CHAPTER 3

Recombinant DNA Technology

DNA Isolation and Purification
Electrophoresis Separates DNA Fragments by Size
Restriction Enzymes Cut DNA; Ligase Joins DNA
Methods of Detection for Nucleic Acids
  Radioactive Labeling of Nucleic Acids and Autoradiography
  Fluorescence Detection of Nucleic Acids
  Chemical Tagging with Biotin or Digoxigenin
Complementary Strands Melt Apart and Reanneal
Hybridization of DNA or RNA in Southern and Northern Blots
Fluorescence in Situ Hybridization (FISH)
General Properties of Cloning Vectors
  Useful Traits for Cloning Vectors
Specific Types of Cloning Vectors
Getting Cloned Genes into Bacteria by Transformation
Constructing a Library of Genes
Screening the Library of Genes by Hybridization
Eukaryotic Expression Libraries
Features of Expression Vectors
Subtractive Hybridization
DNA ISOLATION AND PURIFICATION

Basic to all biotechnology research is the ability to manipulate DNA. First and foremost for recombinant DNA work, researchers need a method to isolate DNA from different organisms. Isolating DNA from bacteria is the easiest procedure because bacterial cells have little structure beyond the cell wall and cell membrane. Bacteria such as *E. coli* are the preferred organisms for manipulating any type of gene because of the ease at which DNA can be isolated. *E. coli* maintain both genomic and plasmid DNA within the cell. Genomic DNA is much larger than plasmid DNA, allowing the two different forms to be separated by size.

To release the DNA from a cell, the cell membrane must be destroyed. For bacteria, an enzyme called lysozyme digests the peptidoglycan, which is the main component of the cell wall. Next, a detergent bursts the cell membranes by disrupting the lipid bilayer. For other organisms, bursting the cells depends on their architecture. Tissue samples from animals and plants have to be ground up to release the intracellular components. Plant cells are mechanically sheared in a blender to break up the tough cell walls, and then the wall tissue is digested with enzymes that break the long polymers into monomers. DNA from the tail tip of a mouse is isolated after enzymes degrade the connective tissue. Every organism or tissue needs slight variations in the procedure for releasing intracellular components.

Once released, the intracellular components are separated from the remains of the outer structures by either centrifugation or chemical extraction. Centrifugation separates components according to size, because heavier or larger molecules sediment at a faster rate than smaller molecules. For example, after the cell wall has been digested, its fragments are smaller than the large DNA molecules. Centrifugation causes the DNA to form a pellet, but the soluble cell wall fragments stay in solution. Chemical extraction uses the properties of phenol to remove unwanted proteins from the DNA. Phenol is an acid that dissolves 60% to 70% of all living matter, especially proteins. Phenol is poorly water soluble, and when it is mixed with an aqueous sample of DNA and protein, the two phases separate, much like oil and water. The protein dissolves in the phenol layer and the nucleic acids in the aqueous layer. The two phases are separated by centrifugation, and the aqueous DNA layer is removed from the phenol.

Once the proteins are removed, the sample still contains RNA along with the DNA. Because this is also a nucleic acid, it is not soluble in phenol. Luckily, the enzyme ribonuclease (RNase) digests RNA into ribonucleotides. Ribonuclease treatment leaves a sample of DNA in a solution containing short pieces of RNA and ribonucleotides. When an equal volume of alcohol is added, the extremely large DNA falls out of the aqueous phase and is isolated by centrifugation. The smaller ribonucleotides stay soluble. The DNA is ready for use in various experiments.

**ELECTROPHORESIS SEPARATES DNA FRAGMENTS BY SIZE**

Gel electrophoresis followed by staining with ethidium bromide is used to separate DNA fragments by size (Fig. 3.1). The gel of gel electrophoresis consists of agarose, a polysaccharide extracted from seaweed that behaves like gelatin. Agarose is a powder that dissolves in water only when heated. After the solution cools, the agarose hardens. For visualizing DNA, agarose is formed into a rectangular slab about 1/4 inch thick. Inserting a comb at one end of the slab before it hardens makes small wells or holes. After the gel solidifies, the comb is removed, leaving small wells at one end.

Gel electrophoresis uses electric current to separate DNA molecules by size. The agarose slab is immersed in a buffer-filled tank that has a positive electrode at one end and a negative
electrode at the other. DNA samples are loaded into the wells, and when an electrical field is applied, the DNA migrates through the gel. The phosphate backbone of DNA is negatively charged so it moves away from the negative electrode and toward the positive electrode. Polymerized agarose acts as a sieve with small holes between the tangled chains of agarose. The DNA must migrate through these gaps. Agarose separates the DNA by size because larger pieces of DNA are slowed down more by the agarose.

To visualize the DNA, the agarose gel is removed from the tank and immersed into a solution of ethidium bromide. This dye intercalates between the bases of DNA or RNA, although less dye binds to RNA because it is single-stranded. When the gel is exposed to ultraviolet light, it fluoresces bright orange. DNA molecules of the same size usually form a tight band, and the size can be determined by comparing to a set of molecular weight standards run in a different well. Because the standards are of known size, the experimental DNA fragment can be compared directly.

The size of DNA being examined affects what type of gel is used. The standard is agarose, but for very small pieces of DNA, from 50 to 1000 base pairs, polyacrylamide gels are used instead. These gels are able to resolve DNA fragments that vary by only one base pair and are essential to sequencing DNA (see Chapter 4). For very large DNA fragments (10 kilobases to 10 megabases), agarose is used, but the current is alternated at two different angles. Pulsed field gel electrophoresis (PFGE), as this is called, allows very large pieces of DNA to migrate further than if the current only flows in one direction. Each change in direction loosens large pieces of DNA that are stuck inside agarose pores, letting them migrate further. Finally, gradient gel electrophoresis can be used to resolve fragments that are very close in size. A concentration gradient of acrylamide, buffer, or electrolyte can reduce compression (i.e., crowding of similar sized fragments) due to secondary structure and/or slow the smaller fragments at the lower end of the gel.

**FIGURE 3.1**

Electrophoresis of DNA

(A) Photo of electrophoresis supplies. Electrophoresis chamber holds an agarose gel in the center portion, and the rest of the tank is filled with buffer solution. The red and black leads are then attached to an electrical source. FisherBiotec Horizontal Electrophoresis Systems, Midigel System; Standard; 13 × 16-cm gel size; 800-mL buffer volume; Model No. FB-SB-1316.

(B) Agarose gel separation of DNA. To visualize DNA, the agarose gel containing the separated DNA fragments is soaked in a solution of ethidium bromide, which intercalates between the bases of DNA. Under UV light, the DNA bands fluoresce a bright orange color. The size of the fragments can be calculated by comparing them to the standards on the right.
RESTRICTION ENZYMES CUT DNA; LIGASE JOINS DNA

The ability to isolate, separate, and visualize DNA fragments would be useless unless some method was available to cut the DNA into fragments of different sizes. In fact, naturally occurring restriction enzymes or restriction endonucleases are the key to making DNA fragments. These bacterial enzymes bind to specific recognition sites on DNA and cut the backbone of both strands. They evolved to protect bacteria from foreign DNA, such as from viral invaders. The enzymes do not cut their own cell’s DNA because they are methylation sensitive, that is, if the recognition sequence is methylated, then the restriction enzyme cannot bind. Bacteria produce modification enzymes that recognize the same sequence as the corresponding restriction enzyme. These methylate each recognition site in the bacterial genome. Therefore, the bacteria can make the restriction enzyme without endangering their own DNA.

Restriction enzymes have been exploited to cut DNA at specific sites, since each restriction enzyme has a particular recognition sequence. Differences in cleavage site determine the type of restriction enzyme. Type I restriction enzymes cut the DNA strand 1000 or more base pairs from the recognition sequence. Type II restriction enzymes cut in the middle of the recognition sequence and are the most useful for genetic engineering. Type II restriction enzymes can either cut both strands of the double helix at the same point, leaving blunt ends, or they can cut at different sites on each strand leaving single-stranded ends, sometimes called sticky ends (Fig. 3.2). The recognition sequences of Type II restriction enzymes are usually inverted repeats, so that the enzyme cuts between the same bases on both strands. Since the repeats are inverted, the cuts may be staggered, thus generating single-stranded overhangs. Some commonly used restriction enzymes for biotechnology applications are listed in Table 3.1.

The number of base pairs in the recognition sequence determines the likelihood of cutting. Finding a particular sequence of four nucleotides is much more likely than finding a six base-pair recognition sequence. So to generate fewer, longer fragments, restriction enzymes with six or more base-pair recognition sequences are used. Conversely, four base-pair enzymes give more, shorter fragments from the same original segment of DNA.

When two different DNA samples are cut with the same sticky-end restriction enzyme, all the fragments will have identical overhangs. This allows DNA fragments from two sources (e.g., two different organisms) to be linked together (Fig. 3.3). Fragments are linked or ligated using DNA ligase, the same enzyme that ligates the Okazaki fragments during replication (see Chapter 4). The most common ligase used is actually from T4 bacteriophage. Ligase catalyzes linkage between the 3’-OH of one strand and the 5’-PO₄ of the other DNA strand. Ligase is much more efficient with overhanging sticky ends, but can also link blunt ends much more slowly. (Specific fragments for ligation are often isolated by agarose gel electrophoresis as described earlier.)

Restriction enzymes are naturally occurring enzymes that recognize a particular DNA sequence and cut the phosphate backbone. When two pieces of DNA are cut by the same restriction enzyme, the two ends have compatible overhangs that can be reconnected by ligase.


### Table 3.1 Table of Common Restriction Enzymes

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Source Organism</th>
<th>Recognition Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>HpaII</td>
<td><em>Haemophilus parainfluenzae</em></td>
<td>C/C GG GGC/C</td>
</tr>
<tr>
<td>MboI</td>
<td><em>Moraxella bovis</em></td>
<td>/GATC GATC/</td>
</tr>
<tr>
<td>NdeII</td>
<td><em>Neisseria denitrificans</em></td>
<td>/GATC GATC/</td>
</tr>
<tr>
<td>EcoRI</td>
<td><em>Escherichia coli</em> <em>RY13</em></td>
<td>G/AATTC CTTAA/G</td>
</tr>
<tr>
<td>EcoRII</td>
<td><em>Escherichia coli</em> <em>RY13</em></td>
<td>/CCWGG GGWCC/</td>
</tr>
<tr>
<td>EcoRV</td>
<td><em>Escherichia coli</em> <em>J62/pGL74</em></td>
<td>GAT/ATC CTA/TAG</td>
</tr>
<tr>
<td>BamHI</td>
<td><em>Bacillus amyloliquefaciens</em></td>
<td>G/GATCC CCTAG/G</td>
</tr>
<tr>
<td>Saull</td>
<td><em>Staphylococcus aureus</em></td>
<td>CC/TNAGG GGANT/CC</td>
</tr>
<tr>
<td>BglII</td>
<td><em>Bacillus globigii</em></td>
<td>GCCNNNN/NGGC CGGN/NNNNCCG</td>
</tr>
<tr>
<td>NotI</td>
<td><em>Nocardioid titidis-caviarum</em></td>
<td>GC/GGCCGC CGCCGG/CG</td>
</tr>
<tr>
<td>DraII</td>
<td><em>Deinococcus radiophilus</em></td>
<td>RG/GNCCY YCCNG/GR</td>
</tr>
</tbody>
</table>

\(, \) position where enzyme cuts.

N, any base; R, any purine; Y, any pyrimidine; W, A or T.

### METHODS OF DETECTION FOR NUCLEIC ACIDS

Recombinant DNA methodologies require the ability to detect DNA. One of the easiest ways to detect the amount of DNA or RNA in solution is to measure the absorbance of ultraviolet light (Fig. 3.4). DNA absorbs ultraviolet light because of the ring structures in the bases. Single-stranded RNA and free nucleotides also absorb ultraviolet light. In fact, they absorb more light because their structures are looser. The amount of absorption is compared with a known set of standards and the concentration of DNA can be determined.

### Box 3.1 Restriction Fragment Length Polymorphisms Identify Individuals

Restriction enzymes are useful for many different applications. Because the DNA sequence is different in each organism, the pattern of restriction sites will also be different. The source of isolated DNA can be identified by this pattern. If genomic DNA is isolated from one organism and cut with one particular restriction enzyme, a specific set of fragments can be separated and identified by electrophoresis. If DNA from a different organism is cut by the same restriction enzyme, a different set of fragments will be generated. This technique can be applied to DNA from two individuals from the same species. Although the DNA sequence differences will be small, restriction enzymes can be used to identify these differences. If the sequence difference falls in a restriction enzyme recognition site, it gives a restriction fragment length polymorphism (RFLP) (Fig. A). When the restriction enzyme patterns are compared, the number and size of one or two fragments will be affected for each base difference that affects a cut site.

(Continued)
Box 3.1 Restriction Fragment Length Polymorphisms Identify Individuals—cont’d

**FIGURE A** RFLP Analysis

DNA from related organisms shows small differences in sequence that cause changes in restriction sites. In the example shown, cutting a segment of DNA from the first organism yields six fragments of different sizes (labeled a–f on the gel). If the equivalent region of DNA from a related organism were digested with the same enzyme, a similar pattern would be expected. Here a single-nucleotide difference is present, which eliminates one of the restriction sites. Consequently, digesting this DNA produces only five fragments. Fragments c and d are no longer seen, but form a new band labeled cd.
Radioactive Labeling of Nucleic Acids and Autoradiography

Ultraviolet light absorption is a general method for detecting DNA, but does not distinguish between different DNA molecules. DNA can also be detected with radioactive isotopes.

The concentration of DNA in a liquid can be determined by measuring the absorbance of UV light at 260 nm.

Radioactive Labeling of Nucleic Acids and Autoradiography

Ultraviolet light absorption is a general method for detecting DNA, but does not distinguish between different DNA molecules. DNA can also be detected with radioactive isotopes.
During replication, radioactive precursors such as $^{32}$P in the form of a phosphate group and $^{35}$S in the form of phosphorothioate can be incorporated. Because native DNA does not contain sulfur atoms, one of the oxygen atoms of a phosphate group is replaced with sulfur to make phosphorothioate. Most radioactive molecules used in laboratories are short lived. $^{32}$P has a half-life of 14 days and $^{35}$S has a half-life of 68 days, so the isotopes degrade fairly fast. Although radioactive DNA is invisible, photographic film will turn black when exposed to the radioactive DNA. Radioactively labeled DNA is considered “hot,” whereas unlabeled DNA is considered “cold.”

The radioactive nucleotide precursors can be supplied to rapidly growing bacterial cultures. During replication, the radioactive precursor is incorporated into new DNA (see Chapter 4). The DNA is isolated from the bacteria and run on a gel. Autoradiography identifies the location of radioactively labeled DNA in the gel (Fig. 3.6). If the gel is thin, like most polyacrylamide gels, it is dried with heat and vacuum. If the gel is thick, like agarose gels, the DNA is transferred to a nylon membrane using capillary action (see Fig. 3.9, later). The dried gel or nylon membrane is placed next to photographic film. As the radioactive phosphate decays, the radiation turns the photographic film black. Only the areas next to radioactive DNA will have black spots or bands. The use of film detects where the hot DNA is on a gel, and the use of ethidium bromide shows where all of the DNA, hot or cold, is. These two methods allow distinguishing one DNA fragment from another.
Radioactive DNA can also be detected using **scintillation counting**. Here a small sample of the radioactive DNA is mixed with scintillation fluid. When the radioactive isotope decays, it emits a beta particle. Scintillation fluid emits a flash of light when excited by the beta particle. The scintillation counter detects light flashes with a photocell, counting them over a specified amount of time. Radioactive DNA concentrations can be determined by comparing to a set of known standards. Scintillation counting cannot detect the unlabeled or cold DNA, nor can it distinguish between multiple fragments of hot DNA, because it merely measures the total radioactivity in the sample.

Radioactive isotopes are incorporated into the DNA backbone during replication. Autoradiography or scintillation counting identifies the radioactive label.
**Fluorescence Detection of Nucleic Acids**

Autoradiography has its merits, but working with and disposing of radioactive waste is costly, both monetarily and environmentally. Using fluorescently tagged nucleotides was developed as a better method of DNA detection (Fig. 3.7). Fluorescent tags absorb light of one wavelength, which excites the atoms, increasing the energy state of the tag. This excited state releases a photon of light at a different (longer) wavelength and returns to the ground state. The emitted photon is detected with a photodetector. There are many different fluorescent tags, and each emits a different wavelength of light. Some photodetector systems are sensitive enough to distinguish between these different tags; therefore, if different bases have different fluorescent labels, the photodetector can determine which base is present. This is the basis for most modern DNA sequencing machines (see Chapter 4).

**Chemical Tagging with Biotin or Digoxigenin**

Biotin is a vitamin and digoxigenin is a steroid from the foxglove plant. Using these two chemicals allows scientists to label DNA without radioactivity or costly photodetectors. Biotin or digoxigenin are chemically linked to uracil; therefore, DNA must be synthesized with the labeled uracil replacing thymine. The DNA is synthesized *in vitro* as described in Chapter 4. A single-stranded DNA template, DNA polymerase, a short DNA primer, and a mixture of dATP, dGTP, dCTP, plus dUTP linked to either biotin or digoxigenin are combined. DNA polymerase synthesizes the complementary strand to the template, incorporating biotin- or digoxigenin-linked uracil opposite all the adenines.

The labeled DNA is visualized in a two-step process (Fig. 3.8). First, for biotin, a molecule of *avidin* binds to the tag. For digoxigenin, a specific antibody binds to the tag. Both avidin and the digoxigenin antibody are conjugated to *alkaline phosphatase*, an enzyme that

---

**FIGURE 3.7 Fluorescent Labeling of DNA**

(A) Fluorescent tagging of DNA. During synthesis, a nucleotide linked to a fluorescent tag is incorporated at the 3' end of the DNA. A beam of light excites the fluorescent tag, which in turn, releases light of a longer wavelength. (B) Energy levels in fluorescence. The fluorescent molecule attached to the DNA has three different energy levels, $S_0$, $S_1'$, and $S_1$. The $S_0$ or ground state is the state before exposure to light. When the fluorescent molecule is exposed to a light photon, the fluorescent tag absorbs the energy and enters the first excited state, $S_1'$. Between $S_1'$ and $S_1$, the fluorescent tag relaxes slightly, but doesn’t emit any light. Eventually the high-energy state releases its excess energy by emitting a longer wavelength photon. This release of fluorescence returns the molecule to the ground state.
removes phosphates from a variety of substrates. Several different chromogenic molecules act as substrates for alkaline phosphatase, but the most widely used one is X-Phos. Once alkaline phosphatase removes the phosphate group from X-Phos, the intermediate molecule reacts with oxygen and forms a blue precipitate. This blue color reveals the location of the labeled DNA. Another substrate of alkaline phosphatase is Lumi-Phos, which is chemiluminescent and emits visible light when the phosphate is removed. Much like autoradiography, when photographic film is placed over labeled DNA treated with Lumi-Phos, dark bands form wherever the Lumi-Phos glows.

**COMPLEMENTARY STRANDS MELT APART AND REANNEAL**

The complementary antiparallel strands of DNA form an elegant molecule that is able to unzip or melt and come back together or reanneal (Fig. 3.9). The hydrogen bonds that hold the two halves together are relatively weak. Heating a sample of DNA will dissolve the hydrogen bonds, resulting in two complementary single strands. If the same sample of DNA is slowly cooled, the two strands will reanneal so that G matches with C and A matches with T, as before.

The proportion of G-C base pairs affects how much heat is required to melt a double helix of DNA. G-C base pairs have three hydrogen bonds to melt, whereas A-T base pairs have only two. Consequently, DNA with a higher percentage of GC will require more energy to melt than DNA with fewer GC base pairs. The **GC ratio** is defined as follows:

\[
\text{GC ratio} = \frac{G + C}{A + G + C + T} \times 100\%
\]

The ability to zip and unzip DNA is crucial to cellular function, and has also been exploited in biotechnology. Replication (see Chapter 4) and transcription (see Chapter 2)
rely on strand separation to generate either new DNA or RNA strands, respectively. In molecular biology research, many techniques, from PCR to library screening, exploit the complementary nature of DNA strands.

**HYBRIDIZATION OF DNA OR RNA IN SOUTHERN AND NORTHERN BLOTS**

If two different double helixes of DNA are melted, the single strands can be mixed together before cooling and reannealing. If the two original DNA molecules have similar sequences, a single strand from one may pair with the opposite strand from the other DNA molecule. This is known as **hybridization** and can be used to determine whether sequences in two separate samples of DNA or RNA are related. In hybridization experiments, the term **probe molecule** refers to a known DNA sequence or gene that is used to screen the experimental sample or **target DNA** for similar sequences.

**Southern blots** are used to determine how closely DNA from one source is related to a DNA sequence from another source. The technique involves forming hybrid DNA molecules by mixing DNA from the two sources. A Southern blot has two components, the probe sequence (e.g., a known gene of interest from one organism) and the target DNA (often from a different organism). A typical Southern blot begins by isolating the target DNA from one organism, digesting it with a restriction enzyme that gives fragments from about 500 to 10,000 base pairs in length, and separating these fragments by electrophoresis. The separated fragments will be double-stranded, but if the gel is incubated in a strong acid, the DNA separates into single strands. Using capillary action, the single strands can be transferred to a membrane as shown in **Fig. 3.10**. The DNA remains single-stranded once attached to the membrane.

Next, the probe is prepared. First, the known sequence or gene must be isolated and labeled in some way (see earlier discussion). Identifying genes has become easier now that many genomes have been entirely sequenced. For example, a scientist can easily obtain a copy of a human gene for use as a probe to find similar genes in other organisms. Alternatively, using sequence data, a unique oligonucleotide probe can be designed that only recognizes the gene of interest (see Chapter 4). If an oligonucleotide has a common sequence, it will bind to many other sequences. Therefore, oligonucleotide probes must be long enough to have sequences that bind to only one (or very few) specific site(s) in the target genome. To prepare DNA probes for a Southern blot, they are labeled using radioactivity, biotin, or digoxigenin (see earlier discussion). Finally, the labeled DNA is denatured at high temperature to make it single-stranded. (Synthetic oligonucleotides do not require treatment, as they are already single-stranded.)

To perform the Southern blot, the single-stranded probe is incubated with the membrane carrying the single-stranded target DNA (Fig. 3.11). These are incubated at a temperature that allows...
hybrid DNA strands to form but allows a low amount of mismatching. The temperature, and hence the level of mismatching tolerated, can be varied depending on how specific a search is being run. If the probe is radioactive, then the membrane is exposed to photographic film. If the probe is labeled with biotin or digoxigenin, the membrane may be treated with chemiluminescent substrate to detect the labeled probe, and then exposed to photographic film. Dark bands on the film reveal the positions of fragments with similar sequence to the probe. Alternatively, biotin or digoxigenin labels may be visualized by treatment with a chromogenic substrate. In this case blue bands will form directly on the membrane at the position of the related sequences.

Northern blots are also based on nucleic acid hybridization. The difference is that RNA is the target in a Northern blot. The probe for a Northern blot is either a fragment of a gene or a unique oligonucleotide just as in a Southern blot. The target RNA is usually messenger RNA. In eukaryotes, screening mRNA is more efficient because genomic DNA has a lot of introns, which may interfere with probes binding to the correct sequence. Besides, mRNA is already single-stranded, so the agarose gel does not have to be treated with strong acid. Much like a Southern blot, Northern blots begin by separating mRNA by size using electrophoresis. The mRNA is transferred to a nylon membrane and incubated with a single-stranded labeled probe. As before, the probe can be labeled with biotin, digoxigenin, or radioactivity. The membrane is processed and exposed to film or chromogenic substrate.

A variation of these hybridization techniques is the dot blot (Fig. 3.12). Here the target sample is not separated by size. The DNA or mRNA target is simply attached to the nylon membrane as a small dot. As in Southern blots, the DNA sample must be made single-stranded before it is attached to the membrane. As before, the dot-blot membrane is allowed to hybridize with a labeled probe. The membranes are processed and exposed to film. If the dot of DNA or mRNA contains a sequence similar to the probe, the film will turn black in that area. Dot blots are a quick and easy way to determine if the target sample has a related sequence, before more detailed analysis by Southern or Northern blotting. Another advantage of dot blots is that multiple samples can be processed in a smaller amount of space.

Southern blots form hybrid DNA molecules to determine if a sample of DNA has a homologous sequence to another DNA probe.

Northern blots determine if a sample of mRNA has a homologous sequence to a DNA probe. In large genomes, using mRNA is more efficient because all the introns are removed.
Recombinant DNA Technology

**Box 3.2 Zoo Blot Compare Sequences among Different Species**

Southern blots are used for a variety of experiments. If DNA from a variety of different organisms is examined for sequences similar to a human probe sequence, the Southern blot is called a **zoo blot**. The more closely related the organisms, the more likely a related sequence will be found. For example, if the human sequence were from a hemoglobin gene, the results might show a related sequence in chimpanzee, horse, and pig, but not in yeast or bacteria. The most interesting use for zoo blots is to test for noncoding versus coding DNA (Fig. B). Sequences of noncoding DNA evolve rapidly compared to coding sequences. Hence a probe that recognizes genomic DNA from many different related organisms is usually from a coding region. If the probe sequence is noncoding DNA, then very few matches will occur.

![Zoo Blot Diagram](image)

**FIGURE B Zoo Blot**

A specialized form of Southern blotting, called zoo blotting, is used to distinguish coding DNA from noncoding regions. The target DNA includes several samples of genomic DNA from different animals, hence the term “zoo.” The probe is a segment of human DNA that may or may not be from a coding region. Since base sequences of noncoding DNA mutate and change rapidly, whereas coding sequences do not mutate as rapidly, a probe that recognizes genomic DNA from many different organisms is usually a coding region. On the left, the only hybrid seen was between the probe and the human DNA; therefore, the probe is probably noncoding. In the example on the right, the probe binds to related sequences in other animals; therefore, this probe is probably from a coding region.

<table>
<thead>
<tr>
<th>Loading slots</th>
<th>Probe was non-coding; binds only to itself</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rat Cow Dog Human</td>
<td></td>
</tr>
<tr>
<td>Probe IS CODING DNA; all species have a related sequence</td>
<td></td>
</tr>
<tr>
<td>Rat Cow Dog Human</td>
<td></td>
</tr>
</tbody>
</table>

**FLUORESCENCE IN SITU HYBRIDIZATION (FISH)**

The previously discussed hybridization techniques rely on purified DNA or RNA run in an agarose gel. In **fluorescence in situ hybridization (FISH)**, the probe is hybridized directly to DNA or RNA within the cell. As described earlier, the probe is a small segment of DNA that has been labeled with fluorescent tags in order to be visualized. The target DNA or RNA is located within the cell and requires some special processing. The target cells may be extremely thin sections of tissue from a particular organism. For example, when a person has a biopsy, a small piece of tissue is removed for analysis. This tissue is preserved and then cut into extremely thin sections to be analyzed under a microscope. These can be used to determine the presence of a gene with FISH. Another source of target cells for FISH is cultured mammalian or insect cells (see Chapter 1). Additionally, blood can be isolated and processed to isolate the white blood cells. (Note: Red blood cells do not contain a nucleus and therefore do not contain DNA.) Chromosomes from white blood cells can be isolated and dropped onto a glass slide. Blood sample DNA may be found both in interphase, with the chromosomes spread out, or in metaphase, with the chromosomes highly condensed.

Whether the target DNA is from blood, cells cultured in dishes, or actual tissue sections, the cells must be heated to make the DNA single-stranded. Samples where RNA is the target do not need to be heated. The fluorescently labeled probe hybridizes to complementary sequences in the DNA or RNA, and when the cells are illuminated at the appropriate wavelength, the probe location on the chromosome can be identified by fluorescence. Figure 3.13 shows FISH analysis of the **RUNX1** gene that is amplified in certain cases of human acute leukemia due to polysomy (that is, multiple copies) of chromosome 21.
GENERAL PROPERTIES OF CLONING VECTORS

Cloning vectors are specialized plasmids (or other genetic elements) that will hold any piece of foreign DNA for further study or manipulation. The numbers and types of plasmids available for cloning have grown. In addition, other DNA elements are now used, including viruses and artificial chromosomes. Once a fragment of DNA has been cloned and inserted into a suitable vector, large amounts of DNA can be manufactured, the sequence can be determined, and any

BOX 3.3 DISCOVERY OF RECOMBINANT DNA

In 1972, two researchers met at a conference in Hawaii to discuss plasmids, the small rings of extrachromosomal DNA found in bacteria. Herbert W. Boyer, PhD, was a faculty member at the University of California, San Diego, and he was studying restriction and modification enzymes. He had just presented his research on EcoRI. Stanley N. Cohen, MD, was a faculty member at Stanford, and he was interested in how plasmids could confer resistance to different antibiotics. His lab perfected laboratory transformation of *Escherichia coli* using calcium chloride to permeabilize the cells. After the talks ended, the two met over corned beef sandwiches and combined their two ideas.

They isolated different fragments of DNA from animals, other bacteria, and viruses and, using restriction enzymes, ligated the fragments into a small plasmid from *E. coli*. This was the first recombinant DNA to be made. Finally, they transformed the engineered plasmid back into *E. coli*. The cells expressed the normal plasmid genes as well as those inserted into the plasmid artificially. This sparked the revolution in genetic engineering, and since then every biotechnology lab has used some variation of their technique. Boyer and Cohen applied for a patent on recombinant DNA technology. In fact, Boyer cofounded Genentech with Robert Swanson, a venture capitalist. Genentech is one of the first biotechnology companies in the United States, and under Boyer and Swanson, the company produced human somatostatin in *E. coli*.

**FIGURE 3.13**

Gene Location on Chromosomes by FISH

(A) and (B) FISH of interphase nuclei with a dual-color DNA probe that shows *RUNX1* (= *AML1*) in red and *TEL* (telomerase) in green. (A) Patient 1 showed multiple *RUNX1* signals and lacked one *TEL* signal (arrow). (B) Patient 2 also showed extra copies of *RUNX1*, but had two normal *TEL* signals (arrows). (C) Partial GTG-banding karyotype and CGH profile of chromosomes 21 from both patients, showing that the amplification threshold is exceeded for the 21q22 region where *RUNX1* is located. From: Garcia-Casado et al. (2006). High-level amplification of the *RUNX1* gene in two cases of childhood acute lymphoblastic leukemia. *Cancer Genet Cytogenet* **170**, 171–174. Reprinted with permission.

FISH is a technique where a labeled probe is incubated with cells that have had their DNA denatured by heat. The probe hybridizes to its homologous sequence on the chromosome.
genes in the fragment can be expressed in other organisms. Studying human genes in humans is virtually impossible because of the ethical ramifications. In contrast, studying a human gene expressed in bacteria provides useful information that can often be applied to humans. Modern biotechnology depends on the ability to express foreign genes in model organisms. Before discussing how a gene is cloned, the properties of vectors are considered.

**Useful Traits for Cloning Vectors**

Although many specialized vectors now exist, the following properties are convenient and found in most modern generalized cloning plasmids:

- Small size, making them easy to manipulate once they are isolated
- Easy to transfer from cell to cell (usually by transformation)
- Easy to isolate from the host organism
- Easy to detect and select
- Multiple copies helps in obtaining large amounts of DNA
- Clustered restriction sites (polylinker) to allow insertion of cloned DNA
- Method to detect presence of inserted DNA (e.g., alpha complementation)

Most bacterial plasmids satisfy the first three requirements. The next key trait of cloning vectors is an easy way to detect their presence in the host organism. Bacterial cloning plasmids often have antibiotic resistance genes that make bacteria resistant to particular antibiotics. When treated with the antibiotic, only bacteria with the plasmid-borne resistance gene will survive. Other bacteria will die. Other traits have been exploited to detect plasmids. Vectors derived from the yeast \(2 \mu\) plasmid often carry genes for essential amino acids, such as leucine. The host strain of yeast is defective in the corresponding gene and unable to grow on media lacking the amino acid, unless the plasmid is present.

Plasmids vary in their **copy number**. Some plasmids exist in just one or a few copies in their host cells whereas others exist in multiple copies. Such multicopy plasmids are in general more useful as the amount of plasmid DNA is higher, making them easier to isolate and purify. The type of origin of replication controls the copy number, since this region on the plasmid determines how often DNA polymerase binds and induces replication.

Most cloning vectors have several unique restriction enzyme sites. Usually these sites are grouped in one location called the **polylinker** or **multiple cloning site (MCS)**; Fig. 3.14. This allows researchers to open the cloning vector at one site without disrupting any of the vector’s replication genes. Fragments of foreign DNA are digested with enzymes matching those in the polylinker. Ligase connects the vector and insert. Specific restriction enzyme sites can be added using PCR primers or synthetic DNA oligomers (see Chapter 4).

Some vectors have ways to detect whether or not they contain an insert. The simplest way to do this is **insertional inactivation** of an antibiotic gene (Fig. 3.15). Here, the vector has two different antibiotic resistance genes. The foreign DNA is inserted into one of the antibiotic-resistant genes. Thus the host bacteria will be resistant to one antibiotic and sensitive to the other.

Alternatively, **alpha complementation** may be used (see Fig. 3.15). The vector has a short portion of the \(\beta\)-galactosidase gene (the alpha fragment), and the bacterial chromosome has the rest of the gene. If both give rise to proteins, the subunits combine to form functional \(\beta\)-galactosidase. If DNA is inserted into the plasmid-borne gene segment, the encoded subunit is not made and \(\beta\)-galactosidase is not produced. When
β-galactosidase is expressed, the bacteria can degrade X-gal, which turns the bacterial colony blue. If a piece of DNA is inserted into the alpha fragment gene, the bacteria cannot split X-gal and stay white.

Once an appropriate vector has been chosen for the gene of interest or other insert, the two pieces are ligated into one construct. The term construct refers to any recombinant DNA molecule that has been assembled by genetic engineering. If both the vector and insert are cut with the same restriction enzyme, the two pieces have complementary ends and require only ligase to link them. Tricks are used to make two pieces of DNA with unrelated ends compatible. Sometimes, short oligonucleotides are synthesized and added onto the ends of the insert to make them compatible with the vector. These short oligonucleotides are called linkers, and they add one or a few new restriction enzyme sites to the ends of a segment of DNA.
SPECIFIC TYPES OF CLONING VECTORS

Because *E. coli* is the main host organism used for manipulating DNA, most vectors are based on plasmids or viruses that can survive in *E. coli* or similar bacteria. Most vectors have bacterial origins of replication and antibiotic resistance genes. The polylinker or multiple cloning site is usually placed between prokaryotic promoter and terminator sequences. Some vectors may also supply the ribosome binding site, so that any inserted coding sequence will be expressed as a protein. Many other features are present in specialized cloning vectors. The following discussion will introduce some of the different categories of vectors with their essential features (Fig. 3.16).

Many vectors are based on the 2µ circle of yeast. The native version of the 2µ circle has been modified in a variety of ways for use as a cloning vector. A shuttle vector contains origins of replication for two organisms plus any other sequences necessary to survive in either organism (see Fig. 3.16B). Shuttle vectors that are based on the 2µ plasmid have the components needed for survival in yeast and bacteria, plus antibiotic resistance and a polylinker. The Cen sequence is a eukaryotic centromere (Cen) sequence that keeps the plasmid in the correct location during mitosis and meiosis in yeast. Because yeast cells are eukaryotic and also have such a thick cell wall, most antibiotics do not kill yeast. Therefore, a different strategy is used to detect the presence of plasmids in yeast. A gene for synthesis of an amino acid, such as leucine, allows strains of yeast that require leucine to grow.

**Bacteriophage vectors** are viral genomes that have been modified so that large pieces of nonviral DNA can be packaged in the virus particle. Lambda bacteriophages have linear genomes with two cohesive ends—cos sequences (lambda cohesive ends). These are 12-base overlapping sticky ends. When inside the virus coat, the cohesive ends are coated with protein to prevent them from annealing. After lambda attaches to *E. coli*, it inserts just the linear DNA. The proteins that protect the cohesive ends are lost, and the genome circularizes with the help of DNA ligase. The circular form is the replicative form (RF), and it replicates by the rolling circle mechanism (see Chapter 4). Expression of various lambda genes produces the proteins that assemble into protein coats. Each coat is packaged with one genome, and after many of these are assembled, the *E. coli* host explodes, releasing the new bacteriophage to infect other cells.

The lambda bacteriophage is a widely used cloning vector (see Fig. 3.16C). The middle segment of the lambda genome has been deleted and a polylinker has been added. An insert of 37 to 52 kb can be ligated into the polylinker and packaged into viral particles. In order to work with the bacteriophage DNA without killing the entire *E. coli* culture, one or more genes necessary for packaging are deleted. When the researcher wants to form fully packaged bacteriophages, coat proteins from helper virus can be added (Fig. 3.17). The helper viruses do not contain foreign DNA, but supply the missing genes for the coat proteins. Because coat proteins self-assemble in vitro, helper lysates are mixed with recombinant lambda DNA and complete virus particles containing DNA are produced. This is known as in vitro packaging.

**Cosmid vectors** can hold pieces of DNA up to 45 kb in length (see Fig. 3.16D). These are highly modified lambda vectors with all the sequences between the cos sites removed and replaced with the insert. The DNA of interest is ligated between the two cos sites using restriction enzymes and ligase. This construct is packaged into a lambda particle.
**Figure 3.16 Various Cloning Vectors**

(A) Typical bacterial cloning vector. This vector has bacterial sequences to initiate replication and transcription. In addition, it has a multiple cloning site embedded within the *lacZ* α gene so that the insert can be identified by alpha-complementation. The antibiotic resistance gene allows the researcher to identify any *E. coli* cells that have the plasmid. (B) Yeast shuttle vector. This vector can survive in either bacteria or yeast because it has both yeast and bacterial origin of replication, a yeast centromere, and selectable markers for yeast and bacteria. As with most cloning vectors, there is a polylinker. (C) Lambda replacement vectors. Because lambda phage is easy to grow and manipulate, its genome has been modified to accept foreign DNA inserts. The region of the genome shown in green is nonessential for lambda growth and packaging. This region can be replaced with large inserts of foreign DNA (up to about 23 kb). (D) Cosmids. Cosmids are small multicopy plasmids that carry cos sites. They are linearized and cut so that each half has a cos site (not shown). Next, foreign DNA is inserted to relink the two halves of the cosmids DNA. This construct is packaged into lambda virus heads and used to infect *E. coli*. (E) Artificial chromosomes. Yeast artificial chromosomes have two forms, a circular form for growing in bacteria and a linear form for growing in yeast. The circular form is maintained like any other plasmid in bacteria, but the linear form must have telomere sequences to be maintained in yeast. The linear form can hold up to 2000 kb of cloned DNA and is very useful for genomics research.
A lambda cloning vector containing cloned DNA must be packaged in a phage head before it can infect *E. coli*. First, one culture of *E. coli* cells is infected with a mutant lambda that lacks the gene for one of the head proteins called E. A different culture of *E. coli* is infected with a different mutant, which lacks the phage head protein D. The two cultures are induced to lyse, which releases the tails, assembly proteins, and head proteins, but no complete heads because of the missing proteins. When these are mixed with a lambda replacement vector, the three spontaneously form complete viral particles containing DNA. These are then used to infect *E. coli*.

**Artificial chromosomes** hold the largest pieces of DNA (see Fig. 3.16E). These include yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), and P1 bacteriophage artificial chromosomes (PACs). They are used to contain lengths of DNA from 150 kb to 2000 kb. YACs hold the largest amount of DNA, up to about 2000 kb. YACs have yeast centromeres and yeast telomeres for maintenance in yeast. BACs can be circularized and grown in bacteria; therefore, they have a bacterial origin of replication and antibiotic resistance genes. The flexibility of artificial chromosomes makes them most useful for sequencing entire genomes, especially those of higher organisms with vast amounts of DNA.

Many different cloning vectors are available to biotechnology research. The smaller genes are studied using bacterial plasmids or shuttle vectors, whereas the larger genes are studied in bacteriophage vectors, cosmids, and artificial chromosomes.

Shuttle vectors have sequences that enable them to survive in two different organisms such as yeast and bacteria.

Bacteriophage vectors have critical genes removed so that the bacteriophage cannot destroy the host cell by producing phage particles. Adding a helper phage restores this activity.
GETTING CLONED GENES INTO BACTERIA BY TRANSFORMATION

Once the gene of interest is cloned into a vector, the construct can be put back into a bacterial cell through a process called transformation (Fig. 3.18). Here the “naked” DNA that was constructed in the laboratory is mixed with competent E. coli cells. To make the cells competent, that is, able to take up naked DNA, the cell wall must be temporarily opened up. In the laboratory, E. coli cells are treated with high concentrations of calcium ions on ice, and then shocked at a higher temperature for a few minutes. Most of the cells die during the treatment, but some survive and take up the DNA. Another method to make E. coli cells competent is to expose them to a high-voltage shock. Electroporation opens the cell wall, allowing the DNA to enter. This method is much faster and more versatile. Electroporation is used for other types of bacteria as well as yeast.

When a mixture of different clones is transformed into bacteria as in a gene library (see later discussion), cells that take up more than one construct usually lose one of them. For example, if genes A and B are both cloned into the same kind of vector and both cloned genes get into the same bacterial cell, the bacteria will lose one plasmid and keep the other. This is due to plasmid incompatibility, which prevents one bacterial cell from harboring two of the same type of plasmid. Incompatibility stems from conflicts between two plasmids with identical or related origins of replication. Only one is allowed to replicate in any given cell. If a researcher wants a cell to have two cloned genes, then two different types of plasmids could be used, or both genes could be put onto the same plasmid.

CONSTRUCTING A LIBRARY OF GENES

Gene libraries are used to find new genes, to sequence entire genomes, and to compare genes from different organisms (Fig. 3.19). Gene libraries are made when the entire DNA from one particular organism is digested into fragments using restriction enzymes, and then each of the fragments is cloned into a vector and transformed into an appropriate host.

The basic steps used to construct a library are:

1. Isolate the chromosomal DNA from an organism, such as E. coli, yeast, or humans.
2. Digest the DNA with one or two different restriction enzymes.
3. Linearize a suitable cloning vector with the same restriction enzyme(s).
4. Mix the cut chromosome fragments with the linearized vector and ligate.
5. Transform this mixture into *E. coli*.
6. Isolate large numbers of *E. coli* transformants.

The type of restriction enzyme affects the type of library. Because restriction sites are not evenly spaced in the genome, some inserts will be large and others small. Using a restriction enzyme that recognizes only four base pairs will give a mixture of mostly small fragments, whereas a restriction enzyme that has a six or eight base-pair recognition sequence will generate larger fragments. (Note that finding a particular four base-pair sequence in a genome is more likely than finding a six base-pair sequence.) Even if an enzyme that recognizes a four base-pair recognition sequence is used to digest the entire genome, there may still be segments that are too large to be cloned. Conversely, clustered restriction sites will cause some genes will be cut into several pieces. To avoid this, partial digestion is often used. The enzyme is allowed to cut the DNA for only a short time, and many of the restriction enzyme sites are not cut, leaving larger pieces for the library. In addition, it is usual to construct another library using a different restriction enzyme.

Gene libraries are used for many purposes because they contain almost all the genes of a particular organism.
SCREENING THE LIBRARY OF GENES BY HYBRIDIZATION

Once the library is assembled, researchers often want to identify a particular gene or segment of DNA within the library. Sometimes the gene of interest is similar to one from another organism. Sometimes the gene of interest contains a particular sequence. For example, many enzymes use ATP to provide energy. Enzymes that bind ATP share a common signature sequence whether they come from bacteria or humans. This sequence can be used to find other enzymes that bind ATP. Such common sequence motifs may also suggest that a protein will bind various cofactors, other proteins, and DNA, to name a few examples.

Screening DNA libraries by hybridization requires preparing the library DNA and preparing the labeled probe. A gene library is stored as a bacterial culture of *E. coli* cells, each having a plasmid with a different insert. The culture is grown up, diluted, and plated onto many different agar plates so that the colonies are spaced apart from one another. The colonies are transferred to a nylon filter and the DNA from each colony is released from the cells by lysing them with detergent. The cellular components are rinsed from the filters. The DNA sticks to the nylon membrane and is then denatured to form single strands (Fig. 3.20).

If a scientist is looking for a particular gene in the target organism, the probe for the library may be the corresponding gene from a related organism. The probe is usually just a segment of the gene, because a smaller piece is easier to manipulate. The probe DNA can be synthesized and labeled either with radioactivity or with chemiluminescence. The probe is heated to make it single-stranded and mixed with the library DNA on the nylon filters. The probe hybridizes with matching sequences in the library. The level of match needed for binding can be adjusted by incubating at various temperatures. The higher the temperature, the more stringent, that is, the more closely matched the sequences must be. The lower the temperature, the less stringent. If the probe is labeled with radioactivity, photographic film will turn black where the probe and library DNA hybridized. The black spot is aligned with the original bacterial colony. Usually the most likely colony plus its neighbors are selected, grown, plated, and rescreened with the same probe to ensure that a single transformant is isolated. Then the DNA from this isolate can be analyzed by sequencing (see Chapter 4).

EUKARYOTIC EXPRESSION LIBRARIES

In expression libraries, the vector has sequences required for transcription and translation of the insert. This means that the insert DNA is expressed as RNA, and this may be translated into a protein. An expression library, in essence, generates a protein from every cloned insert, whether it is a real gene or not. When studying eukaryotic DNA, expression libraries are constructed using complement DNA (cDNA) to help ensure the insert is truly a gene. Eukaryotic DNA, especially in higher plants and animals, is largely noncoding, with coding regions spaced far apart. Even genes are interrupted with noncoding introns. cDNA is a double-stranded DNA copy of mRNA made by using reverse transcriptase. Reverse transcriptase was first identified in retroviruses (see Chapter 1). It is used in eukaryotic research to eliminate the introns and generate a version of a gene consisting solely of an uninterrupted coding sequence.
In contrast, bacteria have very little noncoding DNA and their genes are not interrupted by introns; therefore, genomic DNA can be used directly in expression libraries.

Eukaryotic DNA is first made into cDNA in order to construct an expression library (Fig. 3.21). To make cDNA, the messenger RNA is isolated from the organism of interest by

**FIGURE 3.21 Making a cDNA Library from Eukaryotic mRNA**

First, eukaryotic cells are lysed and the mRNA is purified. Next, reverse transcriptase plus primers containing oligo(dT) stretches are added. The oligo(dT) hybridizes to the adenine in the mRNA poly(A) tail and acts as a primer for reverse transcriptase. This enzyme makes the complementary DNA strand, forming an mRNA/cDNA hybrid. The mRNA strand is digested with ribonuclease H, and DNA polymerase I is added to synthesize the opposite DNA strand, thus creating double-stranded cDNA. Next, S1 nuclease is added to trim off any single-stranded ends, and linkers are added to the ends of the dsDNA. The linkers have convenient restriction enzyme sites for cloning into an expression vector.
binding to a column containing poly(T) (i.e., a DNA strand consisting of repeated thymines). This isolates only mRNA because poly(T) anneals to the poly(A) tail of eukaryotic mRNA. The mRNA is made double-stranded using reverse transcriptase, which synthesizes a DNA complement to mRNA. The hybrid mRNA:DNA molecule is converted to double-stranded DNA using RNase H and DNA polymerase I. RNase H cuts nicks into the mRNA backbone, and DNA polymerase displaces the mRNA strand and synthesizes DNA (see Fig. 3.21).

The cDNA is then ligated into an expression vector with sequences that initiate transcription and translation of the insert. In some cases, the insert will have its own translation start site (e.g., a full-length cDNA). If, as often occurs, the insert does not contain a translation start, then the reading frame becomes an issue. Because the genetic code is triplet, each insert can be translated in three different reading frames. A protein may be produced for all three reading frames, but only one frame will actually produce the correct protein. To ensure obtaining inserts with the correct reading frame, each cDNA is cloned in all three reading frames by using linkers with several different restriction sites. The number of transformants to screen is therefore increased greatly.

The cloned genes are transformed into bacteria, which express the foreign DNA. The bacteria are grown on agar and the colonies are then transferred to a nylon membrane and lysed. The proteins released are attached to the nylon membranes and are screened in various ways. Most often, an antibody to the protein of interest is used (see Chapter 6). This recognizes the protein and can be identified using a secondary antibody that is conjugated to a detection system. Usually alkaline phosphatase is conjugated to the secondary antibody. The whole complex can be identified because alkaline phosphatase cleaves X-Phos, leaving a blue color where the bacterial colony expressed the right protein (Fig. 3.22). *E. coli* cannot perform most of the posttranslational modifications that eukaryotic proteins often undergo. Therefore, the proteins are not always in their native form. Nonetheless, appropriate antibodies can detect most proteins of interest.

### FEATURES OF EXPRESSION VECTORS

Because foreign proteins, especially if made in large amounts, can be toxic to *E. coli*, the promoter used to express the foreign gene is critical. If too much foreign protein is made, the host cell may die. To control protein production, expression vectors have promoters with on/off switches; therefore, the host cell is grown up first and the foreign protein is expressed later. One commonly used promoter is a mutant version of the lac promoter (Fig. 3.23). This *lacUV* promoter drives a very high level of transcription, but only under induced conditions. It has the following elements: a binding site for RNA polymerase, a binding site for the LacI repressor protein, and a transcription start site. The vector has strong transcription stop sites downstream of the polylinker region. The vector also has the gene for LacI so that high levels of repressor protein are made, thus keeping the cloned genes repressed. Like all vectors, there is an origin of replication and antibiotic resistance gene for selection in bacteria. When a gene library is cloned behind this promoter, the genes are not expressed due to high levels of LacI repressor. When an inducer, such as IPTG, is added, LacI is released from the DNA and RNA polymerase transcribes the foreign, cloned, genes.

Another common promoter in expression vectors is the lambda left promoter, or *P*~L~. It has a binding site for the lambda repressor. The gene of interest or library fragment is not expressed unless the repressor is removed. Rather than using its natural inducer, a mutant

Complementary DNA or cDNA is constructed by isolating mRNA and making a DNA copy with reverse transcriptase.

Expression libraries express the foreign DNA insert as a protein because expression vectors contain sequences for both transcription and translation. The protein of interest is identified by incubating the library with an antibody to the protein of interest.
version of the repressor has been isolated that releases its binding site at high temperatures. So when the culture is shifted to 42 °C, the repressor falls off the DNA and RNA polymerase transcribes the cloned genes.

Another expression system uses a promoter whose RNA polymerase binding site only recognizes RNA polymerase from the bacteriophage T7. Bacterial RNA polymerase will not transcribe the gene of interest. This system is designed to work only in bacteria that have the gene for T7 RNA polymerase integrated into the chromosome and under the control of an inducible promoter.

Some expression vectors contain a small segment of DNA that encodes a protein tag. These are primarily used when the gene of interest is already cloned, rather than for screening libraries. The gene of interest must be cloned
in frame with the DNA for the protein tag. The tag can be of many varieties, but **6HIS**, **Myc**, and **FLAG®** tag are three popular forms (Fig. 3.24). 6HIS is a stretch of six histidine residues put at the beginning or end of the protein of interest. The histidines bind strongly to nickel. This allows the tagged protein to be isolated by binding to a column with nickel attached. Myc and FLAG are epitopes that allow the expressed protein to be purified by binding to the corresponding antibody. The antibodies may be attached to a column, used for a Western blot, or seen *in vivo* by staining the cells with fluorescently tagged versions of the Myc or FLAG antibodies. (The histidine tag can also be recognized with a specific antibody, if desired.)

The most important feature of expression vectors is a tightly controlled promoter region. The proteins of the expression library are expressed only under certain conditions, such as presence of an inducer, removal of a repressor, or change in temperature.

Small tags can be fused into the protein of interest using expression vectors. These tags allow the protein of interest to be isolated and purified.

**FIGURE 3.24 Using Tags to Isolate Proteins**

Some expression vectors have DNA sequences that code for short protein tags. The 6HIS tag (A) codes for six histidine residues. When fused in-frame with the coding sequence for the cloned gene, the tag is fused to the protein. The 6HIS tag specifically binds to nickel ions; therefore, binding to a nickel ion column isolates 6HIS-tagged proteins. Other tags, such as Myc or FLAG (B), are specific antibody epitopes that work in a similar manner. Myc-tagged or FLAG-tagged proteins can be isolated or identified by binding to antibodies to Myc or FLAG, respectively.
**SUBTRACTIVE HYBRIDIZATION**

Subtractive hybridization is a screening method that allows researchers to find genes that are “missing.” For example, the gene responsible for a hereditary defect may be totally deleted in one particular victim. A healthy person will have the complete gene. Therefore, the DNA of the two people will be identical, except for an extra segment in the person without the disease. The DNA from the person with the deletion is isolated and cut with one restriction enzyme (Fig. 3.25). The DNA from the healthy person is isolated and cut with a second, different restriction enzyme. An excess amount of mutant DNA is mixed with healthy DNA, and then heating denatures the mixture. Slowly cooling the mixture allows hybrid molecules to form from the normal and mutant DNA. If two mutant fragments anneal, the dsDNA fragment will have sites for restriction enzyme 1 at each end. If two healthy fragments hybridize, the ends will have sites for restriction enzyme 2. If healthy DNA hybridizes with mutant DNA, the two will have nonmatching ends that cannot be cut by either restriction enzyme. All regions of the DNA will be able to form mutant: normal hybrids except for the

---

**FIGURE 3.25 Cloning by Subtractive Hybridization**

The key to subtractive hybridization is to hybridize all the wild-type or “healthy” DNA fragments (pink) with an excess of mutant DNA (purple). In this example, the mutant DNA is digested with restriction enzyme 1 and the wild-type DNA is digested with enzyme 2. Next, the two DNA samples are mixed, denatured, and allowed to anneal. The DNA will hybridize to form mutant/mutant strands (flanked by sites for restriction enzyme 1); mutant/healthy strands (flanked by incompatible ends); or healthy/healthy strands (flanked by sites for enzyme 2). The healthy/healthy hybrids will be rare and should correspond to the region of the deletion. They are easily separated from the other hybrids by cloning into a vector cut with restriction enzyme 2.
region that is missing in the mutant DNA. This region of DNA can only self-hybridize, and the dsDNA formed will have sites for restriction enzyme 2 at the end. These segments can be cloned into a vector that has a corresponding restriction site. Overall, DNA that does not encode the gene of interest is excluded or subtracted by hybridization.

Subtractive hybridization is also used to compare gene expression under two different conditions (Fig. 3.26). For example, a researcher can compare the genes expressed by E. coli at low temperature versus those expressed in heat shock. First, separate cultures of bacteria are grown in both conditions and mRNA is isolated from both cultures. The mRNA from the bacteria kept at low temperature is made into cDNA to provide complementary sequences for hybridization. The cDNA is bound to a filter, denatured to give ssDNA, and incubated with the mRNA from the bacteria grown under the experimental conditions (i.e., heat shock). The mRNA that is present under both conditions will hybridize to the cDNA on the filter. However, mRNA that is expressed only under the experimental condition will not find a complementary sequence and will be left in solution. These unique mRNAs are then made into cDNA, cloned into a vector, and sequenced to identify genes that increase in expression under hot conditions.

Subtractive hybridization can be used to identify a gene that causes a disease or a set of genes that are expressed under certain conditions.

**Figure 3.26** Subtractive Hybridization Captures mRNA Expressed under Specific Conditions

Two different cultures are grown, one under standard conditions (green) and the other under experimental conditions (orange). The mRNA is isolated from each culture, and one sample is converted to cDNA (purple). The cDNA is bound to the filter and denatured to make it single-stranded. The experimental mRNA (orange) is passed through the filter. Only mRNA that is unique to the experimental conditions will fail to find a partner to hybridize. Consequently, this mRNA will come through the filter. Conversely, any mRNA that is present under both growth conditions will be trapped on the filter.
Summary

Recombinant DNA technology is the basis for almost all biotechnology research. Understanding these techniques is tantamount to understanding the rest of the textbook. First, DNA must be isolated from the organism in order to identify novel genes, to recover new recombinant vectors, or to purify a new library clone. The DNA is isolated from the cellular components using enzymes followed by centrifugation, RNA digestion with RNase, and precipitation with ethanol. Each organism will require special adaptations of this basic process in order to remove the cellular and extracellular components.

The purified DNA can be manipulated in many different ways. This chapter describes how restriction enzymes can cut the phosphate backbone of the DNA into smaller fragments, then how these fragments can be visualized by gel electrophoresis. The chapter then describes various methods to label specific DNA pieces so only one particular piece can be visualized. Hybridization is a key technique for FISH, Southern blots, Northern blots, and dot blots, as well as the screening of a genomic library for a particular sequence.

The chapter also outlines the key characteristics of vectors, including plasmids, bacteriophage vectors, cosmids, and artificial chromosomes. These extrachromosomal genetic elements vary in their uses but are very important to getting a foreign gene expressed in a host organism. Vectors require a region that is convenient to adding a foreign piece of DNA such as a multicloning site, they need a gene for selection, and they need some easy way to identify whether or not the vector contains the foreign piece of DNA.

Finally, the chapter describes how to construct a library of genes from an organism. A genomic library simply contains all the DNA of the organism of interest, cut into smaller fragments, and cloned into a vector. The library recombinant clones are then returned to a host cell. In addition, the expression library uses mRNA rather than genomic DNA, and these foreign pieces of DNA are then turned into proteins in the host. Libraries are easily screened for particular sequences of interest by hybridization or antibodies to a protein of interest.

End-of-Chapter Questions

1. Which of the following statements about DNA isolation from *E. coli* is not correct?
   a. Chemical extraction using phenol removes proteins from the DNA.
   b. RNA is removed from the sample by RNase treatment.
   c. Detergent is used to break apart plant cells to extract DNA.
   d. Lysozyme digests peptidoglycan in the bacterial cell wall.
   e. Centrifugation separates cellular components based on size.

2. Which of the following is important for gel electrophoresis to work?
   a. Negatively charged nucleic acids to migrate through the gel.
   b. Ethidium bromide to provide a means to visualize the DNA in the gel.
   c. Agarose or polyacrylamide to separate the DNA based on size.
   d. Known molecular weight standards.
   e. All of the above are important for gel electrophoresis.
3. How are restriction enzymes and ligase used in biotechnology?
   a. Restriction enzymes cut DNA at specific locations, producing ends that can be ligated back together with ligase.
   b. Only restriction enzymes that produce blunt ends after cutting DNA can be ligated with ligase.
   c. Only restriction enzymes that produce sticky ends on the DNA can be ligated with ligase.
   d. Restriction enzymes can both cut DNA at specific sites and ligate them back together.
   e. Restriction enzymes randomly cut DNA, and the cut fragments can be ligated back together with ligase.

4. Which of the following is an appropriate method for detecting nucleic acids?
   a. Measuring absorbance at 260 nm.
   b. Autoradiography of radiolabeled nucleic acids.
   c. Chemiluminescence of DNA labeled with biotin or digoxigenin.
   d. Measuring the light emitted after excitation by fluorescent-labeled nucleic acids on a photodetector.
   e. All of the above are appropriate methods for detecting nucleic acids.

5. Why does the GC content of a particular DNA molecule affect the melting of the two strands?
   a. The G and C bond only requires two hydrogen bonds, thus requiring a lower temperature to “melt” the DNA.
   b. Because G and C base-pairing requires three hydrogen bonds and a higher temperature is required to “melt” the DNA.
   c. The percentage of As and Ts in the molecule is more important to melting temperature than the percentage of Gs and Cs.
   d. The nucleotide content of a DNA molecule is not important to know for biotechnology and molecular biology research.
   e. None of the above.

6. What is the difference between Southern and Northern hybridizations?
   a. Southern blots hybridize a DNA probe to a digested DNA sample but Northern blots hybridize to DNA probes to, usually, mRNA.
   b. Southern blots use an RNA probe to hybridize to DNA but Northern blots use an RNA probe to hybridize to RNA.
   c. Southern blots determine if a particular gene is being expressed but Northern blots determine the homology between mRNA and a DNA probe.
   d. Southern blots determine the homology between mRNA and a DNA probe but Northern blots determine if a particular gene is being expressed.
   e. Southern and Northern blots are essentially the same technique performed in different hemispheres of the world.

(Continued)
7. What might be a use for fluorescence in situ hybridization (FISH)?
   a. For identification of a specific gene in a DNA extraction by hybridization to a DNA probe.
   b. For identification of a specific gene by hybridization to a DNA probe within live cells that have had their DNA denatured by heat.
   c. For identification of an mRNA within an RNA extraction by hybridization to a DNA probe.
   d. For identification of both mRNA and DNA in cellular extracts using an RNA probe.
   e. None of the above.

8. Which of the following are useful traits of cloning vectors?
   a. An antibiotic resistance gene on the plasmid for selection of cells containing the plasmid.
   b. A site that contains unique, clustered restriction enzyme sequences for cloning foreign DNA.
   c. A high copy number plasmid so that large amounts of DNA can be obtained.
   d. Alpha complementation to determine if the foreign DNA was inserted into the cloning site.
   e. All of the above are useful traits.

9. Which of the following vectors holds the largest pieces of DNA?
   a. plasmids
   b. bacteriophage
   c. YACs
   d. PACs
   e. cosmids

10. Besides a high voltage shock, what is another method to make E. coli competent to take up “naked” DNA?
    a. high concentrations of calcium ions followed by high temperature
    b. high concentrations of calcium ions and several hours on ice
    c. large amounts of DNA added directly to a bacterial culture growing at 37°C
    d. high concentrations of minerals followed by high temperature
    e. A high voltage shock is the only way to make E. coli competent.

11. Why are gene libraries constructed?
    a. To find new genes.
    b. To sequence whole genomes.
    c. To compare genes to other organisms.
    d. To create a “bank” of all the genes in an organism.
    e. All of the above.

12. Which of the following statements about gene libraries is correct?
    a. Genes in a library can be compared to genes from other organisms by hybridization with a probe.
    b. A gene library is only necessary to maintain known genes.
    c. Every gene in the library must be sequenced first in order to compare genes in the library to genes from other organisms.
    d. Gene libraries are only created for eukaryotic organisms.
    e. Gene libraries can only be created in prokaryotes.
13. Why must reverse transcriptase be used to create a eukaryotic expression library?
   a. Reverse transcriptase is only used to create prokaryotic expression libraries.
   b. Reverse transcriptase creates cDNA from mRNA in prokaryotes.
   c. Reverse transcriptase ensures the gene is in the correct orientation within the expression vector to create protein.
   d. Reverse transcriptase creates cDNA from mRNA because genes in eukaryotes have large numbers of non-coding regions.
   e. No other enzymes are used to create expression libraries except restriction enzymes.

14. Which of the following are common features of expression vectors?
   a. Small segments of DNA that encode tags for protein purification.
   b. Transcriptional start and stop sites.
   c. A tightly controlled promoter than can only be induced under certain circumstances.
   d. Antibiotic resistance gene.
   e. All of the above are common features of expression vectors.

15. How is subtractive hybridization useful?
   a. To eliminate genes from a gene library.
   b. To create expression libraries based on genes that are currently being expressed.
   c. To identify and construct new probes for Southern and Northern hybridizations.
   d. To identify sets of genes that are only expressed under certain conditions.
   e. All of the above are useful traits of subtractive hybridization.

Further Reading