG in the sickle-cell gene. How would you use this information and the Southern blot method to analyze fetal cells in amniotic fluid to determine whether the fetus carries sickle-cell anemia? What problems might you encounter in using this method?

13. Describe the procedure called “chromosome jumping.” How was this procedure used to map the cystic fibrosis gene?

14. Describe a procedure using the PCR technique that could be used to determine whether a normal individual is a carrier of the cystic fibrosis ΔF508 mutation. What problems could you anticipate with this method?