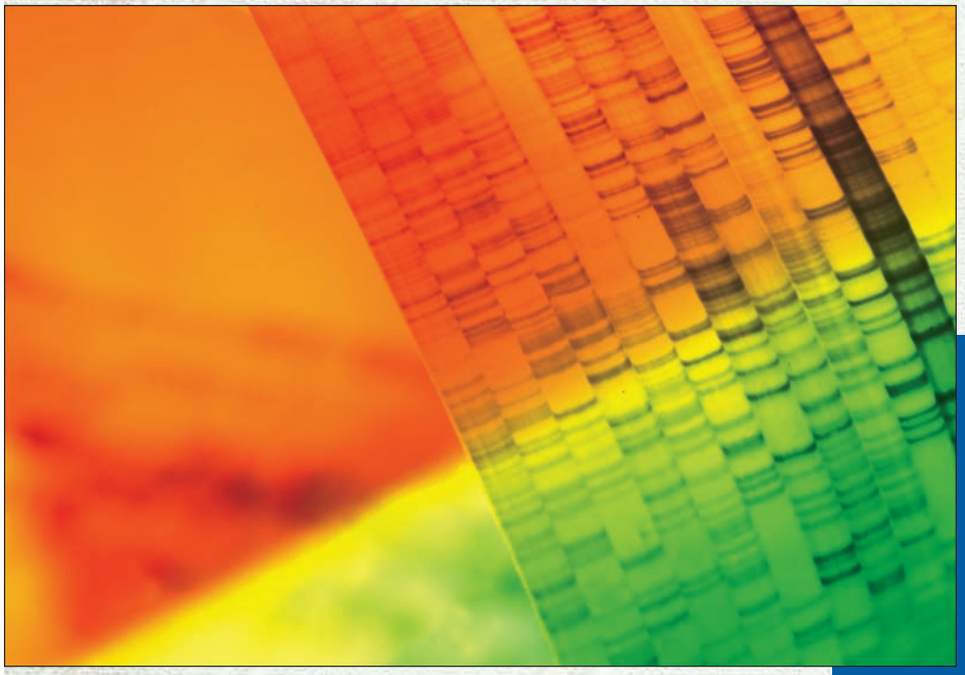
CHAPTER 5

Molecular Tools for Studying Genes and Gene Activity



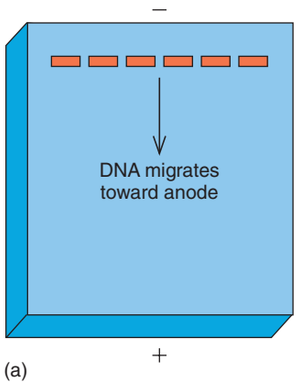
Autoradiograph of DNA sequencing gel. © *Peter Dazeley/Stone/Getty.*

In this chapter we will describe the most popular techniques that molecular biologists use to investigate the structure and function of genes. Most of these start with cloned genes. Many use gel electrophoresis. Many also use labeled tracers, and many rely on nucleic acid hybridization. We have already examined gene cloning techniques. Let us continue by briefly considering three other mainstays of molecular biology research: molecular separations including gel electrophoresis; labeled tracers; and hybridization.

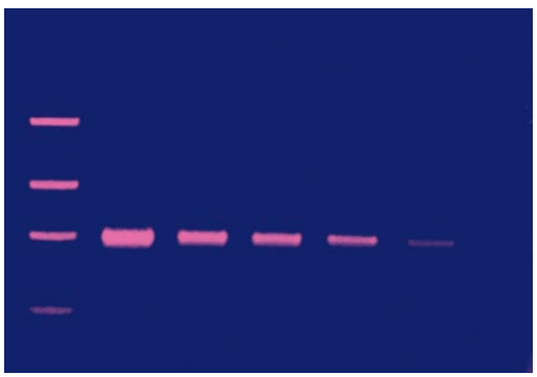
5.1 Molecular Separations

Gel Electrophoresis

It is very often necessary in molecular biology research to separate proteins or nucleic acids from each other. For example, we may need to purify a particular enzyme from a crude cellular extract in order to use it or to study its properties. Or we may want to purify a particular RNA or DNA molecule that has been produced or modified in an enzymatic reaction, or we may simply want to separate a series of RNAs or DNA fragments from each other. We will describe here some of the most common techniques used in such molecular separations, including gel electrophoresis of both nucleic acids and proteins, ion exchange chromatography, and gel filtration chromatography.



(a)



(b)

Figure 5.1 DNA gel electrophoresis. (a) Scheme of the method: This is a horizontal gel made of agarose (a substance derived from seaweed, and the main component of agar). The agarose melts at high temperature, then gels as it cools. A "comb" is inserted into the molten agarose; after the gel cools, the comb is removed, leaving slots, or wells (orange). The DNA is then placed in the wells, and an electric current is run through the gel. Because the DNA is an acid, it is negatively charged at neutral pH and electrophoreses, or migrates, toward the positive pole, or anode. **(b)** A photograph of a gel after electrophoresis showing the DNA fragments as bright bands. DNA binds to a dye that fluoresces orange under ultraviolet light, but the bands appear pink in this photograph.
(*Source:* (*b*) Reproduced with permission from Life Technologies, Inc.)

Gel electrophoresis can be used to separate different nucleic acid or protein species. We will begin by considering DNA gel electrophoresis. In this technique one makes an agarose gel with slots in it, as shown in Figure 5.1. The slots are formed by pouring a hot (liquid) agarose solution into a shallow box equipped with a removable "comb" with teeth that point downward into the agarose. Once the agarose has gelled, the comb is removed, leaving rectangular holes, or slots, in the gel. One puts a little DNA in a slot and runs an electric current through the gel at neutral pH. The DNA is negatively charged because of the phosphates in its backbone, so it migrates toward the positive pole (the *anode*) at the end of the gel. The secret of the gel's ability to separate DNAs of different sizes lies in friction. Small DNA molecules experience little frictional drag from solvent and gel molecules, so they migrate rapidly. Large DNAs, by contrast, encounter correspondingly more friction, so their mobility is lower. The result is that the electric current will distribute the DNA fragments according to their sizes: the largest near the top, the smallest near the bottom. Finally, the DNA is stained with a fluorescent dye and the gel is examined under ultraviolet illumination. Figure 5.2 depicts the results of such analysis on fragments of phage DNA of known size. The mobilities of these fragments are plotted versus the log of their molecular weights (or number of base pairs). Any unknown DNA can be electrophoresed in parallel with the standard fragments, and its size can be estimated if it falls within the range of the standards. For example, a DNA with a mobility of 20 mm in Figure 5.2 would contain about 910 bp. The same principles apply to electrophoresing RNAs of various sizes.

Solved Problem

Problem 1

Following is a graph showing the results of a gel electrophoresis experiment on double-stranded DNA fragments having sizes between 0.3 and 1.2 kb.

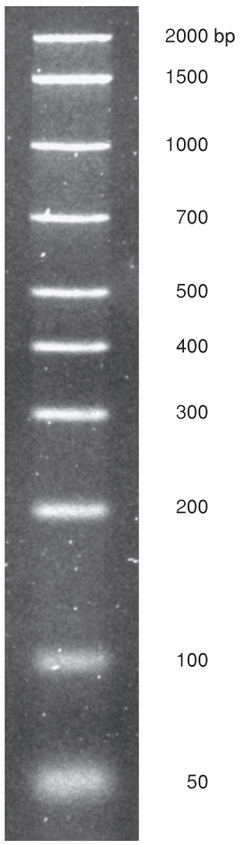
On the basis of this graph, answer the following questions:

a. What is the size of a fragment that migrated 16 mm in this experiment?

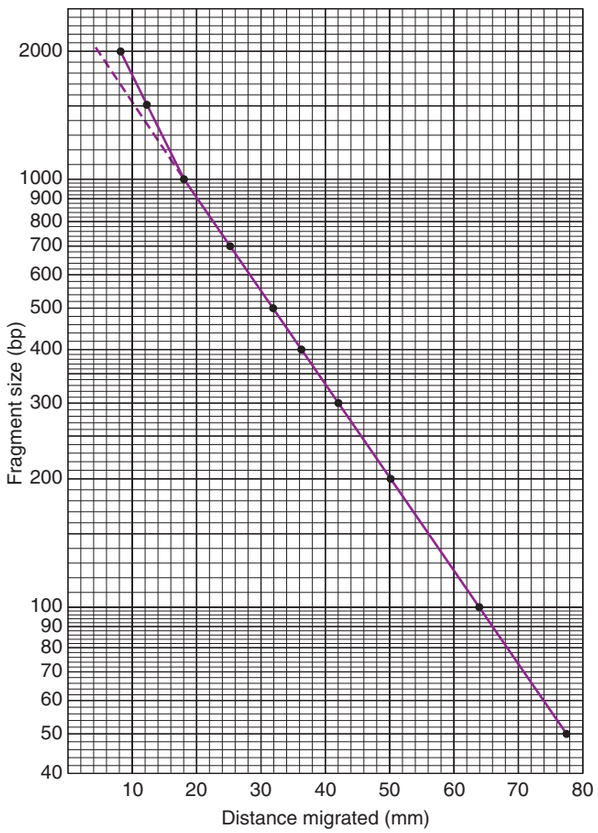
b. How far would a 0.5-kb fragment migrate in this experiment?

Solution

a. Draw a vertical dashed line from the 16-mm point on the *x* axis up to the experimental line. From the point where that vertical line intersects the experimental line, draw a horizontal dashed line to the

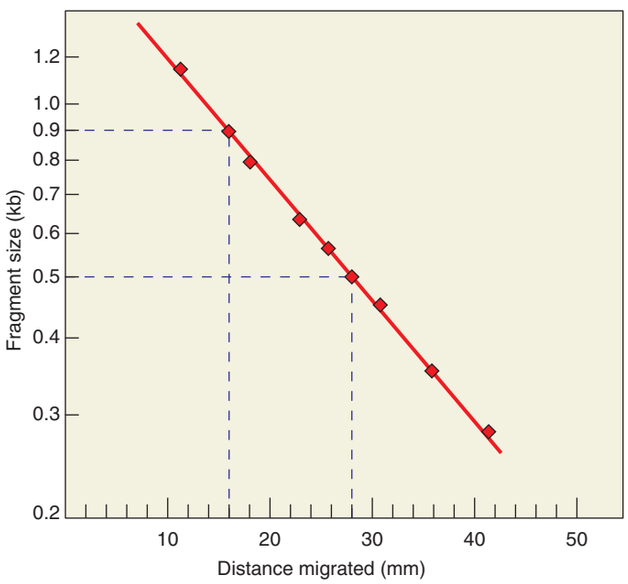


(a)



(b)

Figure 5.2 Analysis of DNA fragment size by gel electrophoresis. **(a)** Photograph of a stained gel of commercially prepared fragments after electrophoresis. The bands that would be orange in a color photo show up white in a black-and-white photo taken with an orange filter. The sizes of the fragments (in bp) are given at right. Note that this photo has been enlarged somewhat, so the mobilities of the bands appear a little higher than they really were. **(b)** Graph of the migration of the DNA fragments versus their sizes in base pairs. The vertical axis is logarithmic rather than linear, because the electrophoretic mobility (migration rate) of a DNA fragment is inversely proportional to the log of its size. However, notice the departure from this proportionality at large fragment sizes, represented by the difference between the solid line (actual results) and the dashed line (theoretical behavior). This suggests the limitations of conventional electrophoresis for measuring the sizes of very large DNAs. (*Source:* (*a*) Courtesy Bio-Rad Laboratories.)



y axis. This line intersects the *y* axis at the 0.9-kb point. This shows that fragments that migrate 16 mm in this experiment are 0.9 kb (or 900 bp) long.

b. Draw a horizontal dashed line from the 0.5-kb point on the *y* axis across to the experimental line. From the point where that horizontal line intersects the experimental line, draw a vertical dashed line down to the *x* axis. This line intersects the *x* axis at the 28-mm point. This shows that 0.5-kb fragments migrate 28 mm in this experiment. ■

Determining the size of a large DNA by gel electrophoresis requires special techniques. One reason is that the relationship between the log of a DNA's size and its electrophoretic mobility deviates strongly from linearity if the DNA is very large. A hint of this deviation is apparent at the top left of Figure 5.2b. Another reason is that double-stranded DNA is a relatively rigid rod—very long

and thin. The longer it is, the more fragile it is. In fact, large DNAs break very easily; even seemingly mild manipulations, like swirling in a beaker or pipetting, create shearing forces sufficient to fracture them. To visualize this, think of DNA as a piece of uncooked spaghetti. If it is short—say a centimeter or two—you can treat it roughly without harming it, but if it is long, breakage becomes almost inevitable.

In spite of these difficulties, molecular biologists have developed a kind of gel electrophoresis that can separate DNA molecules up to several million base pairs (**megabases, Mb**) long and maintain a relatively linear relationship between the log of their sizes and their mobilities. Instead of a constant current through the gel, this method uses pulses of current, with relatively long pulses in the forward direction and shorter pulses in the opposite, or even sideways, direction. This **pulsed-field gel electrophoresis** (**PFGE**) is valuable for measuring the sizes of DNAs even as large as some of the chromosomes found in yeast. Figure 5.3 presents the results of pulsed-field gel electrophoresis on yeast chromosomes. The 16 visible bands represent chromosomes containing 0.2–2.2 Mb.

Electrophoresis is also often applied to proteins, in which case the gel is usually made of polyacrylamide. We therefore call it **polyacrylamide gel electrophoresis,** or **PAGE.** To determine the polypeptide makeup of a complex protein, the experimenter must treat the protein so that the polypeptides, or subunits, will electrophorese independently. This is usually done by treating the protein with a detergent (**sodium dodecyl sulfate,** or **SDS**) to **denature** the subunits so they no longer bind to one another. The SDS has two added advantages: (1) It coats all the polypeptides with negative charges, so they all electrophorese toward the anode. (2) It masks the natural charges of the subunits, so they all electrophorese according to their molecular masses and not by their native charges. Small polypeptides fit easily through the pores in the gel, so they migrate rapidly. Larger polypeptides migrate more slowly. Researchers also usually employ a reducing agent to break covalent bonds between subunits.

Figure 5.4 shows the results of **SDS-PAGE** on a series of polypeptides, each of which is attached to a dye so they can be seen during electrophoresis. Ordinarily, the polypeptides would all be stained after electrophoresis with a dye such as Coomassie Blue.

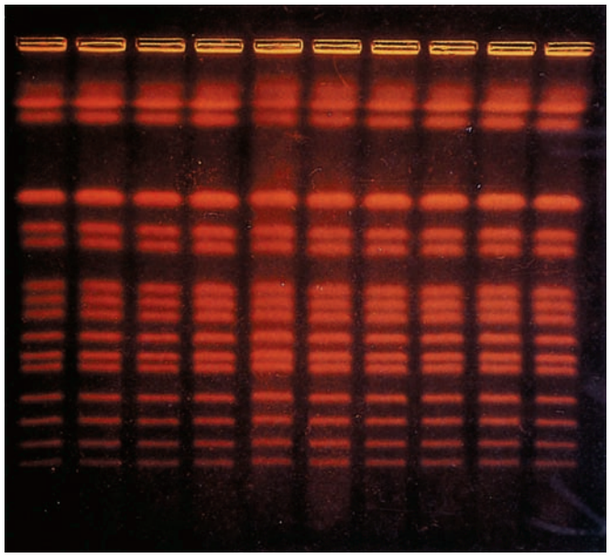


Figure 5.3 Pulsed-field gel electrophoresis of yeast chromosomes. Identical samples of yeast chromosomes were electrophoresed in 10 parallel lanes and stained with ethidium bromide. The bands represent chromosomes having sizes ranging from 0.2 Mb (at bottom) to 2.2 Mb (at top). Original gel is about 13 cm wide by 12.5 cm long. (*Source*: Courtesy Bio-Rad Laboratories/CHEF-DR(R)II pulsed-field electrophoresis systems.)

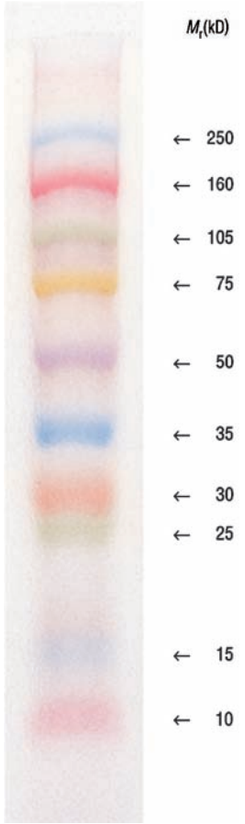


Figure 5.4 SDS-polyacrylamide gel electrophoresis. Polypeptides of the molecular masses shown at right were coupled to dyes and subjected to SDS-PAGE. The dyes allow us to see each polypeptide during and after electrophoresis. (*Source:* Courtesy of Amersham Pharmacia Biotech.)

SUMMARY DNAs, RNAs, and proteins of various masses can be separated by gel electrophoresis. The most common gel used in nucleic acid electrophoresis is agarose, but polyacrylamide is usually used in protein electrophoresis. SDS-PAGE is used to separate polypeptides according to their masses.

Two-Dimensional Gel Electrophoresis

SDS-PAGE gives very good resolution of polypeptides, but sometimes a mixture of polypeptides is so complex that we need an even better method to resolve them all. For example, we may want to separate all of the thousands of polypeptides present at a given time in a given cell type. This is very commonly done now as part of a subfield of molecular biology known as proteomics, which we will discuss in Chapter 24.

To improve on the resolving power of a one-dimensional SDS-PAGE procedure, molecular biologists have developed two-dimensional methods. In one simple method, described in Chapter 19, one can simply run nondenaturing gel electrophoresis (no SDS) in one dimension at one pH and one polyacrylamide gel concentration, then in a second dimension at a second pH and a second polyacrylamide concentration. Proteins will electrophorese at different rates at different pH values because their net charges change with pH. They will also behave differently at different polyacrylamide concentrations according to their sizes. But individual polypeptides cannot be analyzed by this method because the lack of detergent makes it impossible to separate the polypeptides that make up a complex protein.

An even more powerful method is commonly known as **two-dimensional gel electrophoresis,** even though it involves a bit more than the name implies. In the first step, the mixture of proteins is electrophoresed through a narrow tube gel containing molecules called ampholytes that set up a pH gradient from one end of the tube to the other. A negatively charged molecule will electrophorese toward the anode until it reaches its **isoelectric point,** the pH at which it has no net charge. Without net charge, it is no longer drawn toward the anode, or the cathode, for that matter, so it stops. This step is called **isoelectric focusing** because it focuses proteins at their isoelectric points in the gel.

In the second step, the gel is removed from the tube and placed at the top of a slab gel for ordinary SDS-PAGE. Now the proteins that have been partially resolved by isoelectric focusing are further resolved according to their sizes by SDS-PAGE. Figure 5.5 presents two-dimensional gel electrophoresis separations of *E. coli* proteins grown in the presence and absence of benzoic acid. Proteins from the

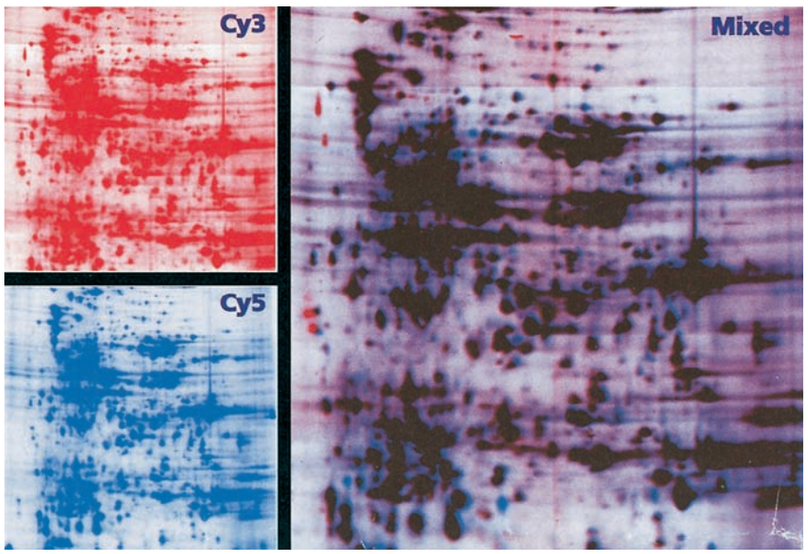


Figure 5.5 Two-dimensional gel electrophoresis. In this experiment, the investigators grew *E. coli* cells in the presence or absence of benzoic acid. Then they stained a lysate of the cells grown in the absence of benzoic acid with the red fluorescent dye Cy3, so the proteins from that lysate would fluoresce red. They stained a lysate of the cells grown in the presence of benzoic acid with the blue fluorescent dye Cy5, so those proteins would fluoresce blue. Finally, they performed two-dimensional gel electrophoresis on **(a)** the proteins from cells grown in the absence of benzoic acid, **(b)** on the proteins grown in the presence of benzoic acid, and **(c)** on a mixture of the two sets of proteins. In panel (c), the proteins that accumulate only in the absence of benzoic acid fluoresce red, those that accumulate only in the presence of benzoic acid fluoresce blue, and those that accumulate under both conditions fluoresce both red and blue, and so appear purple or black. (*Source:* Courtesy of Amersham Pharmacia Biotech.)

cells grown without benzoic acid were stained with the red fluorescent dye Cy3, and proteins from the cells grown with benzoic acid were stained with the blue fluorescent dye Cy5. Two-dimensional gel electrophoresis of these two sets of proteins, separately and together allows us to see which proteins are prevalent in the presence or absence of benzoic acid, and which are prevalent under both conditions.

SUMMARY High-resolution separation of polypeptides can be achieved by two-dimensional gel electrophoresis, which uses isoelectric focusing in the first dimension and SDS-PAGE in the second.

Ion-Exchange Chromatography

Chromatography is a term that originally referred to the pattern one sees after separating colored substances on paper (**paper chromatography**). Nowadays, many different types of chromatography exist for separating biological substances. **Ion-exchange chromatography** uses a resin to separate substances according to their charges. For example, DEAE-Sephadex chromatography uses an ion-exchange resin that contains positively charged diethylaminoethyl (DEAE) groups. These positive charges attract negatively charged substances, including proteins. The greater the negative charge, the tighter the binding.

In Chapter 10, we will see an example of DEAE-Sephadex chromatography in which the experimenters separated three forms of an enzyme called RNA polymerase. They made a slurry of DEAE-Sephadex and poured it into a column. After the resin had packed down, they loaded the sample, a crude cellular extract containing the RNA polymerases. Finally, they **eluted,** or removed, the substances that had bound to the resin in the column by passing a solution of gradually increasing ionic strength (or salt concentration) through the column. The purpose of this salt gradient was to use the negative ions in the salt solution to compete with the proteins for ionic binding sites on the resin, thus removing the proteins one by one. This is why we call it ion-exchange chromatography.

As the ionic strength of the elution buffer increases, samples of solution flowing through the column are collected using a fraction collector. This device works by positioning test tubes, one at a time, beneath the column to collect a given volume of solution. As each tube finishes collecting its fraction of the solution, it moves aside and a new tube moves into position to collect its fraction. Finally, each fraction is assayed (tested) to determine how much of the substance of interest it contains. If the substance is an enzyme, the fractions are assayed for that particular enzyme activity. It is also useful to measure the ionic strength of each fraction to determine what salt concentration is necessary to elute each of the enzymes of interest.

One can also use a negatively charged resin to separate positively charged substances, including proteins. For example, phosphocellulose is commonly used to separate proteins by cation-exchange chromatography. Note that it is not essential for a protein to have a net positive charge to bind to a cation-exchange resin like phosphocellulose. Most proteins have a net negative charge, yet they can still bind to a cation exchange resin if they have a significant center of positive charge. Figure 5.6 depicts the results of a hypothetical ion-exchange chromatography experiment in which two forms of an enzyme are separated.

SUMMARY Ion-exchange chromatography can be used to separate substances, including proteins, according to their charges. Positively charged resins like DEAE-Sephadex are used for anion-exchange chromatography, and negatively charged resins like phosphocellulose are used for cation-exchange chromatography.

Gel Filtration Chromatography

Standard biochemical separations of proteins usually require more than one step, and, because valuable protein is lost at each step, it is important to minimize the number of these steps. One way to do this is to design a strategy that enables each step to take advantage of a different property of the protein of interest. Thus, if anion-exchange chromatography is the first step and cation-exchange chromatography is the second, a third step that separates proteins on some other basis besides charge is needed. Protein size is an obvious next choice.

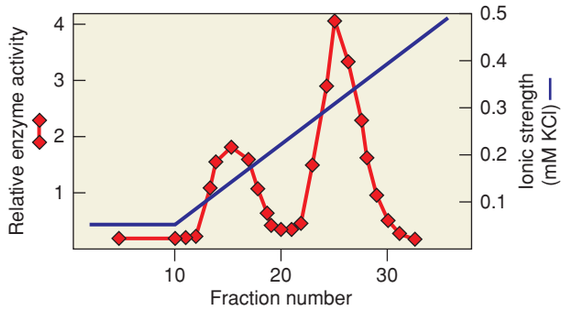
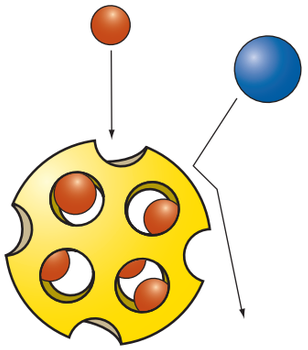
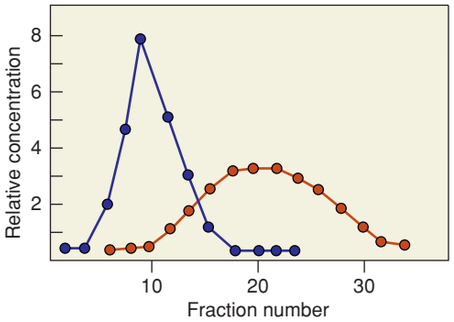


Figure 5.6 Ion-exchange chromatography. Begin by loading a cell extract containing two different forms of an enzyme onto an ion-exchange column. Then pass a buffer of increasing ionic strength through the column and collect fractions (32 fractions in this case). Assay each fraction for enzyme activity (red) and ionic strength (blue), and plot the data as shown. The two forms of the enzyme are clearly separated by this procedure.



(a)



(b)

Figure 5.7 Gel filtration chromatography. (a) Principle of the method. A resin bead is schematically represented as a "whiffle ball" (yellow). Large molecules (blue) cannot fit into the beads, so they are confined to the relatively small buffer volume outside the beads. Thus, they emerge quickly from the column. Small molecules (red), by contrast, can fit into the beads and so have a large buffer volume available to them. Accordingly, they take a longer time to emerge from the column. **(b)** Experimental results. Add a mixture of large and small molecules from panel (a) to the column, and elute them by passing buffer through the column. Collect fractions and assay each for concentration of the large (blue) and small (red) molecules. As expected, the large molecules emerge earlier than the small ones.

Gel filtration chromatography is one method that separates molecules based on their physical dimensions. Gel filtration resins such as Sephadex are porous beads of various sizes that can be likened to "whiffle balls," hollow plastic balls with holes in them. Imagine a column filled with tiny whiffle balls. When one passes a solution containing different size molecules through this column, the small molecules will easily enter the holes in the whiffle balls (the pores in the beads) and therefore flow through the column slowly. On the other hand, large molecules will not be able to enter any of the beads and will flow more quickly through the column. They emerge with the so-called **void volume**—the volume of buffer surrounding the beads, but not included in the beads. Intermediate-size molecules will enter some beads and not others and so will have an intermediate mobility. Thus, large molecules will emerge first from the column, and small molecules will emerge last. Many different resins with different size pores are available for separating different size molecules. Figure 5.7 illustrates this method.

SUMMARY Gel filtration chromatography uses columns filled with porous resins that let in smaller substances, but exclude larger ones. Thus, the smaller substances are slowed in their journey through the column, but larger substances travel relatively rapidly through the column.

Affinity Chromatography

One of the most powerful separation techniques is **affinity chromatography,** in which the resin contains a substance (an *affinity reagent*) to which the molecule of interest has strong and specific affinity. For example, the resin may be coupled to an antibody that recognizes a specific protein, or it may contain an unreactive analog of an enzyme's substrate. In the latter case, the enzyme will bind strongly to the analog, but will not metabolize it. After virtually all the contaminating proteins have flowed through the column because they have no (or weak) affinity for the affinity reagent, the molecule of interest can be eluted from the column using a solution of a substance that competes with binding between the molecule of interest and the affinity reagent. For example, a solution of the enzyme analog could be used. In this case, the analog in solution will compete with the analog on the resin for binding to the enzyme and the enzyme will elute from the column.

The power of affinity chromatography lies in the specificity of binding between the affinity reagent on the resin and the molecule to be purified. Indeed, it is possible to design an affinity chromatography procedure to purify a protein in a single step because that protein is the only one in the cell that will bind to the affinity reagent. In Chapter 4 we saw a good example: the use of a nickel column to purify a protein tagged with oligohistidine. Because all of the other proteins in the cell are natural and are therefore not tagged with oligohistidine, the tagged protein is the only one that will stick to the affinity reagent, nickel. In that case, one could elute the protein from the column with a nickel solution, but that would yield a protein-nickel complex, rather than a pure protein. So investigators use a histidine analog, imidazole, which also disrupts binding between the affinity reagent and the protein of interest—by binding to the nickel on the column.

When the molecule to be purified (e.g., an oligohistidine-tagged protein) is the only one that binds to the affinity resin, column chromatography is not even needed. Instead, the investigator can simply mix the resin with a cell extract, spin down the resin in a centrifuge, throw away the remaining solution (the **supernatant**), leaving the

protein of interest bound to the resin in a pellet at the bottom of the centrifuge tube. After rinsing the pellet with buffer, the protein of interest can be released from the resin (e.g., with a solution of imidazole, if a nickel resin is used), and the resin can be spun down again. This time, the protein of interest will be in the supernatant, which can be removed and saved. This procedure is simpler and faster than traditional chromatography.

SUMMARY Affinity chromatography is a powerful purification technique that exploits an affinity reagent with strong and specific affinity for a molecule of interest. That molecule binds to a column coupled to the affinity reagent but all or most other molecules flow through without binding. Then the molecule of interest can be eluted from the column with a solution of a substance that disrupts the specific binding.

5.2 Labeled Tracers

Until recently, "labeled" has been virtually synonymous with "radioactive" because radioactive tracers have been available for decades, and they are easy to detect. Radioactive tracers allow vanishingly small quantities of substances to be detected. This is important in molecular biology because the substances we are trying to detect in a typical experiment are present in very tiny amounts. Let us assume, for example, that we are attempting to measure the appearance of an RNA product in a transcription reaction. We may have to detect RNA quantities of less than a picogram (pg; only one trillionth of a gram, or 10^{-12} g). Direct measurement of such tiny quantities by UV light absorption or by staining with dyes is not possible because of the limited sensitivities of these methods. On the other hand, if the RNA is radioactive we can measure small amounts of it easily because of the great sensitivity of the equipment used to detect radioactivity. Let us now consider the favorite techniques molecular biologists use to detect radioactive tracers: autoradiography, phosphorimaging, and liquid scintillation counting.

Autoradiography

Autoradiography is a means of detecting radioactive compounds with a photographic emulsion. The form of emulsion favored by molecular biologists is a piece of x-ray film. Figure 5.8 presents an example in which the investigator electrophoreses some radioactive DNA fragments on a gel and then places the gel in contact with the x-ray film and leaves it in the dark for a few hours, or even days. The radioactive emissions from the bands of DNA expose the film, just as visible light would. Thus, when the film is developed, dark bands appear, corresponding to the DNA bands on the gel. In effect, the DNA bands take a picture of themselves, which is why we call this technique *auto*radiography.

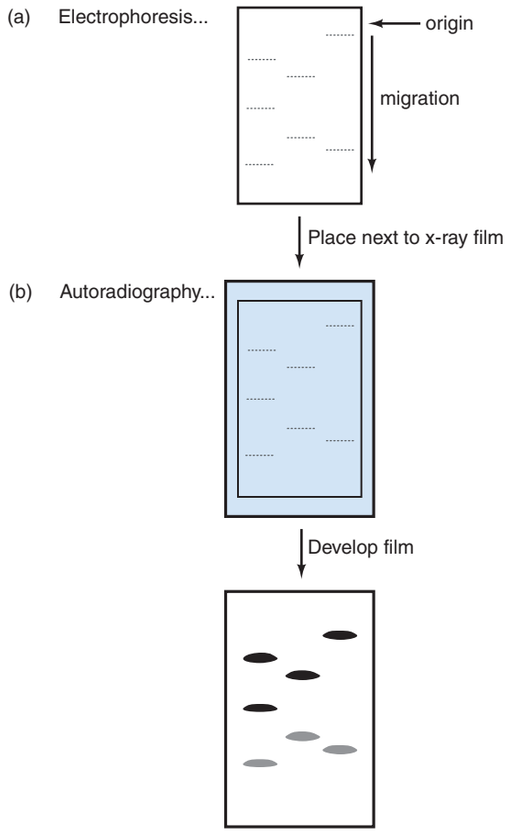


Figure 5.8 Autoradiography. (a) Gel electrophoresis. Electrophorese radioactive DNA fragments in three parallel lanes on a gel, either agarose or polyacrylamide, depending on the sizes of the fragments. At this point the DNA bands are invisible, but their positions are indicated here with dotted lines. **(b)** Autoradiography. Place a piece of x-ray film in contact with the gel and leave it for several hours, or even days if the DNA fragments are only weakly radioactive. Finally, develop the film to see where the radioactivity has exposed the film. This shows the locations of the DNA bands on the gel. In this case, the large, slowly migrating bands are the most radioactive, so the bands on the autoradiograph that correspond to them are the darkest.

To enhance the sensitivity of autoradiography, at least with ^{32}P, one can use an **intensifying screen.** This is a screen coated with a compound that fluoresces when it is excited by β electrons at low temperature. (β electrons are the radioactive emissions from the common radioisotopes used in molecular biology: ^{3}H, ^{14}C, ^{35}S, and ^{32}P.) Thus, one can put a radioactive gel (or other medium) on one side of a photographic film and the intensifying screen on the other. Some β electrons expose the film directly, but others pass right through the film and would be lost without the

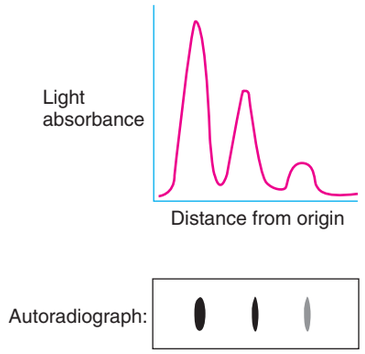


Figure 5.9 Densitometry. An autoradiograph is pictured beneath a densitometer scan of the same film. Notice that the areas under the three peaks of the scan are proportional to the darkness of the corresponding bands on the autoradiograph.

screen. When these high-energy electrons strike the screen, they cause fluorescence, which is detected by the film.

What if the goal is to measure the exact amount of radioactivity in a fragment of DNA? One can get a rough estimate by looking at the intensity of a band on an autoradiograph, and an even better estimate by scanning the autoradiograph with a **densitometer.** This instrument passes a beam of light through a sample—an autoradiograph in this case—and measures the absorbance of that light by the sample. If the band is very dark, it will absorb most of the light, and the densitometer records a large peak of absorbance (Figure 5.9). If the band is faint, most of the light passes through, and the densitometer records only a minor peak of absorbance. By measuring the area under each peak, one can get an estimate of the radioactivity in each band. This is still an indirect measure of radioactivity, however. To get a really accurate reading of the radioactivity in each band, one can scan the gel with a phosphorimager, or subject the DNA to liquid scintillation counting.

Phosphorimaging

The technique of **phosphorimaging** has several advantages over standard autoradiography, but the most important is that it is much more accurate in quantifying the amount of radioactivity in a substance. This is because its response to radioactivity is far more linear than that of an x-ray film. With standard autoradiography, a band with 50,000 radioactive **disintegrations per minute** (**dpm**) may look no darker than one with 10,000 dpm because the emulsion in the film is already saturated at 10,000 dpm. But the **phosphorimager** detects radioactive emissions and analyzes them electronically, so the difference between 10,000 dpm and 50,000 dpm would be obvious. Here is how this technique works: One starts with a radioactive sample—a blot with RNA bands that have hybridized with a labeled probe, for example. This sample is placed

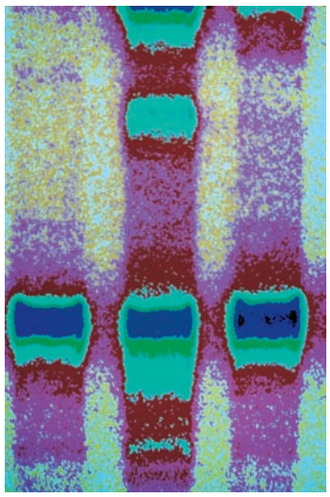


Figure 5.10 False color phosphorimager scan of an RNA blot. After hybridizing a radioactive probe to an RNA blot and washing away unhybridized probe, the blot was exposed to a phosphorimager plate. The plate collected energy from β electrons from the radioactive probe bound to the RNA bands, then gave up this energy when scanned with a laser. A computer converted this energy into an image in which the colors correspond to radiation intensity according to the following color scale: yellow (lowest) < purple < magenta < light blue < green < dark blue < black (highest). (*Source:* © Jay Freis/Image Bank/Getty.)

in contact with a phosphorimager plate, which absorbs β electrons. These electrons excite molecules on the plate, and these molecules remain in an excited state until the phosphorimager scans the plate with a laser. At that point, the β electron energy trapped by the plate is released and monitored by a computerized detector. The computer converts the energy it detects to an image such as the one in Figure 5.10. This is a false color image, in which the different colors represent different degrees of radioactivity, from the lowest (yellow) to the highest (black).

Liquid Scintillation Counting

Liquid scintillation counting uses the radioactive emissions from a sample to create photons of visible light that a photomultiplier tube can detect. To do this, one places the radioactive sample (a band cut out of a gel, for example), into a vial with **scintillation fluid.** This fluid contains a **fluor,** a compound that fluoresces when it is bombarded with radioactive emissions. In effect, it converts the invisible radioactivity into visible light. A liquid scintillation counter is an instrument that lowers the vial into a dark chamber with a photomultiplier tube. There, the tube detects the light resulting from the radioactive emissions exciting the fluor. The instrument counts these bursts of light, or **scintillations,** and records them as **counts per minute** (**cpm**). This is not the same as disintegrations per

minute because the scintillation counter is not 100% efficient. One common radioisotope used by molecular biologists is ^{32}P. The β electrons emitted by this isotope are so energetic that they create photons even without a fluor, so a liquid scintillation counter can count them directly, though at a lower efficiency than with scintillation fluid.

SUMMARY Detection of the tiny quantities of substances used in molecular biology experiments generally requires the use of labeled tracers. If the tracer is radioactive one can detect it by autoradiography, using x-ray film or a phosphorimager, or by liquid scintillation counting.

Nonradioactive Tracers

As we pointed out earlier in this section, the enormous advantage of radioactive tracers is their sensitivity, but now nonradioactive tracers rival the sensitivity of their radioactive forebears. This can be a significant advantage because radioactive substances pose a potential health hazard and must be handled very carefully. Furthermore, radioactive tracers create radioactive waste, and disposal of such waste is increasingly difficult and expensive. How can a nonradioactive tracer compete with the sensitivity of a radioactive one? The answer is, by using the multiplier effect of an enzyme. An enzyme is coupled to a probe that detects the molecule of interest, so the enzyme will produce many molecules of product, thus amplifying the signal. This works especially well if the product of the enzyme is chemiluminescent (light-emitting, like the tail of a firefly), because each molecule emits many photons, amplifying the signal again. Figure 5.11 shows the principle behind one such tracer method. The light can be detected by autoradiography with x-ray film, or by a phosphorimager.

To avoid the expense of a phosphorimager or x-ray film, one can use enzyme substrates that change color instead of becoming chemiluminescent. These **chromogenic substrates** produce colored bands corresponding to the location of the enzyme and, therefore, to the location of the molecule of interest. The intensity of the color is directly related to the amount of that molecule, so this is also a quantitative method.

SUMMARY Some very sensitive nonradioactive labeled tracers are now available. Those that employ chemiluminescence can be detected by autoradiography or by phosphorimaging, just as if they were radioactive. Those that produce colored products can be detected directly, by observing the appearance of colored spots.

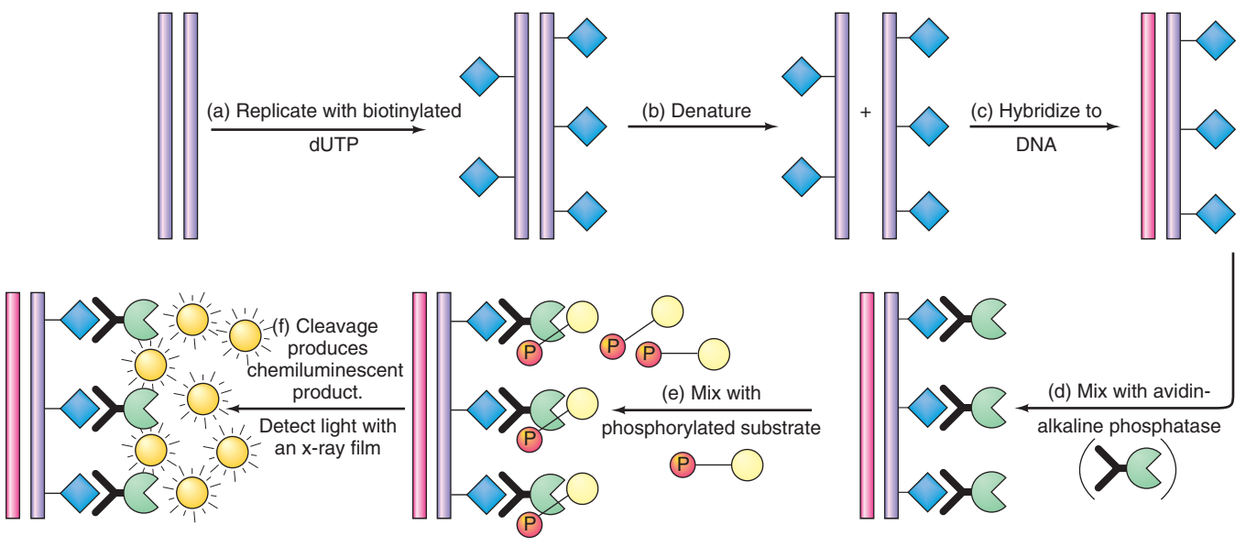


Figure 5.11 Detecting nucleic acids with a nonradioactive probe. This sort of technique is usually indirect; detecting a nucleic acid of interest by hybridization to a labeled probe that can in turn be detected by virtue of its ability to produce a colored or light-emitting substance. In this example, the following steps are executed. **(a)** Replicate the probe DNA in the presence of dUTP that is tagged with the vitamin biotin (blue). This generates biotinylated probe DNA. **(b)** Denature this probe and **(c)** hybridize it to the DNA to be detected (pink). **(d)** Mix the hybrids with a bifunctional reagent containing both avidin and the enzyme alkaline phosphatase (green). The avidin binds tightly and specifically to the biotin in the probe DNA. **(e)** Add a phosphorylated compound that will become chemiluminescent as soon as its phosphate group is removed. **(f)** The alkaline phosphatase enzymes attached to the probe cleave the phosphates from these substrate molecules, rendering them chemiluminescent (light-emitting). The light emitted from the chemiluminescent substrate can be detected with an x-ray film.

5.3 Using Nucleic Acid Hybridization

The phenomenon of hybridization—the ability of one single-stranded nucleic acid to form a double helix with another single strand of complementary base sequence—is one of the backbones of modern molecular biology. We have already encountered plaque and colony hybridization in Chapter 4. Here we will illustrate several further examples of hybridization techniques.

Southern Blots: Identifying Specific DNA Fragments

Many eukaryotic genes are parts of families of closely related genes. How would one determine the number of family members in a particular gene family? If a member of that gene family—even a partial cDNA—has been cloned, one can estimate this number.

One begins by using a restriction enzyme to cut genomic DNA isolated from the organism. It is best to use a restriction enzyme such as *Eco*RI or *Hin*dIII that recognizes a 6-bp cutting site. These enzymes will produce thousands of fragments of genomic DNA, with an average size of about 4000 bp. Next, these fragments are electrophoresed on an agarose gel (Figure 5.12). The result, if the bands are visualized by staining, will be a blurred streak of thousands of bands, none distinguishable from the others (although Figure 5.12, for simplicity's sake shows just a few bands). Eventually, a labeled probe will be hybridized to these bands to see how many of them contain coding sequences for the gene of interest. First, however, the bands are transferred to a medium on which hybridization is more convenient.

Edward Southern was the pioneer of this technique; he transferred, or blotted, DNA fragments from an agarose gel to nitrocellulose by diffusion, as depicted in Figure 5.12. This process has been called **Southern blotting** ever since. Nowadays, blotting is frequently done by electrophoresing the DNA bands out of the gel and onto the blot. Before blotting, the DNA fragments are denatured with alkali so that the resulting single-stranded DNA can bind to the nitrocellulose, forming the Southern blot. Media superior to nitrocellulose are now available; some use nylon supports that are far more flexible than nitrocellulose. Next, the cloned DNA is labeled by adding DNA polymerase to it in the presence of labeled DNA precursors. Then this labeled probe is denatured and hybridized to the Southern blot. Wherever the probe encounters a complementary DNA sequence, it hybridizes, forming a labeled band corresponding to the fragment of DNA containing the gene of interest. Finally, these bands are visualized by autoradiography with x-ray film or by phosphorimaging.

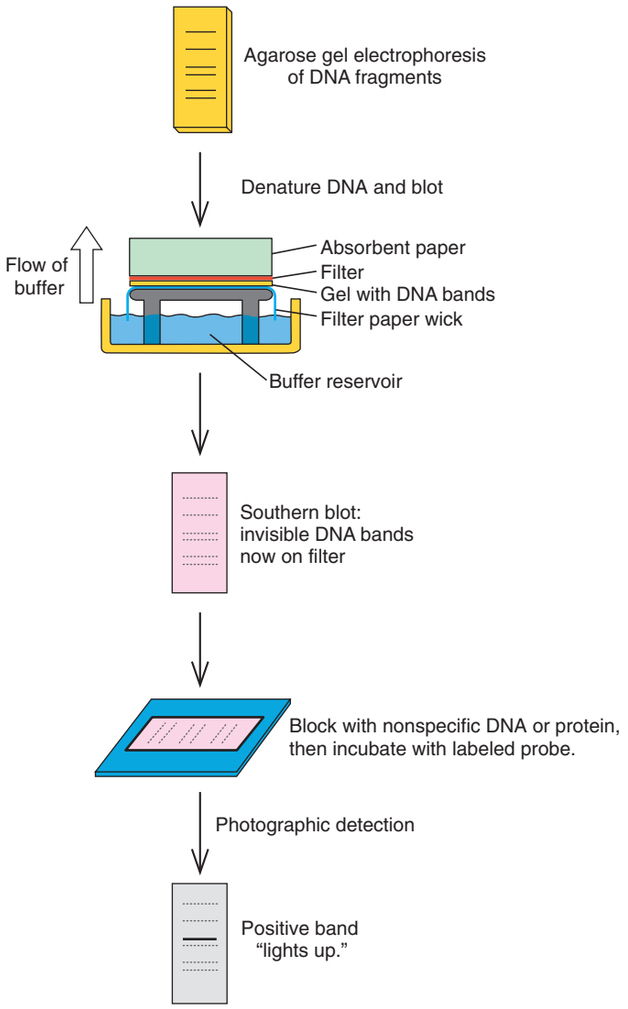


Figure 5.12 Southern blotting. First, electrophorese DNA fragments in an agarose gel. Next, denature the DNA with base and transfer the single-stranded DNA fragments from the gel (yellow) to a sheet of nitrocellulose or another DNA-binding material (red). One can do this in two ways: by diffusion, as shown here, in which buffer passes through the gel, carrying the DNA with it, or by electrophoresis (not shown). Next, hybridize the blot to a labeled probe and detect the labeled bands by autoradiography or phosphorimaging.

If only one band is seen, the interpretation is relatively easy; probably only one gene has a sequence matching the cDNA probe. Alternatively, a gene (e.g., a histone or ribosomal RNA gene) could be repeated over and over again in tandem, with a single restriction site in each copy of the gene. This would yield a single very dark band. If multiple bands are seen, multiple genes are probably present, but it is difficult to tell exactly how many. One gene can give more than one band if it contains one or more cutting sites for the restriction enzyme used. One can minimize this problem by using a short probe, such as a 100–200-bp restriction fragment of the cDNA, for

example. Chances are, a restriction enzyme that cuts on average only every 4000 bp will not cut within the 100–200-bp region of the genes that hybridize to such a probe. If multiple bands are still obtained with a short probe, they probably represent a gene family whose members' sequences are similar or identical in the region that hybridizes to the probe.

SUMMARY Labeled DNA (or RNA) probes can be used to hybridize to DNAs of the same, or very similar, sequence on a Southern blot. The number of bands that hybridize to a short probe gives an estimate of the number of closely related genes in an organism.

DNA Fingerprinting and DNA Typing

Southern blots are not just a research tool. They are widely used in forensic laboratories to identify individuals who have left blood or other DNA-containing material at the scenes of crimes. Such **DNA typing** has its roots in a discovery by Alec Jeffreys and his colleagues in 1985. These workers were investigating a DNA fragment from the gene for a human blood protein, α-globin, when they discovered that this fragment contained a sequence of bases repeated several times. This kind of repeated DNA is called a **minisatellite.** More interestingly, they found similar minisatellite sequences in other places in the human genome, again repeated several times. This simple finding turned out to have far-reaching consequences, because individuals differ in the pattern of repeats of the basic sequence. In fact, they differ enough that two individuals have only a remote chance of having exactly the same pattern. That means that these patterns are like fingerprints; indeed, they are called **DNA fingerprints.**

A DNA fingerprint is really just a Southern blot. To make one, investigators first cut the DNA under study with a restriction enzyme such as *Hae*III. Jeffreys chose this enzyme because the repeated sequence he had found did not contain a *Hae*III recognition site. That means that *Hae*III will cut on either side of the minisatellite regions, but not inside, as shown in Figure 5.13a. In this case, the DNA has three sets of repeated regions, containing four, three, and two repeats, respectively. Thus, three different-size fragments bearing these repeated regions will be produced.

Next, the fragments are electrophoresed, denatured, and blotted. The blot is then probed with a labeled minisatellite DNA, and the labeled bands are detected with x-ray film, or by phosphorimaging. In this case, three labeled bands occur, so three dark bands will appear on the film (Figure 5.13d).

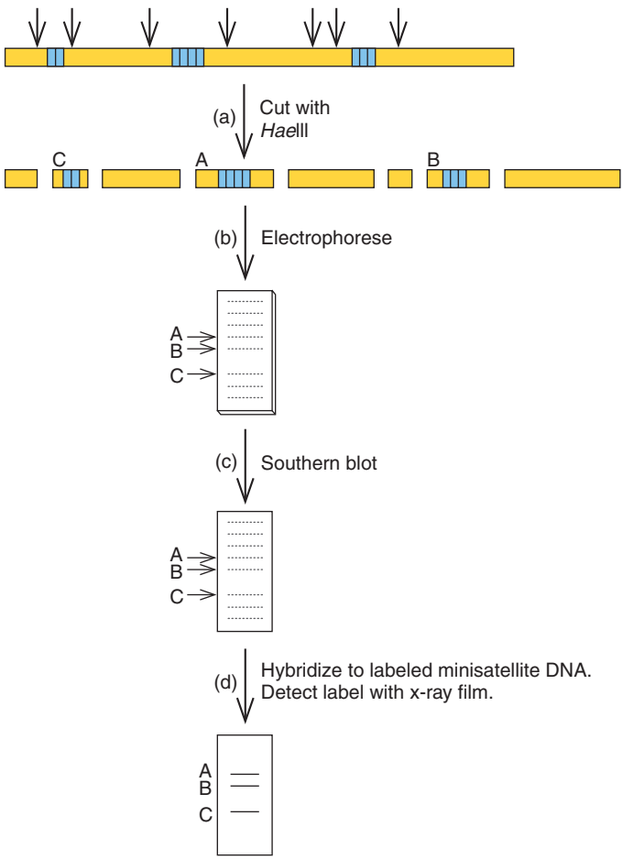


Figure 5.13 DNA fingerprinting. (a) First, cut the DNA with a restriction enzyme. In this case, the enzyme *Hae*III cuts the DNA in seven places (short arrows), generating eight fragments. Only three of these fragments (labeled A, B, and C according to size) contain the minisatellites, represented by blue boxes. The other fragments (yellow) contain unrelated DNA sequences. **(b)** Electrophorese the fragments from part (a), which separates them according to their sizes. All eight fragments are present in the electrophoresis gel, but they remain invisible. The positions of all the fragments, including the three (A, B, and C) with minisatellites are indicated by dotted lines. **(c)** Denature the DNA fragments and Southern blot them. **(d)** Hybridize the DNA fragments on the Southern blot to a radioactive DNA with several copies of the minisatellite. This probe will bind to the three fragments containing the minisatellites, but with no others. Finally, use x-ray film to detect the three labeled bands.

Real animals have a much more complex genome than the simple piece of DNA in this example, so they will have many more than three fragments that contain a minisatellite sequence that will react with the probe. Figure 5.14 shows an example of the DNA fingerprints of several unrelated people and a set of monozygotic twins. As we have already mentioned, this is such a complex pattern of fragments that the patterns for two individuals are extremely unlikely to be identical, unless they come from monozygotic twins. This complexity makes DNA fingerprinting a very powerful identification technique.

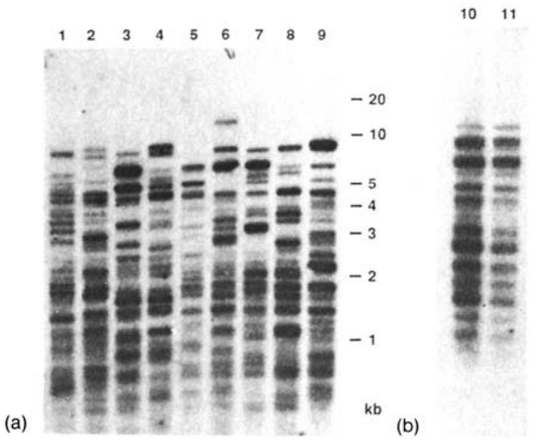


Figure 5.14 DNA fingerprint. (a) The nine lanes contain DNA from nine unrelated white subjects. Note that no two patterns are identical, especially at the upper end. **(b)** The two lanes contain DNA from monozygotic twins, so the patterns are identical (although there is more DNA in lane 10 than in lane 11). (*Source:* G. Vassart et al., A sequence in M13 phage detects hypervariable minisatellites in human and animal DNA. *Science* 235 (6 Feb 1987) p. 683, f. 1. © AAAS.)

Forensic Uses of DNA Fingerprinting and DNA Typing

A valuable feature of DNA fingerprinting is the fact that, although almost all individuals have different patterns, parts of the pattern (sets of bands) are inherited in a Mendelian fashion. Thus, fingerprints can be used to establish parentage. An immigration case in England illustrates the power of this technique. A Ghanaian boy born in England had moved to Ghana to live with his father. When he wanted to return to England to be with his mother, British authorities questioned whether he was a son or a nephew of the woman. Information from blood group genes was equivocal, but DNA fingerprinting of the boy demonstrated that he was indeed her son.

In addition to testing parentage, DNA fingerprinting has the potential to identify criminals. This is because a person's DNA fingerprint is, in principle, unique, just like a traditional fingerprint. Thus, if a criminal leaves some of his cells (blood, semen, or hair, for example) at the scene of a crime, the DNA from these cells can identify him. As Figure 5.14 showed, however, DNA fingerprints are very complex. They contain dozens of bands, some of which smear together, which can make them hard to interpret.

To solve this problem, forensic scientists have developed probes that hybridize to a single DNA locus that varies from one individual to another, rather than to a whole set of DNA loci as in a classical DNA fingerprint. Each probe now gives much simpler patterns, containing only one or a few bands. This is an example of a restriction fragment length polymorphism (RFLP) disussed in detail in Chapter 24. RFLPs occur because the pattern of restriction fragment sizes at a given locus varies from one person to another. Of course, each probe by itself is not as powerful an identification tool as a whole DNA fingerprint with its multitude of bands, but a panel of four or five probes can give enough different bands to be definitive. We sometimes still call such analysis DNA fingerprinting, but a better, more inclusive term is *DNA typing*.

One early, dramatic case of DNA typing involved a man who murdered a man and woman as they slept in a pickup truck, then about forty minutes later went back and raped the woman. This act not only compounded the crime, it also provided forensic scientists with the means to convict the perpetrator. They obtained DNA from the sperm cells in the semen he had left behind, typed it, and showed that the pattern matched that of the suspect's DNA. This evidence helped convince the jury to convict the defendant. Figure 5.15 presents an example of DNA typing that was used to identify another rape suspect. The pattern from the suspect clearly matches that from the sperm DNA. This is the result from only one probe. The others also gave patterns that matched the sperm DNA.

One advantage of DNA typing is its extreme sensitivity. Only a few drops of blood or semen are sufficient to perform a test. However, sometimes forensic scientists have even less to go on—a hair pulled out by the victim, for example. Although the hair by itself may not be enough for DNA typing, it can be useful if it is accompanied by hair follicle cells. Selected segments of DNA from these cells can be amplified by PCR and typed.

In spite of its potential accuracy, DNA typing has sometimes been effectively challenged in court, most famously in the O.J. Simpson trial in Los Angeles in 1995. Defense lawyers have focused on two problems with DNA typing: First, it is tricky and must be performed very carefully to give meaningful results. Second, there has been controversy about the statistics used in analyzing the data. This second question revolves around the use of the product rule in deciding whether the DNA typing result uniquely identifies a suspect. Let us say that a given probe detects a given allele (a set of bands in this case) in one in a hundred people in the general population. Thus, the chance of a match with a given person with this probe is one in a hundred, or 10^{-2}. If we use five probes, and all five alleles match the suspect, we might conclude that the chances of such a match are the product of the chances of a match with each individual probe, or $(10^{-2})^5$ or 10^{-10}. Because fewer than 10^{10} (10 billion) people are now on earth, this would mean this DNA typing would statistically eliminate everyone but the suspect. Prosecutors have used a more conservative estimate that takes into account the fact that members of some ethnic groups have higher probabilities of matches with certain probes. Still, probabilities greater than one in a million are frequently

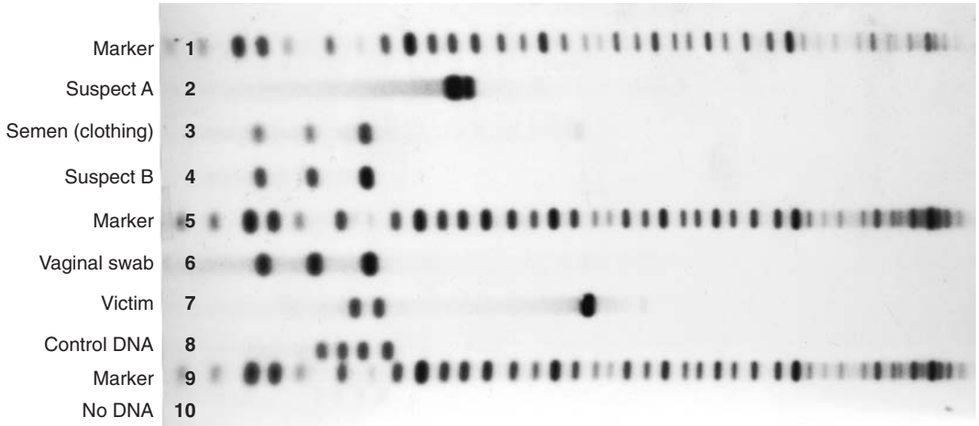


Figure 5.15 Use of DNA typing to help identify a rapist. Two suspects have been accused of attacking and raping a young woman, and DNA analyses have been performed on various samples from the suspects and the woman. Lanes 1, 5, and 9 contain marker DNAs. Lane 2 contains DNA from the blood cells of suspect A. Lane 3 contains DNA from a semen sample found on the woman's clothing. Lane 4 contains DNA from the blood cells of suspect B. Lane 6 contains DNA obtained by swabbing the woman's vaginal canal. (Too little of the victim's own DNA was present to detect.) Lane 7 contains DNA from the woman's blood cells. Lane 8 contains a control DNA. Lane 10 is a control containing no DNA. Partly on the basis of this evidence, suspect B was found guilty of the crime. Note how his DNA fragments in lane 4 match the DNA fragments from the semen in lane 3 and the vaginal swab in lane 6. (*Source:* Courtesy Lifecodes Corporation, Stamford, CT.)

achieved, and they can be quite persuasive in court. Of course, DNA typing can do more than identify criminals. It can just as surely eliminate a suspect (see suspect A in Figure 5.15).

SUMMARY Modern DNA typing uses a battery of DNA probes to detect variable sites in individual animals, including humans. As a forensic tool, DNA typing can be used to test parentage, to identify criminals, or to remove innocent people from suspicion.

In Situ Hybridization: Locating Genes in Chromosomes

This chapter has illustrated the use of probes to identify the band on a Southern blot that contains a gene of interest. Labeled probes can also be used to hybridize to chromosomes and thereby reveal which chromosome has the gene of interest. The strategy of such **in situ hybridization** is to spread the chromosomes from a cell and partially denature the DNA to create single-stranded regions that can hybridize to a labeled probe. One can use x-ray film to detect the label in the spread after it is stained and probed. The stain allows one to visualize and identify the chromosomes, and the darkening of the photographic emulsion locates the labeled probe, and therefore the gene to which it hybridized.

Other means of labeling the probe are also available. Figure 5.16 shows the localization of the muscle glycogen phosphorylase gene to human chromosome 11 using a DNA probe labeled with dinitrophenol, which can be

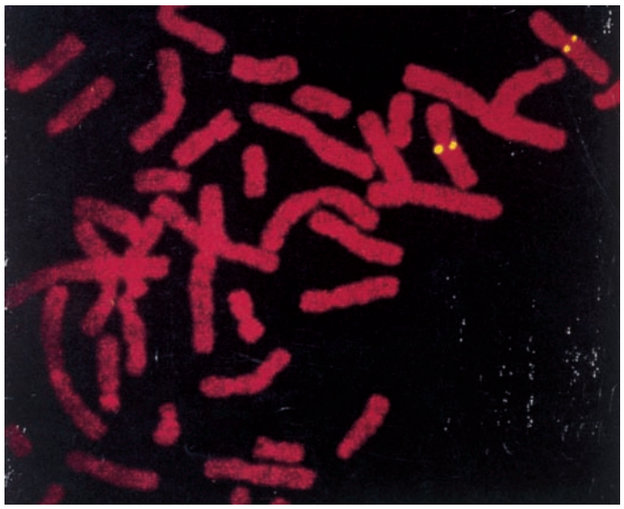


Figure 5.16 Using a fluorescent probe to find a gene in a chromosome by in situ hybridization. A DNA probe specific for the human muscle glycogen phosphorylase gene was coupled to dinitrophenol. A human chromosome spread was then partially denatured to expose single-stranded regions that can hybridize to the probe. The sites where the DNP-labeled probe hybridized were detected indirectly as follows: A rabbit anti-DNP antibody was bound to the DNP on the probe; then a goat antirabbit antibody, coupled with fluorescein isothiocyanate (FITC), which emits yellow fluorescent light, was bound to the rabbit antibody. The chromosomal sites where the probe hybridized show up as bright yellow fluorescent spots against a red background that arises from staining the chromosomes with the fluorescent dye propidium iodide. This analysis identifies chromosome 11 as the site of the glycogen phosphorylase gene. (*Source:* Courtesy Dr. David Ward, *Science* 247 (5 Jan 1990) cover. © AAAS.)

detected with a fluorescent antibody. The chromosomes are counterstained with propidium iodide, so they will fluoresce red. Against this background, the yellow fluorescence of the antibody probe on chromosome 11 is easy to see. This technique is known as **fluorescence in situ hybridization (FISH).**

SUMMARY One can hybridize labeled probes to whole chromosomes to locate genes or other specific DNA sequences. This type of procedure is called in situ hybridization; if the probe is fluorescently labeled, the technique is called fluorescence in situ hybridization (FISH).

Immunoblots (Western Blots)

Immunoblots (also known as **Western blots,** keeping to the Southern nomenclature system) follow the same experimental pattern as Southern blots: The investigator electrophoreses molecules and then blots these molecules to a membrane where they can be identified readily. However, immunoblots involve electrophoresis of proteins instead of nucleic acids. We have seen that DNAs on Southern blots are detected by hybridization to labeled oligonucleotide or polynucleotide probes. But hybridization is appropriate only for nucleic acids, so how are the blotted proteins detected? Instead of a nucleic acid, one uses an antibody (or antiserum) specific for a particular protein. That antibody binds to the target protein on the blot. Then a labeled secondary antibody (for example, a goat antibody that recognizes all rabbit antibodies in the IgG class), or a labeled IgG-binding protein such as protein A, can be used to label the band with the target protein, by binding to the antibody already attached there. (The fact that antibodies are products of the immune system gives rise to the term "immunoblot.") Immunoblots can tell us whether or not a particular protein is present in a mixture, and can also give at least a rough idea of the quantity of that protein.

Why bother with a secondary antibody or protein A; why not just use a labeled primary antibody? The main reason is that this would require individually labeling every different antibody used to probe a series of immunoblots. It is much simpler and cheaper to use unlabeled primary antibody, and buy a stock of labeled secondary antibody or protein A that can bind to and detect any primary antibody. Figure 5.17 illustrates the process of making and probing an immunoblot for a particular protein.

SUMMARY Proteins can be detected and quantified in complex mixtures using immunoblots (or Western blots). Proteins are electrophoresed, then blotted to a membrane and the proteins on the blot are probed with specific antibodies that can be detected with labeled secondary antibodies or protein A.

DNA Sequencing

In 1975, Frederick Sanger and his colleagues, and Alan Maxam and Walter Gilbert developed two different methods for determining the exact base sequence of a cloned piece of DNA. These spectacular breakthroughs revolutionized molecular biology and won the 1980 Nobel prize in chemistry for Gilbert and Sanger. They have allowed molecular biologists to determine the sequences of thousands of genes and many whole genomes, including almost 3 billion base pairs of the human genome. Modern DNA sequencing derives from the Sanger method, so that is the one we will describe here.

The Sanger Chain-Termination Sequencing Method The original method of sequencing a piece of DNA by the

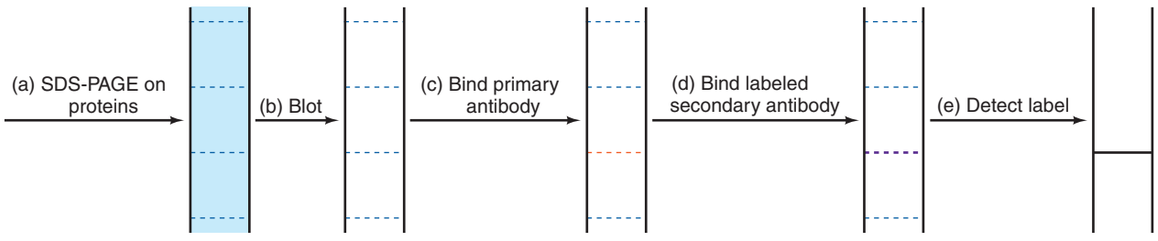


Figure 5.17 Immunoblotting (Western blotting). (a) An immunoblot begins with separation of a mixture of proteins by SDS-PAGE. **(b)** Next, the separated proteins, represented by dotted lines, are blotted to a membrane. **(c)** The blot is probed with a primary antibody specific for a protein of interest on the blot. Here, the antibody has reacted with one of the protein bands (red), but the reaction is undetectable so far. **(d)** A labeled secondary antibody (or protein A) is used to detect the primary antibody, and therefore the protein of interest. Here, the presence of the secondary antibody attached to the primary antibody is denoted by the change in color of the band from red to purple, but this reaction is also undetectable so far. **(e)** Finally, the labeled band is detected—using an x-ray film or a phosphorimager if the label is radioactive. If the label is nonradioactive, it can be detected as described in Figure 5.11.

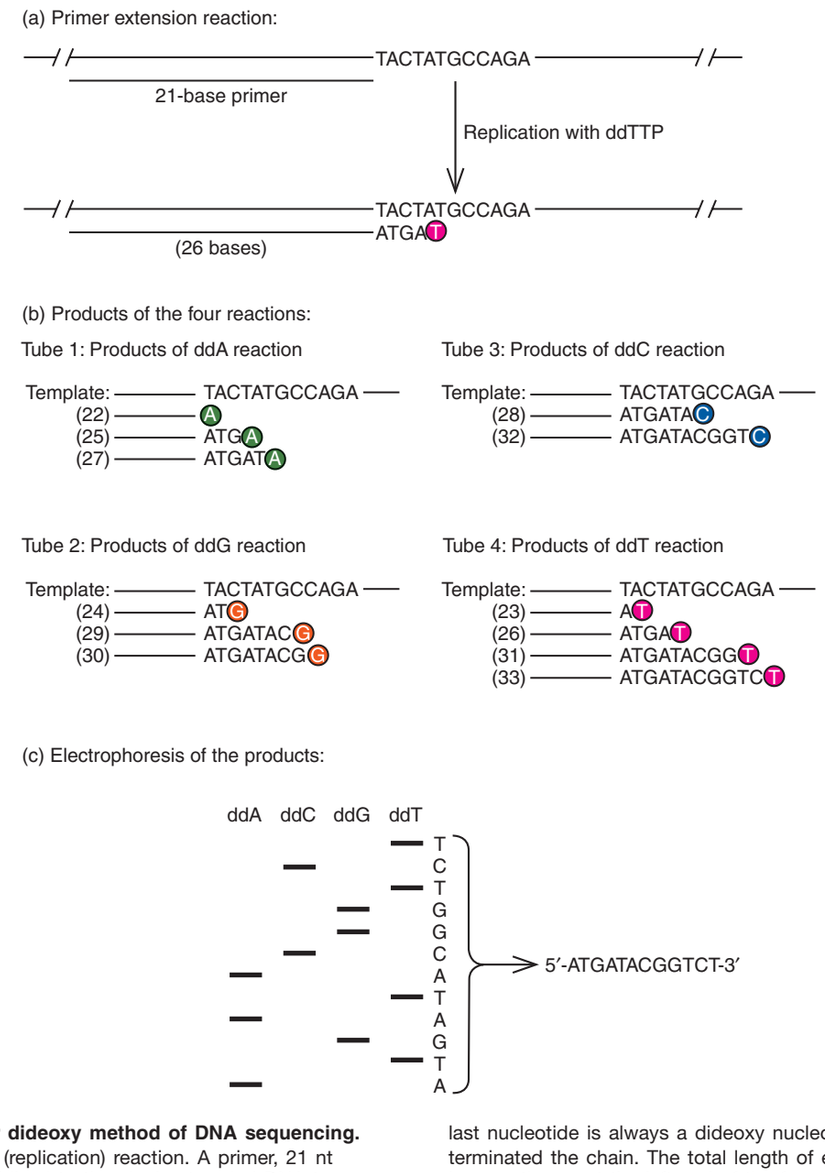


Figure 5.18 The Sanger dideoxy method of DNA sequencing. **(a)** The primer extension (replication) reaction. A primer, 21 nt long in this case, is hybridized to the single-stranded DNA to be sequenced, then mixed with the Klenow fragment of DNA polymerase and dNTPs to allow replication. One dideoxy NTP is included to terminate replication after certain bases; in this case, ddTTP is used, and it has caused termination at the second position where dTTP was called for. **(b)** Products of the four reactions. In each case, the template strand is shown at the top, with the various products underneath. Each product begins with the 21-nt primer and has one or more nucleotides added to the 3′-end. The last nucleotide is always a dideoxy nucleotide (color) that terminated the chain. The total length of each product is given in parentheses at the left end of the fragment. Thus, fragments ranging from 22 to 33 nt long are produced. **(c)** Electrophoresis of the products. The products of the four reactions are loaded into parallel lanes of a high-resolution electrophoresis gel and electrophoresed to separate them according to size. By starting at the bottom and finding the shortest fragment (22 nt in the A lane), then the next shortest (23 nt in the T lane), and so forth, one can read the sequence of the product DNA. Of course, this is the complement of the template strand.

Sanger method (Figure 5.18) is presented here to explain the principles. In practice, it is rarely done manually this way anymore. In the next section we will see how the technique has been automated. The original method began with cloning the DNA into a vector, such as M13 phage or a phagemid, that would give the cloned DNA in single-stranded form. These days, one can start with double-stranded DNA and simply heat it to create single-stranded DNAs for sequencing. To the single-stranded DNA one hybridizes an oligonucleotide primer about 20 bases long. This synthetic primer is designed to hybridize to a sequence adjacent to the

multiple cloning site of the vector and is oriented with its 3′-end pointing toward the insert in the multiple cloning site.

Extending the primer using the Klenow fragment of DNA polymerase (Chapter 20) produces DNA complementary to the insert. The trick to Sanger's method is to carry out such DNA synthesis reactions in four separate tubes and to include in each tube a different chain terminator. The chain terminator is a **dideoxy nucleotide** such as **dideoxy ATP (ddATP).** Not only is this terminator 2′-deoxy, like a normal DNA precursor, it is 3′-deoxy as well. Thus, it cannot form a phosphodiester bond because it lacks the necessary 3′-hydroxyl group. That is why we call it a chain terminator; whenever a dideoxy nucleotide is incorporated into a growing DNA chain, DNA synthesis stops.

Dideoxy nucleotides by themselves do not permit any DNA synthesis at all, so an excess of normal deoxy nucleotides must be used, with just enough dideoxy nucleotide to stop DNA strand extension once in a while at random. This random arrest of DNA growth means that some strands will terminate early, others later. Each tube contains a different dideoxy nucleotide: ddATP in tube 1, so chain termination will occur with A's; ddCTP in tube 2, so chain termination will occur with C's; and so forth. Radioactive dATP is also included in all the tubes so the DNA products will be radioactive.

The result is a series of fragments of different lengths in each tube. In tube 1, all the fragments end in A; in tube 2, all end in C; in tube 3, all end in G; and in tube 4, all end in T. Next, all four reaction mixtures are electrophoresed in parallel lanes in a high-resolution polyacrylamide gel under denaturing conditions, so all DNAs are single-stranded. Finally, autoradiography is performed to visualize the DNA fragments, which appear as horizontal bands on an x-ray film.

Figure 5.18c shows a schematic of the sequencing film. To begin reading the sequence, start at the bottom and find the first band. In this case, it is in the A lane, so you know that this short fragment ends in A. Now move to the next longer fragment, one step up on the film; the gel electrophoresis has such good resolution that it can separate fragments differing by only one base in length, at least until the fragments become much longer than this. And the next fragment, one base longer than the first, is found in the T lane, so it must end in T. Thus, so far you have found the sequence AT. Simply continue reading the sequence in this way as you work up the film. The sequence is shown, reading bottom to top, at the right of the drawing. At first you will be reading just the sequence of part of the multiple cloning site of the vector. However, before very long, the DNA chains will extend into the insert—and unknown territory. An experienced sequencer can continue to read sequence from one film for hundreds of bases.

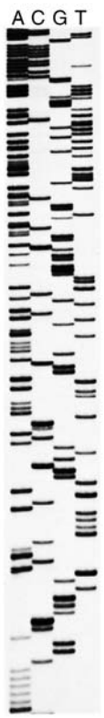


Figure 5.19 A typical sequencing film. The sequence begins CAAAAAACGG. You can probably read the rest of the sequence to the top of the film. (*Source:* Courtesy Life Technologies, Inc., Gaithersburg, MD.)

Figure 5.19 shows a typical sequencing film. The shortest band (at the very bottom) is in the C lane. After that, a series of six bands occurs in the A lane. So the sequence begins CAAAAAA. It is easy to read many more bases on this film; try it yourself.

Automated DNA Sequencing The "manual" sequencing technique just described is powerful, but it is still relatively slow. If one is to sequence a really large amount of DNA, such as the 3 billion base pairs found in the human genome, then rapid, automated sequencing methods are required. Indeed, automated DNA sequencing has been in use for many years. Figure 5.20a describes one such technique, again based on Sanger's chain-termination method. This procedure uses dideoxynucleotides, just as in the manual method, with one important exception. The primers, or, more commonly, the dideoxy nucleotides used in each of the four reactions are tagged with a different fluorescent molecule, so the products from each tube will emit a different color fluorescence when excited by light.

After the extension reactions and chain termination are complete, all four reactions are mixed and electrophoresed together in the same lane on a gel in a short, thin column (Figure 5.20b). Near the bottom of the gel is an analyzer that excites the fluorescent oligonucleotides with a laser beam as they pass by. Then the color of the fluorescent light

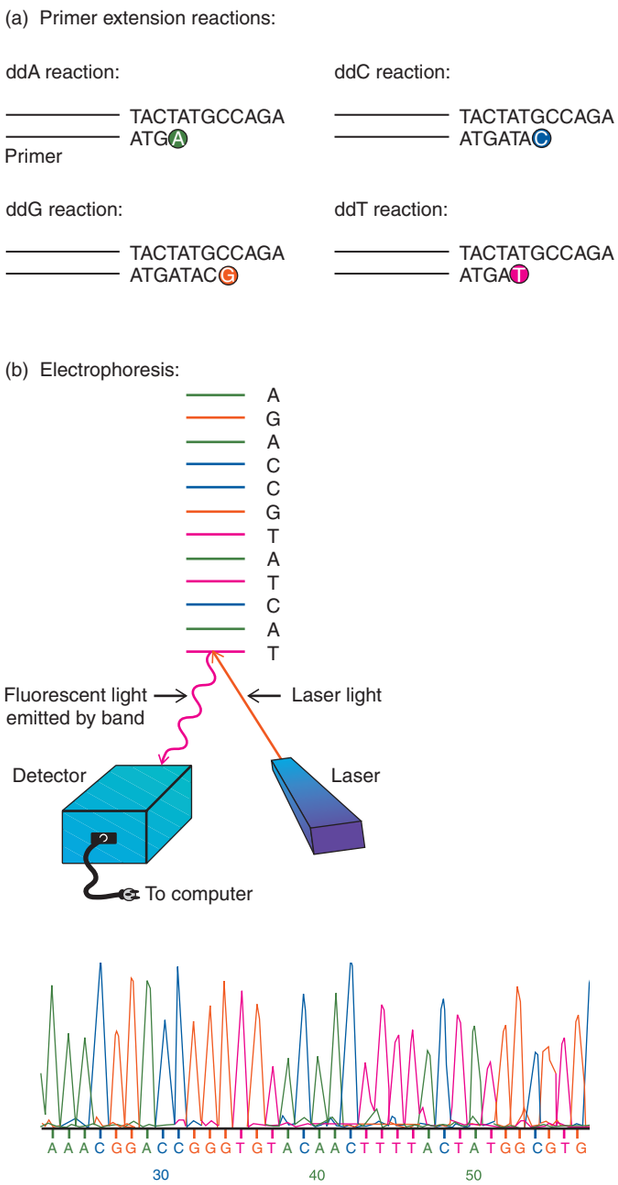


Figure 5.20 Automated DNA sequencing. (a) The primer extension reactions are run in the same way as in the manual method, except that the primers in each reaction are labeled with a different fluorescent molecule that emits light of a distinct color. Only one product is shown for each reaction, but all possible products are actually produced, just as in manual sequencing. **(b)** Electrophoresis and detection of bands. The various primer extension reaction products separate according to size on gel electrophoresis. The bands are color-coded according to the termination reaction that produced them (e.g., green for oligonucleotides ending in ddA, blue for those ending in ddC, and so forth). A laser scanner excites the fluorescent tag on each band as it passes by, and a detector analyzes the color of the resulting emitted light. This information is converted to a sequence of bases and stored by a computer. **(c)** Sample printout of an automated DNA sequencing experiment. Each colored peak is a plot of the fluorescence intensity of a band as it passes through the laser beam. The colors of these peaks, and those of the bands in part (b) and the tags in part (a), were chosen for convenience. They may not correspond to the actual colors of the fluorescent light.

emitted from each oligonucleotide is detected electronically. This information then passes to a computer, which has been programmed to convert the color information to a base sequence. If it "sees" blue, for example, this might mean that this oligonucleotide came from the dideoxy C reaction, and therefore ends in C (actually a ddC). Green may indicate A; orange, G; and red, T. The computer gives a printout of the profile of each passing fluorescent band, color-coded for each base (Figure 5.20c), and stores the sequence of these bases in its memory for later use.

Nowadays, automated sequencers (**sequenators**) may simply print out the sequence or send it directly to a computer for analysis. Large genome projects use many sequenators with 96, or even 384, columns apiece, running simultaneously to obtain millions or even billions of bases of sequence (Chapter 24). One 384-column sequenator can produce 200,000 nt of sequence in one three-hour run.

SUMMARY The Sanger DNA sequencing method uses dideoxy nucleotides to terminate DNA synthesis, yielding a series of DNA fragments whose sizes can be measured by electrophoresis. The last base in each of these fragments is known, because we know which dideoxy nucleotide was used to terminate each reaction. Therefore, ordering these fragments by size—each fragment one (known) base longer than the next—tells us the base sequence of the DNA. Automated sequenators make this process very efficient.

Restriction Mapping

Before sequencing a large stretch of DNA, some preliminary mapping is usually done to locate landmarks on the DNA molecule. These are not genes, but small regions of the DNA—cutting sites for restriction enzymes, for example. A map based on such physical characteristics is called, naturally enough, a **physical map.** (If restriction sites are the only markers involved, we can also call it a **restriction map.**)

To introduce the idea of restriction mapping, let us consider the simple example illustrated in Figure 5.21. We start with a *Hin*dIII fragment 1.6 kb (1600 bp) long (Figure 5.21a). When this fragment is cut with another restriction enzyme (*Bam*HI), two fragments are generated, 1.2 and 0.4 kb long. The sizes of these fragments can be measured by electrophoresis, as pictured in Figure 5.21a. The sizes reveal that *Bam*HI cuts 0.4 kb from one end of the 1.6-kb *Hin*dIII fragment, and 1.2 kb from the other.

Now suppose the 1.6-kb *Hin*dIII fragment is cloned into the *Hin*dIII site of a hypothetical plasmid vector, as illustrated in Figure 5.21b. Because this is not directional cloning, the fragment will insert into the vector in either of the two possible orientations: with the *Bam*HI site on the right (left side of Figure 5.21), or with the *Bam*HI site on the left (right side of the Figure 5.21). How can you determine which orientation exists in a given clone? To answer this question, locate a restriction site asymmetrically situated in the vector, relative to the *Hin*dIII cloning site. In this case, an *Eco*RI site is only 0.3 kb from the *Hin*dIII site. This means that if you cut the cloned DNA pictured on the left with *Bam*HI and *Eco*RI, you will generate two fragments: 3.6 and 0.7 kb long. On the other hand, if you cut the DNA pictured on the right with the same two enzymes, you will generate two fragments: 2.8 and 1.5 kb in size. You can distinguish between these two possibilities easily by electrophoresing the fragments to measure their sizes, as shown at the bottom of Figure 5.21. Usually, DNA is prepared from several different clones, each of them is cut with the two enzymes, and the fragments are electrophoresed side by side with one lane reserved for marker fragments of known sizes. On average, half of the clones will have one orientation, and the other half will have the opposite orientation. You can use similar reasoning to solve the following mapping problem.

Solved Problem

Problem 2

You are mapping a double-stranded circular DNA 17 kb in size. As shown in Figure 5.22a, *Hin*dIII cuts this DNA twice, yielding two fragments (*Hin*dIII-A and *Hin*dIII-B) with sizes of 11 and 6 kb, respectively. *Pst*I also cuts twice, yielding two fragments (*Pst*I-A and *Pst*I-B) with sizes of 12 and 5 kb, respectively. Figure 5.22b shows the cleavage patterns of each of these four fragments cut with the other restriction enzyme. For example, the *Hin*dIII-A fragment can be cleaved with *Pst*I to yield two fragments: 7 and 4 kb in size.

Show the locations of all four restriction sites on the circular DNA. To simplify matters, start by placing the *Hin*dIII-B fragment in the upper right quadrant of the circle, with one of the *Hin*dIII sites at the top of the circle (at 12 o'clock), and the other *Hin*dIII site at about 4 o'clock. Also, place the *Pst*I site in this fragment closer to the 4 o'clock end.

Solution

Two *Hin*dIII fragments can be arranged in a circle in only two possible ways. We just need to decide which of the two is correct. We have already decided arbitrarily to place the smaller of the two *Hin*dIII fragments (*Hin*dIII-B) in the top right quadrant of the circle, oriented with its *Pst*l site nearer the 4 o'clock end. The only question remaining is the orientation of the *Hin*dIII-A fragment. Should its *Pst*I site be closer to the 12 o'clock or 4 o'clock end of the *Hin*dIII-B fragment? To answer this question, we just need to consider what size *Pst*I fragments will result. Suppose we place *Hin*dIII-A with its *Pst*I site closer

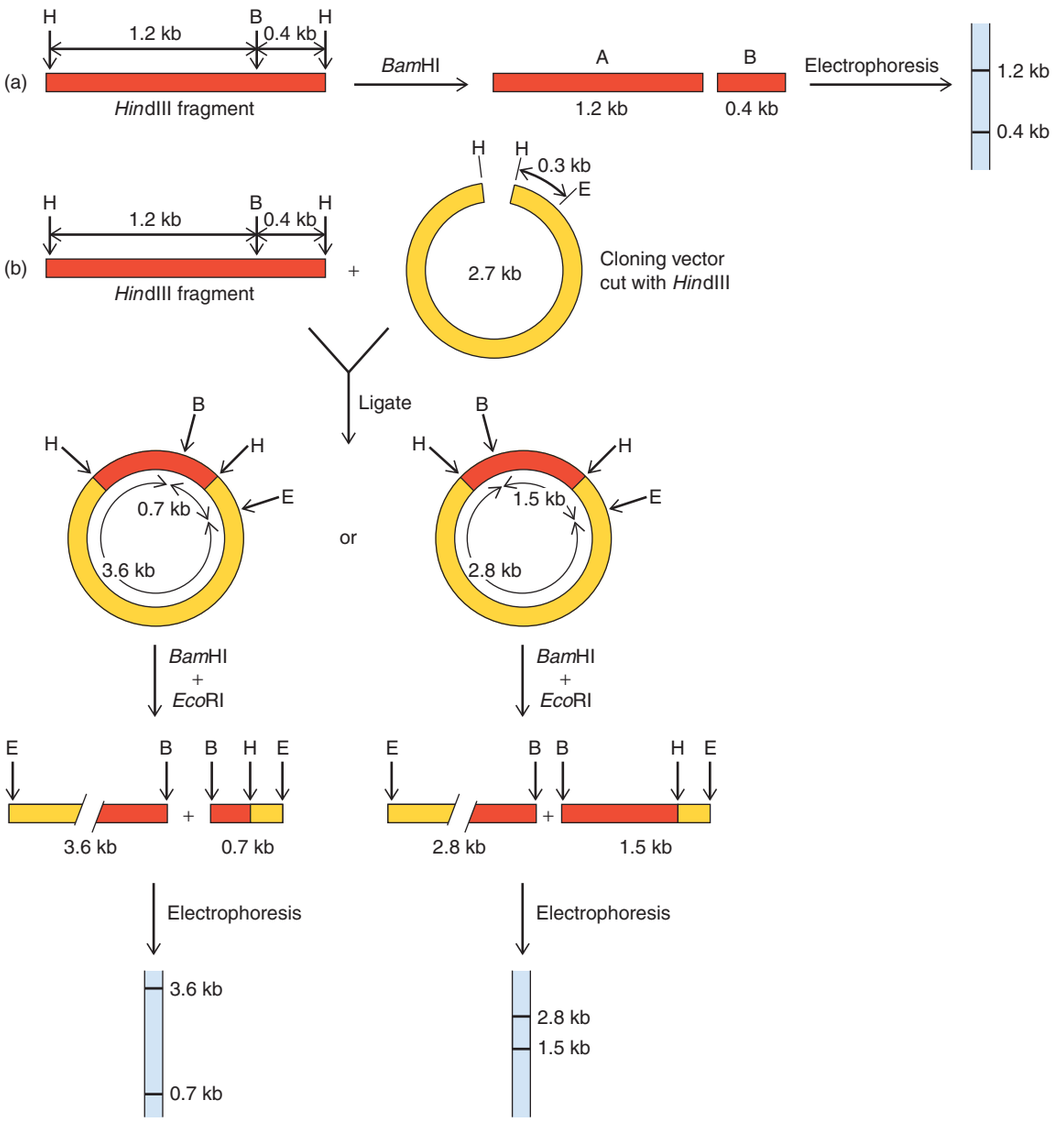


Figure 5.21 A simple restriction mapping experiment. **(a)** Determining the position of a *Bam*HI site. A 1.6-kb *Hin*dIII fragment is cut by *Bam*HI to yield two subfragments. The sizes of these fragments are determined by electrophoresis to be 1.2 kb and 0.4 kb, demonstrating that *Bam*HI cuts once, 1.2 kb from one end of the *Hin*dIII fragment and 0.4 kb from the other end. **(b)** Determining the orientation of the *Hin*dIII fragment in a cloning vector. The 1.6-kb *Hin*dIII fragment can be inserted into the *Hin*dIII site of a cloning vector, in either of two ways: (1) with the *Bam*HI site near an *Eco*RI site in the vector or (2) with the *Bam*HI site remote from an *Eco*RI site in the vector. To determine which, cleave the DNA with both *Bam*HI and *Eco*RI and electrophorese the products to measure their sizes. A short fragment (0.7 kb) shows that the two sites are close together (left). On the other hand, a long fragment (1.5 kb) shows that the two sites are far apart (right).

to 12 o'clock, as shown in Figure 5.23a. Because this site is 4 kb from the *Hin*dIII site, and the *Pst*I site of *Hin*dIII-B is another 5 kb from the *Hin*dIII site, the *Pst*I cleavage of the plasmid would generate a 9-kb fragment. Furthermore, because the whole plasmid is 17 kb, the other *Pst*I fragment would be 17 − 9 = 8 kb. But this is not the result we observe. Therefore, the other possibility must be correct. Let us demonstrate that it is. If the *Hin*dIII-A fragment has its *Pst*I site closer to the 4 o'clock *Hin*dIII site, as in Figure 5.23b, then cleavage of the plasmid with *Pst*I will generate fragments of 12 (7 + 5) and 5 (4 + 1) kb. This is exactly what we observe, so this must be the correct arrangement. ■

These examples are relatively simple, but we use the same kind of logic to solve much more complex mapping

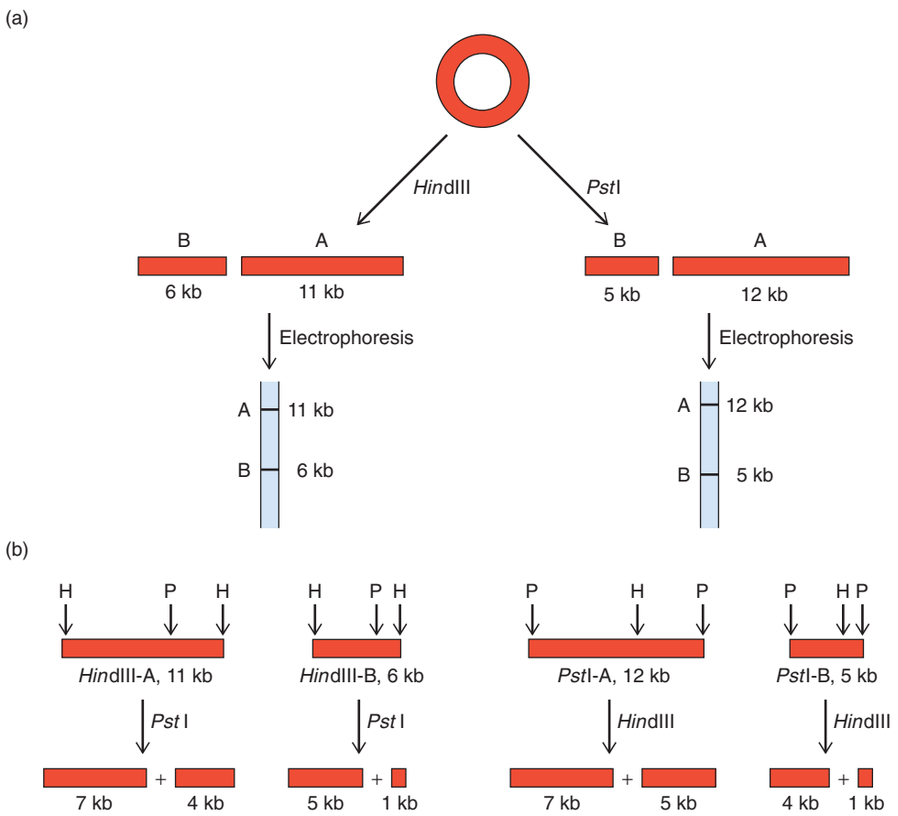


Figure 5.22 Restriction mapping of an unknown DNA. *Abbreviations:* H = *Hin*dIII site; P = *Pst*I site.

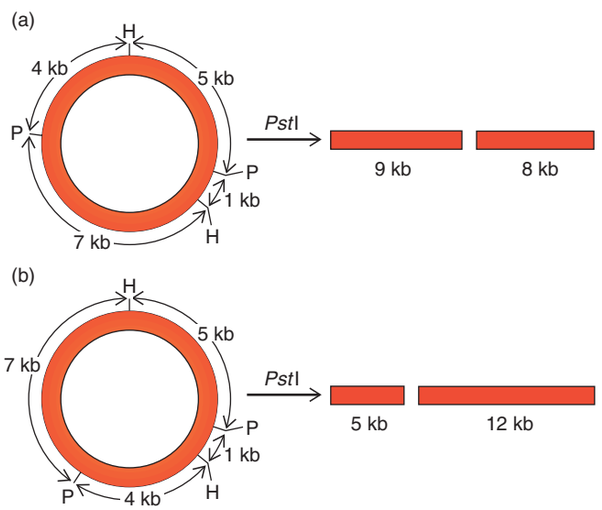


Figure 5.23 Two potential maps of the unknown DNA.
Abbreviations: H = *Hin*dIII site; P = *Pst*I site.

problems. Sometimes it helps to label (radioactively or nonradioactively) one restriction fragment and hybridize it to a Southern blot of fragments made with another restriction enzyme to help sort out the relationships among fragments. For example, consider the linear DNA in Figure 5.24. We might be able to figure out the order of restriction sites without the use of hybridization, but it is not simple. Consider the information we get from just a few hybridizations. If we Southern blot the *Eco*RI fragments and hybridize them to the labeled *Bam*HI-A fragment, for example, the *Eco*RI-A and *Eco*RI-C fragments will become labeled. This demonstrates that *Bam*HI-A overlaps these two *Eco*RI fragments. If we hybridize the blot to the *Bam*HI-B fragment, the *Eco*RI-A and *Eco*RI-D fragments become labeled. Thus, *Bam*HI-B overlaps *Eco*RI-A and *Eco*RI-D. Ultimately, we will discover that no other *Bam*HI fragments besides A and B hybridize to *Eco*RI-A, so *Bam*HI-A and *Bam*HI-B must be adjacent. Using this kind of approach, we can piece together the physical map of the whole 30-kb fragment.

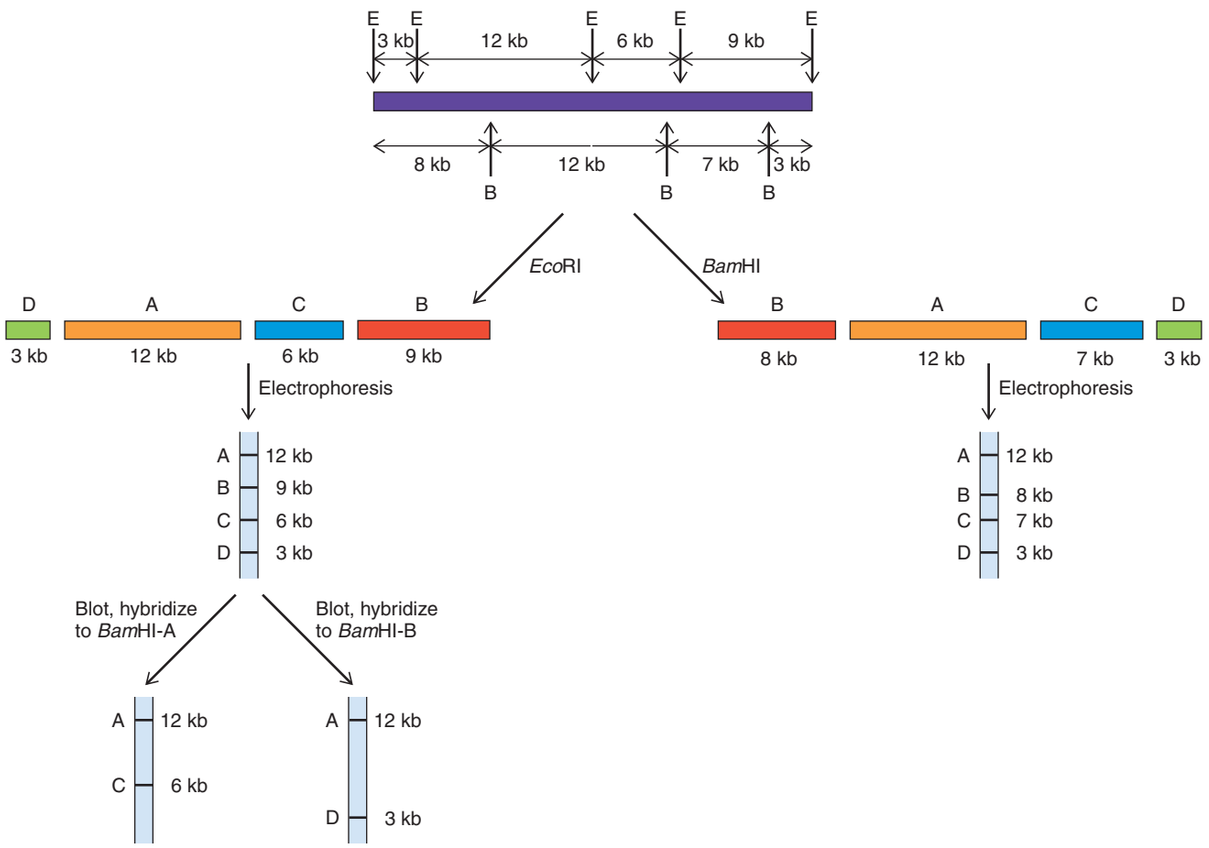


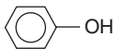
Figure 5.24 Using Southern blots in restriction mapping. A 30-kb fragment is being mapped. It is cut three times each by *Eco*RI (E) and *Bam*HI (B). To aid in the mapping, first cut with *Eco*RI, and electrophorese the four resulting fragments (*Eco*RI-A, -B, -C, and -D); next, Southern blot the fragments and hybridize them to labeled, cloned *Bam*HI-A and -B fragments. The results, shown at lower left, demonstrate that the *Bam*HI-A fragment overlaps *Eco*RI-A and -C, and the *Bam*HI-B fragment overlaps *Eco*RI-A and -D. This kind of information, coupled with digestion of *Eco*RI fragments by *Bam*HI (and vice versa), allows the whole restriction map to be pieced together.

SUMMARY A physical map tells us about the spatial arrangement of physical "landmarks," such as restriction sites, on a DNA molecule. One important strategy in restriction mapping (mapping of restriction sites) is to cut the DNA in question with two or more restriction enzymes in separate reactions, measure the sizes of the resulting fragments, then cut each with another restriction enzyme and measure the sizes of the subfragments by gel electrophoresis. These sizes allow us to locate at least some of the recognition sites relative to the others. We can improve this process considerably by Southern blotting some of the fragments and then hybridizing these fragments to labeled fragments generated by another restriction enzyme. This strategy reveals overlaps between individual restriction fragments.

Protein Engineering with Cloned Genes: Site-Directed Mutagenesis

Traditionally, protein biochemists relied on chemical methods to alter certain amino acids in the proteins they studied, so they could observe the effects of these changes on protein activities. But chemicals are rather crude tools for manipulating proteins; it is difficult to be sure that only one amino acid, or even one kind of amino acid, has been altered. Cloned genes make this sort of investigation much more precise, allowing us to perform microsurgery on a protein. By changing specific bases in a gene, we also change amino acids at corresponding sites in the protein product. Then we can observe the effects of those changes on the protein's function.

Let us suppose that we have a cloned gene in which we want to change a single codon. In particular, the gene codes for a sequence of amino acids that includes a tyrosine. The amino acid tyrosine contains a phenolic group:



To investigate the importance of this phenolic group, we can change the tyrosine codon to a phenylalanine codon. Phenylalanine is just like tyrosine except that it lacks the phenolic group; instead, it has a simple phenyl group:

If the tyrosine phenolic group is important to a protein's activity, replacing it with phenylalanine's phenyl group should diminish that activity.

In this example, let us assume that we want to change the DNA codon TAC (Tyr) to TTC (Phe). How do we perform such **site-directed mutagenesis?** A popular technique, depicted in Figure 5.25 relies on PCR (Chapter 4). We begin with a cloned gene containing a tyrosine codon (TAC) that we want to change to a phenylalanine codon (TTC). The CH_3 symbols indicate that this DNA, like DNAs isolated from most strains of *E. coli,* is methylated on 5′-GATC-3′ sequences. This methylated sequence happens to be the recognition site for the restriction enzyme *Dpn*I, which will come into play later in this procedure. Two methylated *Dpn*I sites are shown, even though many more are usually present because GATC occurs about once every 250 bp in a random sequence of DNA.

The first step is to denature the DNA by heating. The second step is to hybridize mutagenic primers to the DNA. One of these primers is a 25-base oligonucleotide (a 25-mer) with the following sequence:

3′-CGAGTCTGCCAAAGCATGTATAGTA-5′

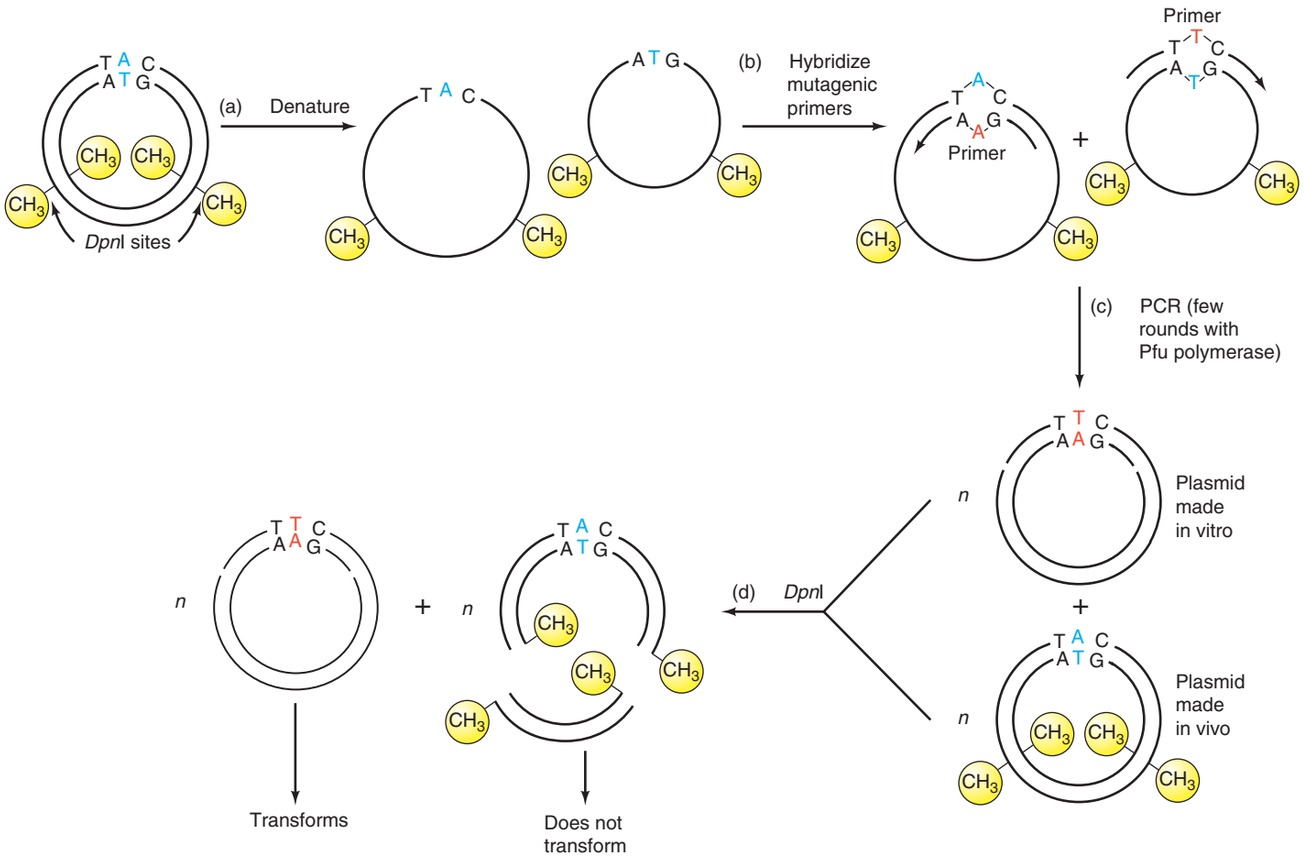


Figure 5.25 PCR-based site-directed mutagenesis. Begin with a plasmid containing a gene with a TAC tyrosine codon that is to be altered to a TTC phenylalanine codon. Thus, the A–T pair (blue) in the original must be changed to a T–A pair. This plasmid was isolated from a normal strain of *E. coli* that methylates the A's of GATC sequences (*Dpn*I sites). The methyl groups are indicated in yellow. **(a)** Heat the plasmid to separate its strands. The strands of the original plasmid are intertwined, so they don't completely separate. They are shown here separating completely for simplicity's sake. **(b)** Hybridize mutagenic primers that contain the TTC codon, or its reverse complement, GAA, to the single-stranded DNA. The altered base in each primer is indicated in red. **(c)** Perform a few rounds of PCR (about eight) with the mutagenic primers to amplify the plasmid with its altered codon. Use a faithful, heat-stable DNA polymerase, such as Pfu polymerase, to minimize mistakes in copying the plasmid. **(d)** Treat the DNA in the PCR reaction with *Dpn*I to digest the methylated wild-type DNA. Because the PCR product was made in vitro, it is not methylated and is not cut. Finally, transform *E. coli* cells with the treated DNA. In principle, only the mutated DNA survives to transform. Check this by sequencing the plasmid DNA from several clones.

This primer was designed to have the same sequence as a piece of the gene's nontemplate strand, except that the central triplet has been changed from ATG to AAG, with the altered base underlined. The other primer is the complementary 25-mer. Both primers incorporate the altered base to change the codon we are targeting. The third step is to use a few rounds of PCR with these primers to amplify the DNA, and incorporate the change we want to make. We deliberately use just a few rounds of PCR to minimize other mutations that might creep in by accident during DNA replication. For the same reason, we use a very faithful DNA polymerase called Pfu polymerase. This enzyme is purified from archaea called *Pyrococcus furiosus* (Latin: *furious fireball*), which live in the boiling hot water surrounding undersea vents. It has the ability to "proofread" the DNA it synthesizes, so it makes relatively few mistakes. A similar enzyme from another hyperthermophilic (extreme heat-loving) archeon is called vent polymerase.

Once the mutated DNA is made, we must either separate it from the remaining wild-type DNA or destroy the latter. This is where the methylation of the wild-type DNA comes in handy. *Dpn*I will cut only GATC sites that are methylated. Because the wild-type DNA is methylated, but the mutated DNA, which was made in vitro, is not, only the wild-type DNA will be cut. Once cut, it is no longer capable of transforming *E. coli* cells, so the mutated DNA is the only species that yields clones. We can check the sequence of DNA from several clones to make sure it is the mutated sequence and not the original, wild-type sequence. Usually, it is mutated.

SUMMARY Using cloned genes, we can introduce changes at will, thus altering the amino acid sequences of the protein products. The mutagenized DNA can be made with double-stranded DNA, two complementary mutagenic primers, and PCR. Simply digesting the PCR product with *Dpn*I removes almost all of the wild-type DNA, so cells can be transformed primarily with mutagenized DNA.

5.4 Mapping and Quantifying Transcripts

One recurring theme in molecular biology has been mapping transcripts (locating their starting and stopping points) and quantifying them (measuring how much of a transcript exists at a certain time). Molecular biologists use a variety of techniques to map and quantify transcripts, and we will encounter several in this book.

You might think that the simplest way of finding out how much transcript is made at a given time would be to label the transcript by allowing it to incorporate labeled nucleotides in vivo or in vitro, then to electrophorese it and detect the transcript as a band on the electrophoretic gel by autoradiography. In fact, this has been done for certain transcripts, both in vivo and in vitro. However, it works in vivo only if the transcript in question is quite abundant and easy to separate from other RNAs by electrophoresis. Transfer RNA and 5S ribosomal RNA satisfy both these conditions and their synthesis has been traced in vivo by simple electrophoresis (Chapter 10). This direct method succeeds in vitro only if the transcript has a clear-cut terminator, so a discrete species of RNA is made, rather than a continuum of species with different 3′-ends that would produce an unintelligible smear, rather than a sharp band. Again, in some instances this is true, most notably in the case of prokaryotic transcripts, but eukaryotic examples are rare. Thus, we frequently need to turn to other, less direct, but more specific methods. Several popular techniques are available for mapping the 5′-ends of transcripts, and one of these also locates the 3′-end. Some of them can also tell how much of a given transcript is in a cell at a given time. These methods rely on the power of nucleic acid hybridization to detect just one kind of RNA among thousands.

Northern Blots

Suppose you have cloned a cDNA (a DNA copy of an RNA) and want to know how actively the corresponding gene (gene X) is expressed in a number of different tissues of organism Y. You could answer that question in several ways, but the method we describe here will also tell the size of the mRNA the gene produces.

You would begin by collecting RNA from several tissues of the organism in question. Then you electrophorese these RNAs in an agarose gel and blot them to a suitable support. Because a similar blot of DNA is called a Southern blot, it was natural to name a blot of RNA a **Northern blot.**

Next, you hybridize the Northern blot to a labeled cDNA probe. Wherever an mRNA complementary to the probe exists on the blot, hybridization will occur, resulting in a labeled band that you can detect with x-ray film. If you run marker RNAs of known size next to the unknown RNAs, you can tell the sizes of the RNA bands that "light up" when hybridized to the probe.

Furthermore, the Northern blot tells you how abundant the gene X transcript is. The more RNA the band contains, the more probe it will bind and the darker the band will be on the film. You can quantify this darkness by measuring the amount of light it absorbs in a densitometer. Or you can quantify the amount of label in the band directly by phosphorimaging. Figure 5.26 shows a Northern blot of RNA from eight different rat tissues, hybridized to a probe for a gene encoding G3PDH (glyceraldehyde-3-phosphate dehydrogenase), which is involved in sugar metabolism. Clearly, transcripts of this gene are most abundant in the heart and skeletal muscle, and least abundant in the lung.

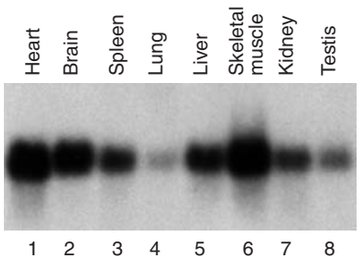


Figure 5.26 A Northern blot. Cytoplasmic mRNA was isolated from the rat tissues indicated at the top, then equal amounts of RNA from each tissue were electrophoresed and Northern blotted. The RNAs on the blot were hybridized to a labeled probe for the rat glyceraldehyde-3-phosphate dehydrogenase (G3PDH) gene, and the blot was then exposed to x-ray film. The bands represent the G3PDH mRNA, and their intensities are indicative of the amounts of this mRNA in each tissue. (*Source:* Courtesy Clontech.)

SUMMARY A Northern blot is similar to a Southern blot, but it contains electrophoretically separated RNAs instead of DNAs. The RNAs on the blot can be detected by hybridizing them to a labeled probe. The intensities of the bands reveal the relative amounts of specific RNA in each.

S1 Mapping

S1 Mapping is used to locate the 5′- or 3′-ends of RNAs and to quantify the amount of a given RNA in cells at a given time. The principle behind this method is to label a single-stranded DNA probe that can hybridize only to the transcript of interest. The probe must span the sequence where the transcript starts or ends. After hybridizing the probe to the transcript, one applies **S1 nuclease,** which degrades only single-stranded DNA and RNA. Thus, the transcript protects part of the probe from degradation. The size of this part can be measured by gel electrophoresis, and the extent of protection tells where the transcript starts or ends. Figure 5.27 shows in detail how S1 mapping can be

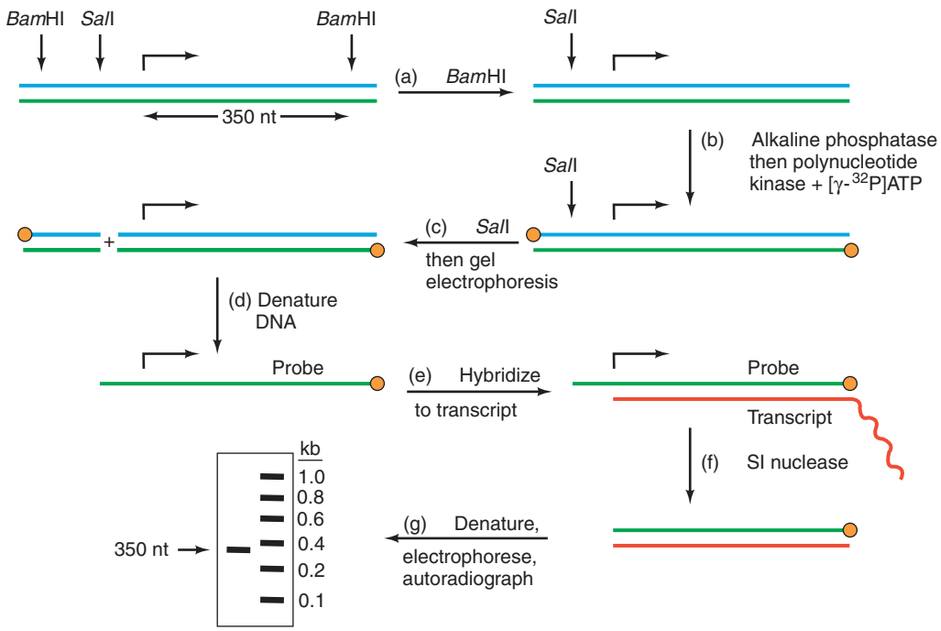


Figure 5.27 S1 mapping the 5′-end of a transcript. Begin with a cloned piece of double-stranded DNA with several known restriction sites. In this case, the exact position of the transcription start site is not known, even though it is marked here (⌐) based on what will be learned from the S1 mapping. It is known that the transcription start site is flanked by two *Bam*HI sites, and a single *Sal*I site occurs upstream of the start site. In step **(a)** cut with *Bam*HI to produce the *Bam*HI fragment shown at upper right. In step **(b)** remove the unlabeled phosphates on this fragment's 5′-hydroxyl groups, then label these 5′-ends with polynucleotide kinase and [γ^{32}P]ATP. The orange circles denote the labeled ends. In step **(c)** cut with *Sal*I and separate the two resulting fragments by electrophoresis. This removes the label from the left end of the double-stranded DNA. In step **(d)** denature the DNA to generate a single-stranded probe that can hybridize with the transcript (red) in step **(e).** In step **(f)** treat the hybrid with S1 nuclease. This digests the single-stranded DNA on the left and the single-stranded RNA on the right of the hybrid from step (e). In step **(g)** denature the remaining hybrid and electrophorese the protected piece of the probe to see how long it is. DNA fragments of known length are included as markers in a separate lane. The length of the protected probe indicates the position of the transcription start site. In this case, it is 350 bp upstream of the labeled *Bam*HI site in the probe.

used to find the transcription start site. First, the DNA probe is labeled at its 5′-end with ^{32}P-phosphate. The 5′-end of a DNA strand usually already contains a nonradioactive phosphate, so this phosphate is removed with an enzyme called alkaline phosphatase before the labeled phosphate is added. Then the enzyme polynucleotide kinase is used to transfer the ^{32}P-phosphate group from [γ-^{32}P]ATP to the 5′-hydroxyl group at the beginning of the DNA strand.

In this example, a *Bam*HI fragment has been labeled on both ends, which would yield two labeled single-stranded probes. However, this would needlessly confuse the analysis, so the label on the left end must be removed. That task is accomplished here by recutting the DNA with another restriction enzyme, *Sal*I, then using gel electrophoresis to separate the short, left-hand fragment from the long fragment that will produce the probe. Now the double-stranded DNA is labeled on only one end, and it can be denatured to yield a labeled single-stranded probe. Next, the probe DNA is hybridized to a mixture of cellular RNAs that contains the transcript of interest. Hybridization between the probe and the complementary transcript will leave a tail of single-stranded DNA on the left, and single-stranded RNA on the right. Next, S1 nuclease is used. This enzyme specifically degrades single-stranded DNA or RNA, but leaves double-stranded polynucleotides, including RNA–DNA hybrids, intact. Thus, the part of the DNA probe, including the terminal label, that is hybridized to the transcript will be protected. Finally, the length of the protected part of the probe is determined by high-resolution gel electrophoresis alongside marker DNA fragments of known length. Because the location of the right-hand end of the probe (the labeled *Bam*HI site) is known exactly, the length of the protected probe automatically tells the location of the left-hand end, which is the transcription start site. In this case, the protected probe is 350 nt long, so the transcription start site is 350 bp upstream of the labeled *Bam*HI site.

One can also use S1 mapping to locate the 3′-end of a transcript. It is necessary to hybridize a 3′-end-labeled probe to the transcript, as shown in Figure 5.28. All other aspects of the assay are the same as for 5′-end mapping.

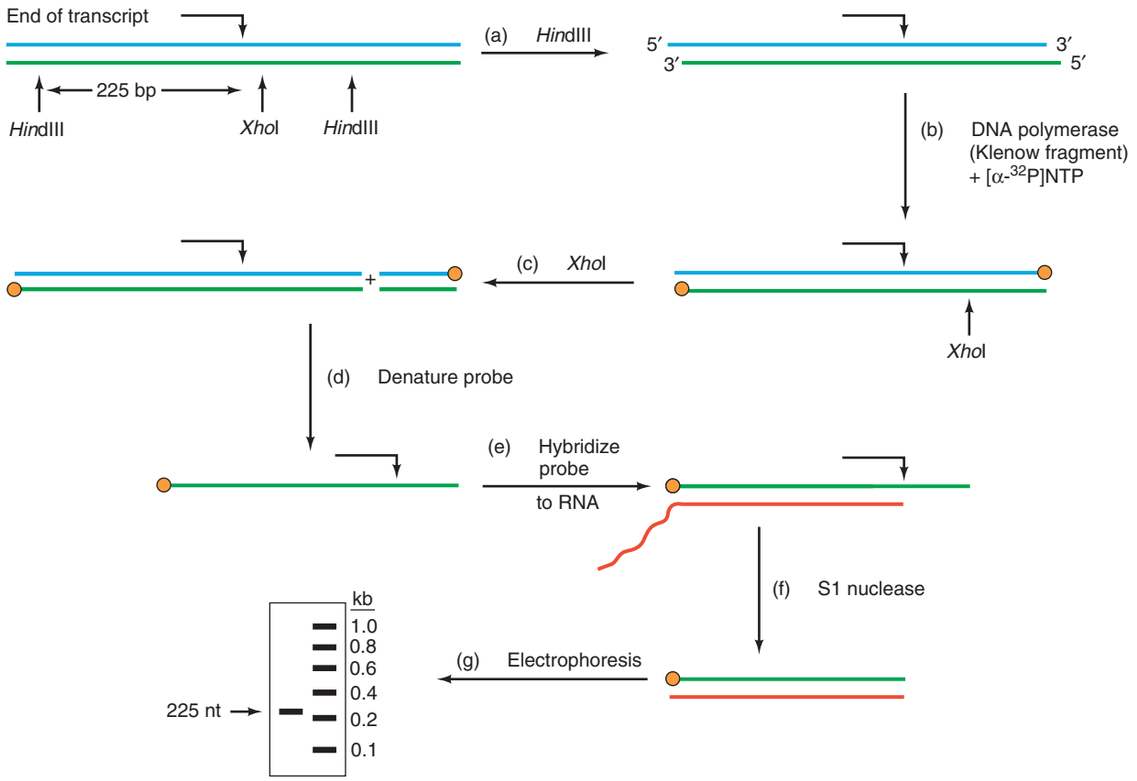


Figure 5.28 S1 mapping the 3′-end of a transcript. The principle is the same as in 5′-end mapping except that a different means of labeling the probe—at its 3′-end instead of its 5′-end—is used (detailed in Figure 5.29). In step **(a)** cut with *Hin*dIII, then in step **(b)** label the 3′-ends of the resulting fragment. The orange circles denote these labeled ends. In step **(c)** cut with *Xho*I and purify the left-hand labeled fragment by gel electrophoresis. In step **(d)** denature the probe and hybridize it to RNA (red) in step **(e).** In step **(f)** remove the unprotected region of the probe (and of the RNA) with S1 nuclease. Finally, in step **(g)** electrophorese the labeled protected probe to determine its size. In this case it is 225 nt long, which indicates that the 3′-end of the transcript lies 225 bp downstream of the labeled *Hin*dIII site on the left-hand end of the probe.

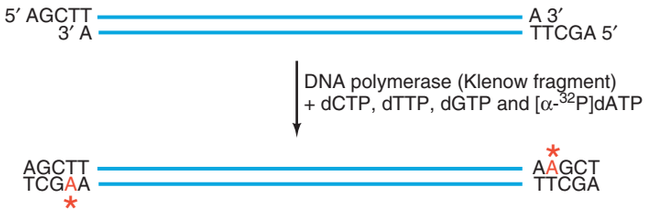


Figure 5.29 3′-end-labeling a DNA by end-filling. The DNA fragment at the top has been created by cutting with *Hin*dIII, which leaves 5′-overhangs at each end, as shown. These can be filled in with a fragment of DNA polymerase called the Klenow fragment (Chapter 20). This enzyme fragment has an advantage over the whole DNA polymerase in that it lacks the normal 5′→3′ exonuclease activity, which could degrade the 5′-overhangs before they could be filled in. The end-filling reaction is run with all four nucleotides, one of which (dATP) is labeled, so the DNA end becomes labeled. If more labeling is desired, more than one labeled nucleotide can be used.

3′-end-labeling is different from 5′-labeling because polynucleotide kinase will not phosphorylate 3′-hydroxyl groups on nucleic acids. One way to label 3′-ends is to perform **end-filling,** as shown in Figure 5.29. When a DNA is cut with a restriction enzyme that leaves a recessed 3′-end, that recessed end can be extended in vitro until it is flush with the 5′-end. If labeled nucleotides are included in this end-filling reaction, the 3′-ends of the DNA will become labeled.

S1 mapping can be used not only to map the ends of a transcript, but to tell the transcript concentration. Assuming that the probe is in excess, the intensity of the band on the autoradiograph is proportional to the concentration of the transcript that protected the probe. The more transcript, the more protection of the labeled probe, so the more intense the band on the autoradiograph. Thus, once it is known which band corresponds to the transcript of interest, its intensity can be used to measure the transcript concentration.

One important variation on the S1 mapping theme is **RNase mapping (RNase protection assay).** This procedure is analogous to S1 mapping and can yield the same information about the 5′- and 3′-ends and concentration of a specific transcript. The probe in this method, however, is made of RNA and can therefore be degraded with RNase instead of S1 nuclease. This technique is very popular, partly because of the relative ease of preparing RNA probes (**riboprobes**) by transcribing recombinant plasmids or phagemids in vitro with purified phage RNA polymerases. Another advantage of using riboprobes is that they can be labeled to very high specific activity by including a labeled nucleotide in the in vitro transcription reaction, yielding a uniformly-labeled, rather than an end-labeled probe. The higher the specific activity of the probe, the more sensitive it is in detecting tiny quantities of transcripts.

SUMMARY In S1 mapping, a labeled DNA probe is used to detect the 5′- or 3′-end of a transcript. Hybridization of the probe to the transcript protects a portion of the probe from digestion by S1 nuclease, which specifically degrades single-stranded polynucleotides. The length of the section of probe protected by the transcript locates the end of the transcript, relative to the known location of an end of the probe. Because the amount of probe protected by the transcript is proportional to the concentration of transcript, S1 mapping can also be used as a quantitative method. RNase mapping is a variation on S1 mapping that uses an RNA probe and RNase instead of a DNA probe and S1 nuclease.

Primer Extension

S1 mapping has been used in some classic experiments we will introduce in later chapters, and it is the best method for mapping the 3′-end of a transcript, but it has some drawbacks. One is that the S1 nuclease tends to "nibble" a bit on the ends of the RNA–DNA hybrid, or even within the hybrid where A–T-rich regions can melt transiently. On the other hand, sometimes the S1 nuclease will not digest the single-stranded regions completely, so the transcript appears to be a little longer than it really is. These can be serious problems if we need to map the end of a transcript with one-nucleotide accuracy. But another method, called **primer extension** can tell the 5′-end (but not the 3′-end) to the very nucleotide.

Figure 5.30 shows how primer extension works. The first step, transcription, generally occurs naturally in vivo. One simply harvests cellular RNA containing the transcript whose 5′-end is to be mapped and whose sequence is known. Next, one hybridizes a labeled oligonucleotide (the primer) of approximately 18 nt to the cellular RNA. Notice that the specificity of this method derives from the complementarity between the primer and the transcript, just as the specificity of S1 mapping comes from the complementarity between the probe and the transcript. In principle, this primer (or an S1 probe) will be able to pick out the transcript to be mapped from a sea of other, unrelated RNAs.

Next, one uses reverse transcriptase to extend the oligonucleotide primer to the 5′-end of the transcript. As presented in Chapter 4, reverse transcriptase is an enzyme that performs the reverse of the transcription reaction; that is, it makes a DNA copy of an RNA template. Hence, it is perfectly suited for the job we are asking it to do: making a DNA copy of the RNA to be mapped. Once this primer extension reaction is complete, one can denature the RNA–DNA hybrid and electrophorese the labeled

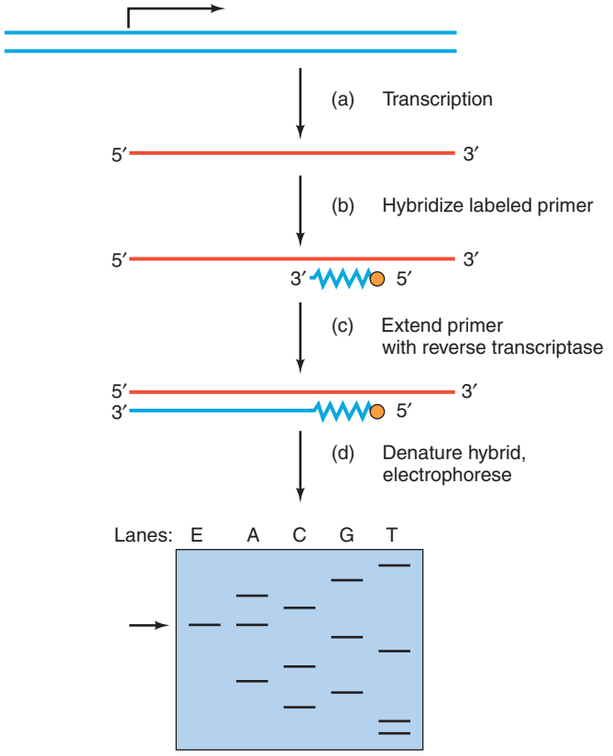


Figure 5.30 Primer extension. (a) Transcription occurs naturally within the cell, so begin by harvesting cellular RNA. **(b)** Knowing the sequence of at least part of the transcript, synthesize and label a DNA oligonucleotide that is complementary to a region not too far from the suspected 5′-end, then hybridize this oligonucleotide to the transcript. It should hybridize specifically to this transcript and to no others. **(c)** Use reverse transcriptase to extend the primer by synthesizing DNA complementary to the transcript, up to its 5′-end. If the primer itself is not labeled, or if it is desirable to introduce extra label into the extended primer, labeled nucleotides can be included in this step. **(d)** Denature the hybrid and electrophorese the labeled, extended primer (experimental lane, E). In separate lanes (lanes A, C, G, and T) run sequencing reactions, performed with the same primer and a DNA from the transcribed region, as markers. In principle, this can indicate the transcription start site to the exact base. In this case, the extended primer (arrow) coelectrophoreses with a DNA fragment in the sequencing A lane. Because the same primer was used in the primer extension reaction and in all the sequencing reactions, this shows that the 5′-end of this transcript corresponds to the middle A (underlined) in the sequence TTCGACTGACAGT.

DNA along with markers on a high-resolution gel such as the ones used in DNA sequencing. In fact, it is convenient to use the same primer used during primer extension to do a set of sequencing reactions with a cloned DNA template. One can then use the products of these sequencing reactions as markers. In the example illustrated here, the product comigrates with a band in the A lane, indicating that the 5′-end of the transcript corresponds to the second A (underlined) in the sequence TTCGACTGACAGT. This is a very accurate determination of the transcription start site.

Just as with S1 mapping, primer extension can also give an estimate of the concentration of a given transcript. The higher the concentration of transcript, the more molecules of labeled primer will hybridize and therefore the more labeled reverse transcripts will be made. The more labeled reverse transcripts, the darker the band on the autoradiograph of the electrophoretic gel.

SUMMARY Using primer extension one can locate the 5′-end of a transcript by hybridizing an oligonucleotide primer to the RNA of interest, extending the primer with reverse transcriptase to the 5′-end of the transcript, and electrophoresing the reverse transcript to determine its size. The intensity of the signal obtained by this method is a measure of the concentration of the transcript.

Run-off Transcription and G-Less Cassette Transcription

Suppose you want to mutate a gene's promoter and observe the effects of the mutations on accuracy and efficiency of transcription. You would need a convenient assay that would tell you two things: (1) whether transcription is accurate (i.e., it initiates in the right place, which you have already mapped in previous primer extension or other experiments); and (2) how much of this accurate transcription occurred. You could use S1 mapping or primer extension, but they are relatively complicated. A simpler method, called **run-off transcription,** will give answers more rapidly.

Figure 5.31 illustrates the principle of run-off transcription. You start with a DNA fragment containing the gene you want to transcribe, then cut it with a restriction enzyme in the middle of the transcribed region. Next, you transcribe this truncated gene fragment in vitro with labeled nucleotides so the transcript becomes labeled. Because you have cut the gene in the middle, the polymerase reaches the end of the fragment and simply "runs off." Hence the name of this method. Now you can measure the length of the run-off transcript. Because you know precisely the location of the restriction site at the 3′-end of the truncated gene (a *Sma*I site in this case), the length of the run-off transcript (327 nt in this case) confirms that the start of transcription is 327 bp upstream of the *Sma*I site.

Notice that S1 mapping and primer extension are well suited to mapping transcripts made in vivo; by contrast, run-off transcription relies on transcription in vitro. Thus, it will work only with genes that are accurately transcribed in vitro and cannot give information about cellular transcript concentrations. However, it is a good method for measuring the rate of in vitro transcription. The more

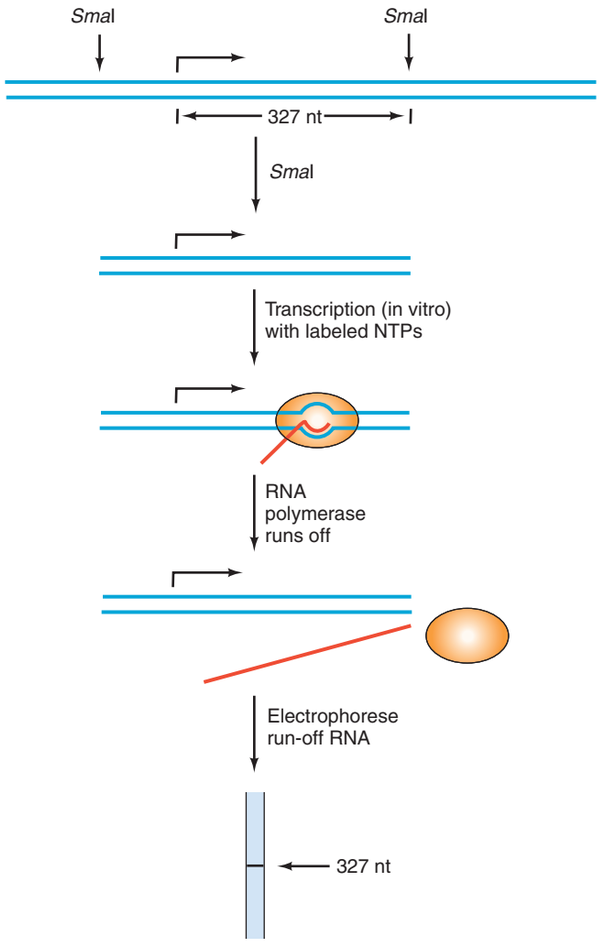


Figure 5.31 Run-off transcription. Begin by cutting the cloned gene, whose transcription is to be measured, with a restriction enzyme. Then transcribe this truncated gene in vitro. When the RNA polymerase (orange) reaches the end of the shortened gene, it falls off and releases the run-off transcript (red). The size of the run-off transcript (327 nt in this case) can be measured by gel electrophoresis and corresponds to the distance between the start of transcription and the known restriction site at the 3′-end of the shortened gene (a *Sma*I site in this case). The more actively this gene is transcribed, the stronger the 327-nt signal.

transcript is made, the more intense will be the run-off transcription signal. Indeed, run-off transcription is most useful as a quantitative method. After you have identified the physiological transcription start site by S1 mapping or primer extension you can use run-off transcription in vitro.

A variation on the run-off theme of quantifying accurate transcription in vitro is the **G-less cassette** assay (Figure 5.32). Here, instead of cutting the gene, a G-less cassette, or stretch of nucleotides lacking guanines in the nontemplate strand is inserted just downstream of the promoter. This template is transcribed in vitro with CTP,

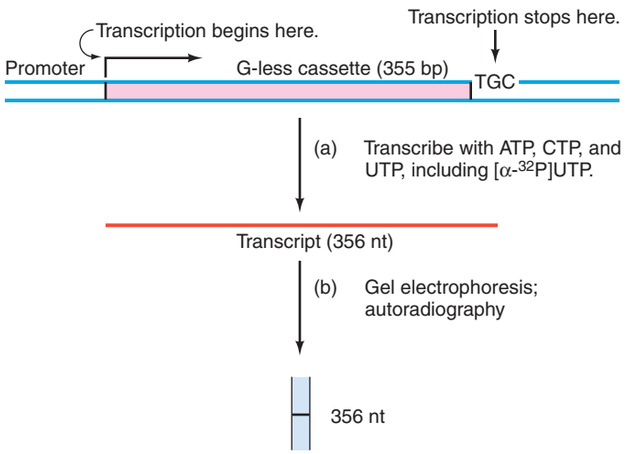


Figure 5.32 G-less cassette assay. (a) Transcribe a template with a G-less cassette (pink) inserted downstream of the promoter in vitro in the absence of GTP. This cassette is 355 bp long, contains no G's in the nontemplate strand, and is followed by the sequence TGC, so transcription stops just before the G, producing a transcript 356 nt long. **(b)** Electrophorese the labeled transcript and autoradiograph the gel and measure the intensity of the signal, which indicates how actively the cassette was transcribed.

ATP, and UTP, one of which is labeled, but no GTP. Transcription will stop at the end of the cassette where the first G is required, yielding an aborted transcript of predictable size (based on the size of the G-less cassette, which is usually a few hundred base pairs long). These transcripts are electrophoresed, and the gel is autoradiographed to measure the transcription activity. The stronger the promoter, the more of these aborted transcripts will be produced, and the stronger the corresponding band on the autoradiograph will be.

SUMMARY Run-off transcription is a means of checking the efficiency and accuracy of in vitro transcription. A gene is truncated in the middle and transcribed in vitro in the presence of labeled nucleotides. The RNA polymerase runs off the end and releases an incomplete transcript. The size of this run-off transcript locates the transcription start site, and the amount of this transcript reflects the efficiency of transcription. In G-less cassette transcription, a promoter is fused to a double-stranded DNA cassette lacking G's in the nontemplate strand, then the construct is transcribed in vitro in the absence of GTP. Transcription aborts at the end of the cassette, yielding a predictable size band on gel electrophoresis.

5.5 Measuring Transcription Rates in Vivo

Primer extension, S1 mapping, and Northern blotting are useful for determining the concentrations of specific transcripts in a cell at a given time, but they do not necessarily tell us the rates of synthesis of the transcripts. That is because the transcript concentration depends not only on its rate of synthesis, but also on its rate of degradation. To measure transcription rates, we can employ other methods, including nuclear run-on transcription and reporter gene expression.

Nuclear Run-On Transcription

The idea of this assay (Figure 5.33a) is to isolate nuclei from cells, then allow them to extend in vitro the transcripts they had already started in vivo. This continuing transcription in isolated nuclei is called **run-on transcription** because the RNA polymerase that has already initiated transcription in vivo simply "runs on" or continues to elongate the same RNA chains. The run-on reaction is usually done in the presence of labeled nucleotides so the products will be labeled. Because initiation of new RNA chains in isolated nuclei does not generally occur, one can be fairly confident that any transcription observed in the isolated nuclei is simply a continuation of transcription that was already occurring in vivo. Therefore, the transcripts obtained in a run-on reaction should reveal not only transcription rates but also give an idea about which genes are transcribed in vivo. To eliminate the possibility of initiation of new RNA chains in vitro, one can add heparin, an anionic polysaccharide that binds to any free RNA polymerase and prevents reinitiation.

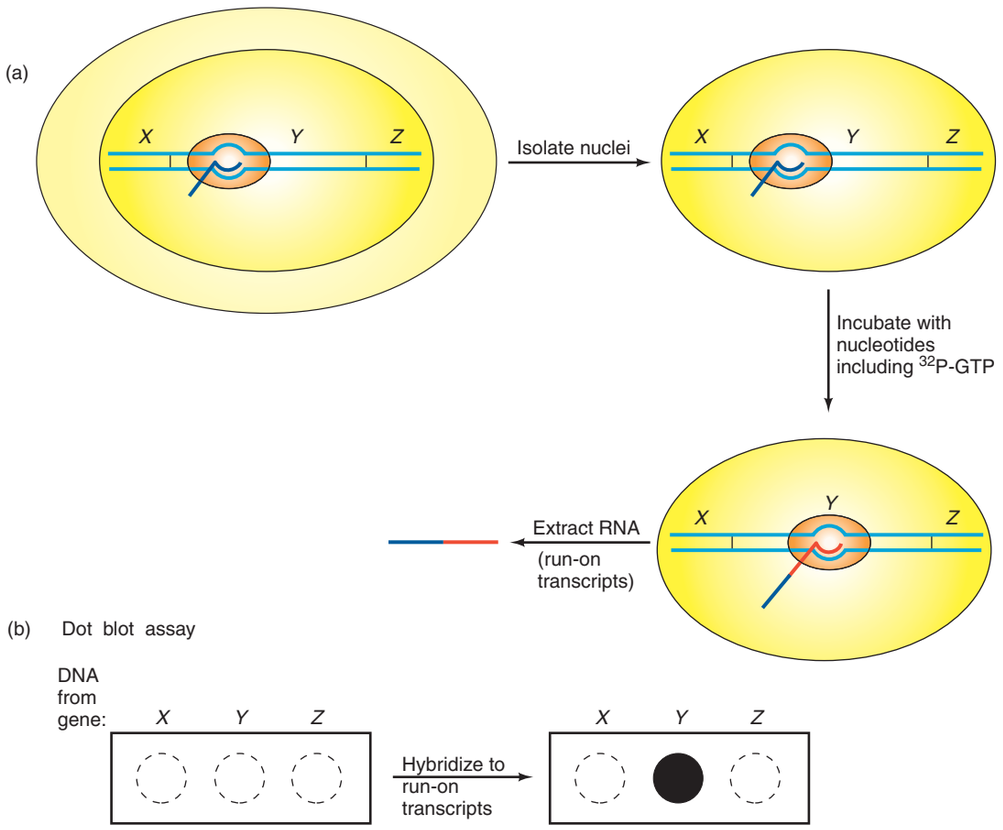


Figure 5.33 Nuclear run-on transcription. (a) The run-on reaction. Start with cells that are in the process of transcribing the *Y* gene, but not the *X* or *Z* genes. The RNA polymerase (orange) is making a transcript (blue) of the *Y* gene. Isolate nuclei from these cells and incubate them with nucleotides so transcription can continue (run-on). Also include a labeled nucleotide in the run-on reaction so the transcripts become labeled (red). Finally, extract the labeled run-on transcripts. **(b)** Dot blot assay. Spot single-stranded DNA from genes *X, Y,* and *Z* on nitrocellulose or another suitable medium, and hybridize the blot to the labeled run-on transcripts. Because gene *Y* was transcribed in the run-on reaction, its transcript will be labeled, and the gene *Y* spot becomes labeled. The more active the transcription of gene *Y,* the more intense the labeling will be. On the other hand, because genes *X* and *Z* were not active, no labeled *X* and *Z* transcripts were made, so the *X* and *Z* spots remain unlabeled.

Once labeled run-on transcripts have been produced, they must be identified. Because few if any of them are complete transcripts, their sizes will not be meaningful. The easiest way to perform the identification is by dot blotting (see Figure 5.33b). Samples of known, denatured DNAs are spotted on a filter and this **dot blot** is hybridized to the labeled run-on RNA. The RNA is identified by the DNA to which it hybridizes. Furthermore, the relative activity of a given gene is proportional to the degree of hybridization to the DNA from that gene. The conditions in the run-on reaction can also be manipulated, and the effects on the products can be measured. For example, inhibitors of certain RNA polymerases can be included to see if they inhibit transcription of a certain gene. If so, the RNA polymerase responsible for transcribing that gene can be identified.

SUMMARY Nuclear run-on transcription is a way of ascertaining which genes are active in a given cell by allowing transcription of these genes to continue in isolated nuclei. Specific transcripts can be identified by their hybridization to known DNAs on dot blots. The run-on assay can also be used to determine the effects of assay conditions on nuclear transcription.

Reporter Gene Transcription

Another way to measure transcription in vivo is to place a surrogate **reporter gene** under control of a specific promoter, and then measure the accumulation of the product of this reporter gene. For example, imagine that you want to examine the structure of a eukaryotic promoter. One way to do this is to make mutations in the DNA region that contains the promoter, then introduce the mutated DNA into cells and measure the effects of the mutations on promoter activity. You can use S1 mapping or primer extension analysis to do this measurement, but you can also substitute a reporter gene for the natural gene, and then assay the activity of the reporter gene product.

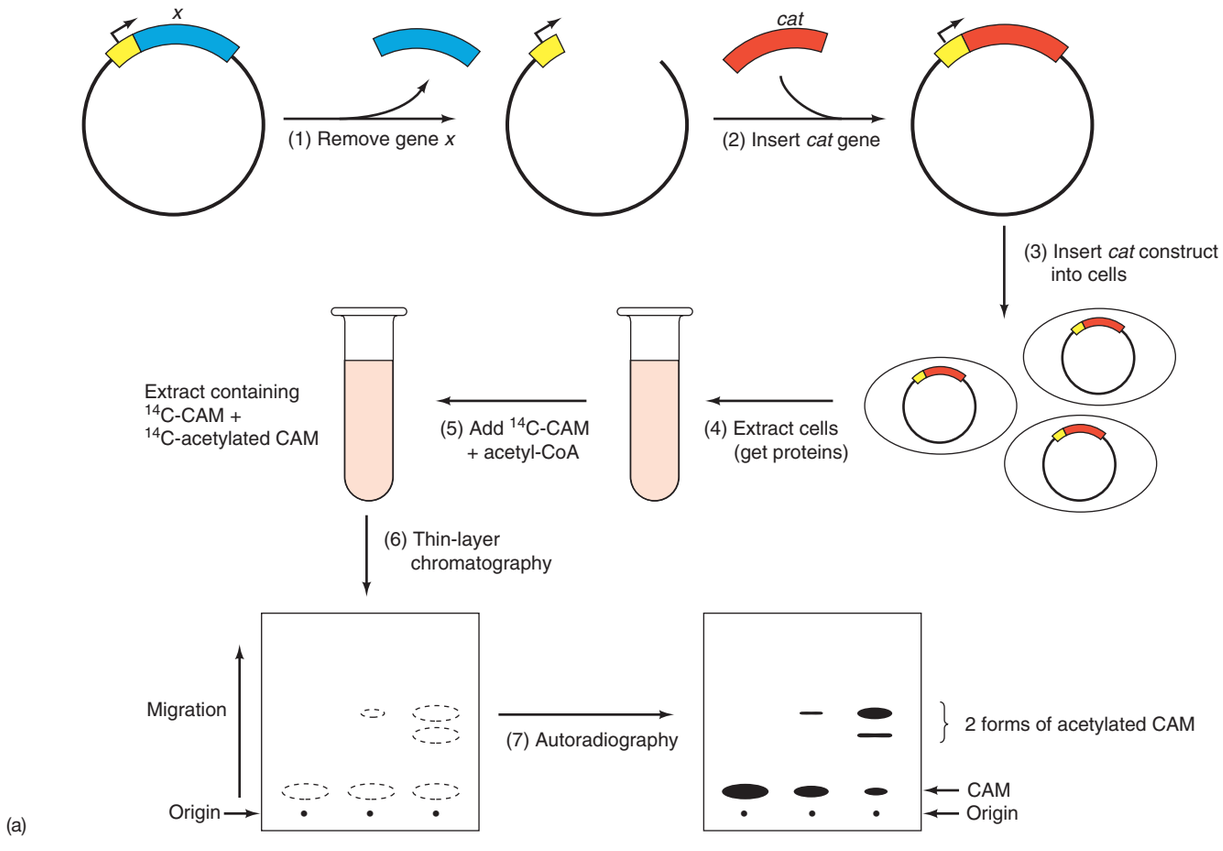
Why do it this way? The main reason is that reporter genes have been carefully chosen to have products that are very convenient to assay—more convenient than S1 mapping or primer extension. One of the most popular reporter genes is *lacZ,* whose product, β-galactosidase, can be measured using chromogenic substrates such as X-gal, which turns blue on cleavage. Another widely used reporter gene is the bacterial gene *(cat)* encoding the enzyme **chloramphenicol acetyl transferase (CAT).** The growth of most bacteria is inhibited by the antibiotic chloramphenicol (CAM), which blocks a key step in protein synthesis (Chapter 18). Some bacteria have developed a means of evading this antibiotic by acetylating it and therefore blocking its activity. The enzyme that carries out this acetylation is CAT. But eukaryotes are not susceptible to this antibiotic, so they have no need for CAT. Thus, the background level of CAT activity in eukaryotic cells is zero. That means that one can introduce a *cat* gene into these cells, under control of a eukaryotic promoter, and any CAT activity observed is due to the introduced gene.

How could you measure CAT activity in cells that have been transfected with the *cat* gene? In one of the most popular methods, an extract of the transfected cells is mixed with radioactive chloramphenicol and an acetyl donor (acetyl-CoA). Then thin-layer chromatography is used to separate the chloramphenicol from its acetylated products. The greater the concentrations of these products, the higher the CAT activity in the cell extract, and therefore the higher the promoter activity. Figure 5.34 outlines this procedure.

(Thin layer chromatography uses a thin layer of adsorbent material, such as silica gel, attached to a plastic backing. One places substances to be separated in spots near the bottom of the thin layer plate, then places the plate into a chamber with a shallow pool of solvent in the bottom. As the solvent wicks upward through the thin layer, substances move upward as well, but their mobilities depend on their relative affinities for the adsorbent material and the solvent.)

Another standard reporter gene is the **luciferase** gene from firefly lanterns. The enzyme luciferase, mixed with ATP and luciferin, converts the luciferin to a chemiluminescent compound that emits light. That is the secret of the firefly's lantern, and it also makes a convenient reporter because the light can be detected easily with x-ray film, or even in a scintillation counter.

In the experiments described here, we are assuming that the amount of reporter gene product is a reasonable measure of transcription rate (the number of RNA chain initiations per unit of time) and therefore of promoter activity. But the gene products come from a two-step process that includes translation as well as transcription. Ordinarily, we are justified in assuming that the translation rates do not vary from one DNA construct to another, as long as we are manipulating only the promoter. That is because the promoter lies outside the coding region. For this reason changes in the promoter cannot affect the structure of the mRNA itself and therefore should not affect translation. However, one can deliberately make changes in the region of a gene that will be transcribed into mRNA and then use a reporter gene to measure the effects of these changes on translation. Thus, depending on where the changes to a gene are made, a reporter gene can detect alterations in either transcription or translation rates.



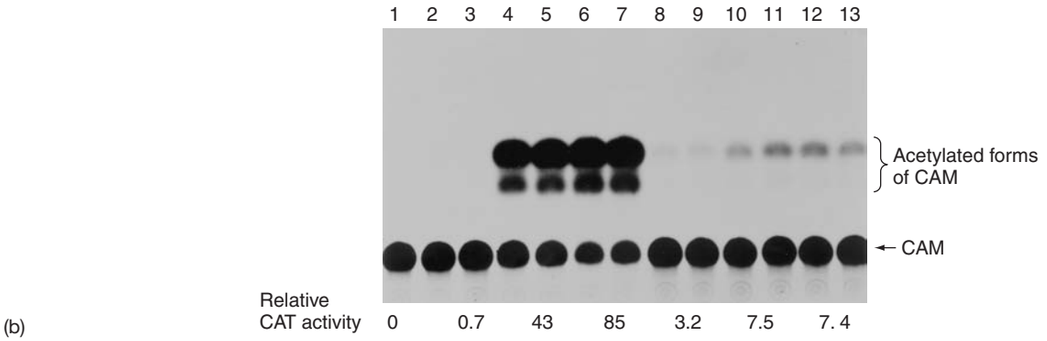


Figure 5.34 Using a reporter gene. (a) Outline of the method. Step 1: Start with a plasmid containing gene *X,* (blue) under control of its own promoter (yellow) and use restriction enzymes to remove the coding region of gene *X.* Step 2: Insert the bacterial *cat* gene under control of the *X* gene's promoter. Step 3: Place this construct into eukaryotic cells. Step 4: After a period of time, make an extract of the cells that contains soluble cellular proteins. Step 5: To begin the CAT assay, add ^{14}C-CAM and the acetyl donor acetyl-CoA. Step 6: Perform thin-layer chromatography to separate acetylated and unacetylated species of CAM. Step 7: Finally, subject the thin layer to autoradiography to visualize CAM and its acetylated derivatives. Here CAM is seen near the bottom of the autoradiograph and two acetylated forms of CAM, with higher mobility, are seen near the top. **(b)** Actual experimental results. Again, the parent CAM is near the bottom, and two acetylated forms of CAM are nearer the top. The experimenters scraped these radioactive species off of the thin-layer plate and subjected them to liquid scintillation counting, yielding the CAT activity values reported at the bottom (averages of duplicate lanes). Lane 1 is a negative control with no cell extract. *Abbreviations:* CAM = chloramphenicol; CAT = chloramphenicol acetyl transferase. (*Source:* (*b*) Qin, Liu, and Weaver. Studies on the control region of the p10 gene of the *Autographa californica* nuclear polyhedrosis virus. *J. General Virology* 70 (1989) f. 2, p. 1276. (Society for General Microbiology, Reading, England.)

SUMMARY To measure the activity of a promoter, one can link it to a reporter gene, such as the genes encoding β-galactosidase, CAT, or luciferase, and let the easily assayed reporter gene products indicate the activity of the promoter. One can also use reporter genes to detect changes in translational efficiency after altering regions of a gene that affect translation.

Measuring Protein Accumulation in Vivo

Gene activity can also be measured by monitoring the accumulation of the ultimate products of genes—proteins. This is commonly done in two ways, **immunoblotting (Western blotting)**, which we discussed earlier in this chapter, and immunoprecipitation.

Immunoprecipitation begins with labeling proteins in a cell by growing the cells with a labeled amino acid, typically [^{35}S] methionine. Then the labeled cells are homogenized and a particular labeled protein is bound to a specific antibody or antiserum directed against that protein. The antibody-protein complex is precipitated with a secondary antibody or protein A coupled to resin beads that can be sedimented in a low-speed centrifuge. Then the precipitated protein is electrophoresed and detected by autoradiography. Note that the antibody and other reagents will also be present in the precipitate, but will not be detected because they are not labeled. The more label in the protein band, the more that protein has accumulated in vivo.

SUMMARY Gene expression can be quantified by measuring the accumulation of the protein products of genes. Immunoblotting and immunoprecipitation are the favorite ways of accomplishing this task.

5.6 Assaying DNA–Protein Interactions

Another of the recurring themes of molecular biology is the study of DNA–protein interactions. We have already discussed RNA polymerase–promoter interactions, and we will encounter many more examples. Therefore, we need methods to quantify these interactions and to determine exactly what part of the DNA interacts with a given protein. We will consider here two methods for detecting protein–DNA binding and three examples of methods for showing which DNA bases interact with a protein.

Filter Binding

Nitrocellulose membrane filters have been used for decades to filter–sterilize solutions. Part of the folklore of molecular biology is that someone discovered by accident that DNA can bind to such nitrocellulose filters because they lost their DNA preparation that way. Whether this story is true or not is unimportant. What is important is that nitrocellulose filters can indeed bind DNA, but only under certain conditions. Single-stranded DNA binds readily to nitrocellulose, but double-stranded DNA by itself does not. On the other hand, protein does bind, and if a protein is bound to double-stranded DNA the protein–DNA complex will bind. This is the basis of the assay portrayed in Figure 5.35.

In Figure 5.35a, labeled double-stranded DNA is poured through a nitrocellulose filter. The amount of label in the filtrate (the material that passes through the filter) and in the filter-bound material is measured, which shows that all the labeled material has passed through the filter into the filtrate. This confirms that double-stranded DNA does not bind to nitrocellulose. In Figure 5.35b, a solution of a labeled protein is filtered, showing that all the protein is bound to the filter. This demonstrates that proteins bind by themselves to the filter. In Figure 5.35c, double-stranded DNA is again labeled, but this time it is mixed with a protein to which it binds. Because the protein binds to the filter, the protein–DNA complex will also bind, and the radioactivity is found bound to the filter, rather than in the filtrate. Thus, filter binding is a direct measure of DNA–protein interaction.

SUMMARY Filter binding as a means of measuring DNA–protein interaction is based on the fact that double-stranded DNA will not bind by itself to a nitrocellulose filter or similar medium, but a protein–DNA complex will. Thus, one can label a double-stranded DNA, mix it with a protein, and assay protein–DNA binding by measuring the amount of label retained by the filter.

Gel Mobility Shift

Another method for detecting DNA–protein interaction relies on the fact that a small DNA has a much higher mobility in gel electrophoresis than the same DNA does when it is bound to a protein. Thus, one can label a short, double-stranded DNA fragment, then mix it with a protein, and electrophorese the complex. Then one subjects the gel to autoradiography to detect the labeled species. Figure 5.36 shows the electrophoretic mobilities of three different species. Lane 1 contains naked DNA, which has a very high mobility because of its small size. Recall from earlier in this chapter that DNA electropherograms are

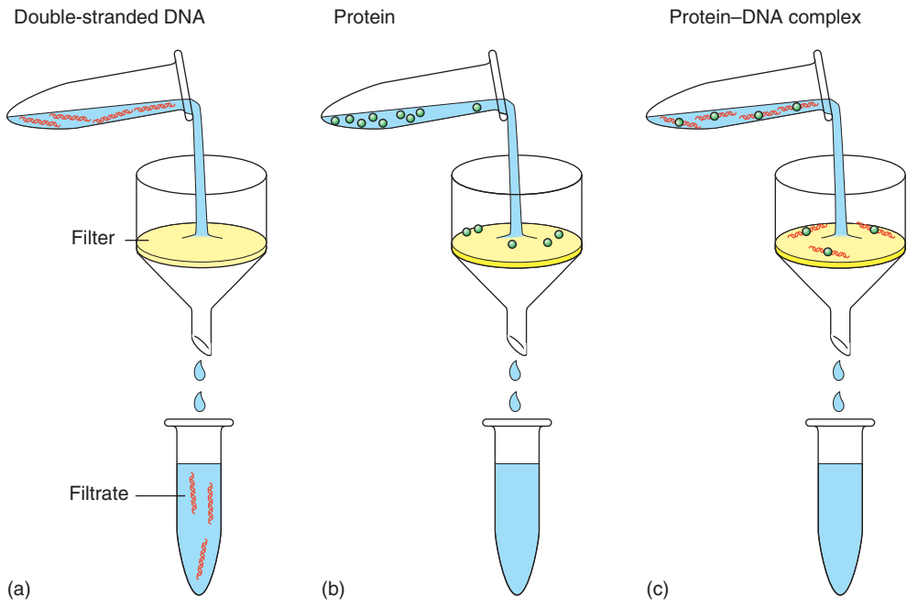


Figure 5.35 Nitrocellulose filter-binding assay. (a) Double-stranded DNA. End-label double-stranded DNA (red), and pass it through a nitrocellulose filter. Then monitor the radioactivity on the filter and in the filtrate by liquid scintillation counting. None of the radioactivity sticks to the filter, indicating that double-stranded DNA does not bind to nitrocellulose. Single-stranded DNA, on the other hand, binds tightly. **(b)** Protein. Label a protein (green), and filter it through nitrocellulose. The protein binds to the nitrocellulose. **(c)** Double-stranded DNA–protein complex. Mix an end-labeled double-stranded DNA (red) with an unlabeled protein (green) to which it binds to form a DNA–protein complex. Then filter the complex through nitrocellulose. The labeled DNA now binds to the filter because of its association with the protein. Thus, double-stranded DNA–protein complexes bind to nitrocellulose, providing a convenient assay for association between DNA and protein.

conventionally depicted with their origins at the top, so high-mobility DNAs are found near the bottom, as shown here. Lane 2 contains the same DNA bound to a protein, and its mobility is greatly reduced. This is the origin of the name for this technique: **gel mobility shift assay** or **electrophoretic mobility shift assay (EMSA).** Lane 3 depicts the behavior of the same DNA bound to two proteins. The mobility is reduced still further because of the greater mass of protein clinging to the DNA. This is called a **supershift.** The protein could be another DNA-binding protein, or a second protein that binds to the first one. It can even be an antibody that specifically binds to the first protein.

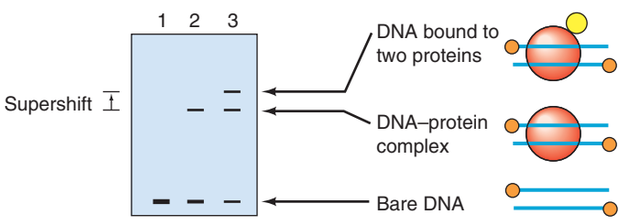
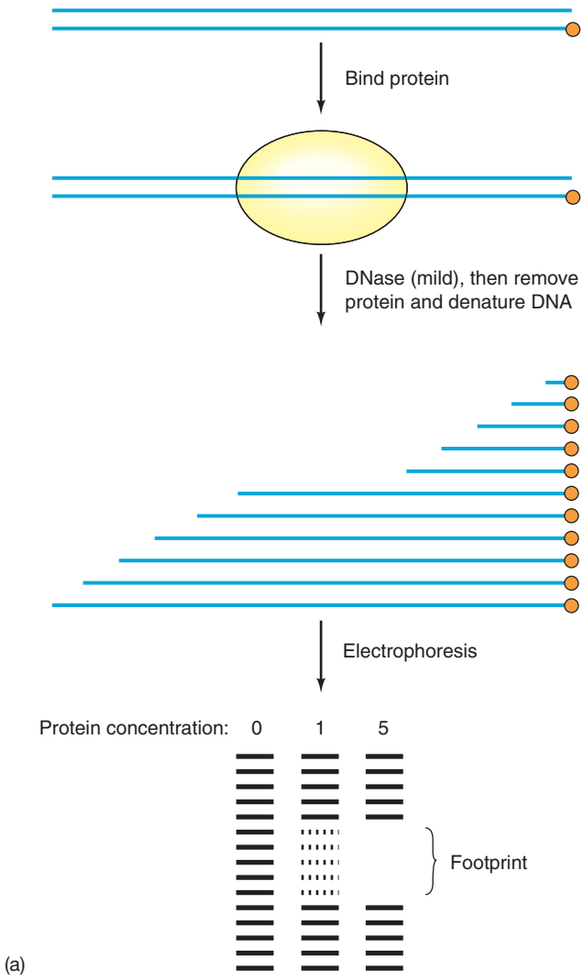


Figure 5.36 Gel mobility shift assay. Subject pure, labeled DNA or DNA–protein complexes to gel electrophoresis, then autoradiograph the gel to detect the DNAs and complexes. Lane 1 shows the high mobility of bare DNA. Lane 2 shows the mobility shift that occurs on binding a protein (red) to the DNA. Lane 3 shows the supershift caused by binding a second protein (yellow) to the DNA–protein complex. The orange dots at the ends of the DNAs represent terminal labels.

SUMMARY A gel mobility shift assay detects interaction between a protein and DNA by the reduction of the electrophoretic mobility of a small DNA that occurs on binding to a protein.

DNase Footprinting

Footprinting is a method for detecting protein–DNA interactions that can tell where the target site lies on the DNA and even which bases are involved in protein binding. Several methods are available, but three are very popular: DNase, dimethylsulfate (DMS), and hydroxyl radical footprinting. **DNase footprinting** (Figure 5.37) relies on the fact that a protein, by binding to DNA, covers the binding site and so protects it from attack by DNase. In this sense, it leaves its "footprint" on the DNA. The first step in a footprinting experiment is to end-label the DNA. Either strand can be labeled, but only one



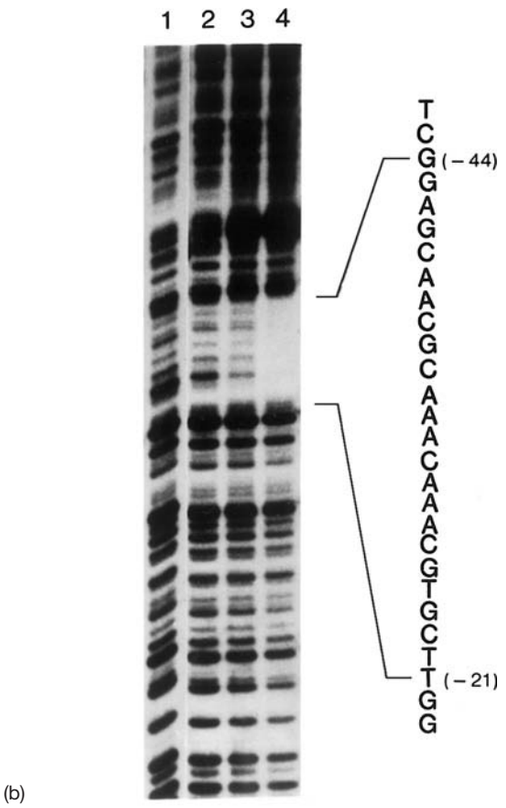
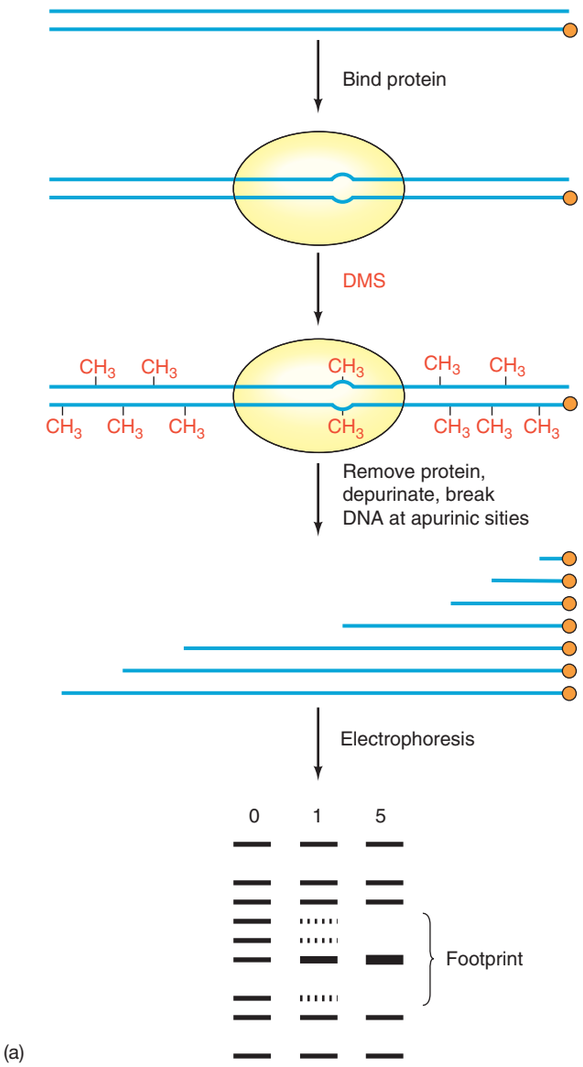


Figure 5.37 DNase footprinting. (a) Outline of method. Begin with a double-stranded DNA, labeled at one end (orange). Next, bind a protein to the DNA. Next, digest the DNA–protein complex under mild conditions with DNase I, so as to introduce approximately one break per DNA molecule. Next, remove the protein and denature the DNA, yielding the end-labeled fragments shown at center. Notice that the DNase cut the DNA at regular intervals except where the protein bound and protected the DNA. Finally, electrophorese the labeled fragments, and perform autoradiography to detect them. The three lanes represent DNA that was bound to 0, 1, and 5 units of protein. The lane with no protein shows a regular ladder of fragments. The lane with one unit of protein shows some protection, and the lane with five units of protein shows complete protection in the middle. This protected area is called the footprint; it shows where the protein binds to the DNA. Sequencing reactions performed on the same DNA in parallel lanes are usually included. These serve as size markers that show exactly where the protein bound. **(b)** Actual experimental results. Lanes 1–4 contained DNA bound to 0, 10, 18, and 90 pmol of protein, respectively (1 pmol = 10^{-12} mol). The DNA sequence was obtained previously by standard dideoxy sequencing. (*Source:* (*b*) Ho et al., Bacteriophage lambda protein *c*II binds promoters on the opposite face of the DNA helix from RNA polymerase. *Nature* 304 (25 Aug 1983) p. 705, f. 3, © Macmillan Magazines Ltd.)

strand per experiment. Next, the protein (yellow in the figure) is bound to the DNA. Then the DNA–protein complex is treated with DNase I under mild conditions (very little DNase), so that an average of only one cut occurs per DNA molecule. Next, the protein is removed from the DNA, the DNA strands are separated, and the resulting fragments are electrophoresed on a high-resolution polyacrylamide gel alongside size markers (not shown). Of course, fragments will arise from the other end of the DNA as well, but they will not be detected because they are unlabeled. A control with DNA alone (no protein) is always included, and more than one protein concentration is usually used so the gradual disappearance of the bands in the footprint region reveals that protection of the DNA depends on the concentration of added protein. The footprint represents the region of DNA protected by the protein, and therefore tells where the protein binds.



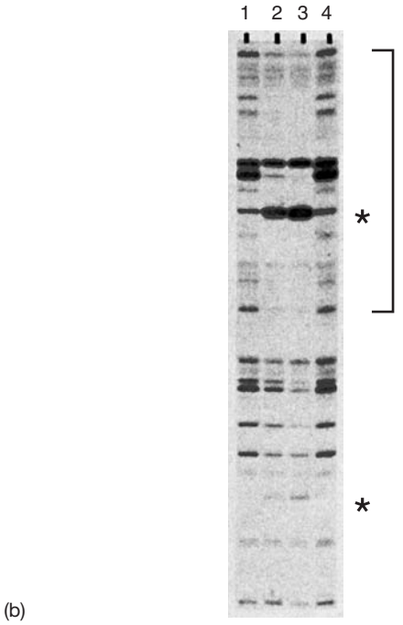


Figure 5.38 DMS footprinting. (a) Outline of the method. As in DNase footprinting, start with an end-labeled DNA, then bind a protein (yellow) to it. In this case, the protein causes some tendency of the DNA duplex to melt in one region, represented by the small "bubble." Next, methylate the DNA with DMS. This adds methyl groups (CH_3, red) to certain bases in the DNA. Do this under mild conditions so that, on average, only one methylated base occurs per DNA molecule (even though all seven methylations are shown together on one strand for convenience here). Next, use piperidine to remove methylated purines from the DNA, then to break the DNA at these apurinic sites. This yields the labeled DNA fragments depicted at center. Electrophorese these fragments and autoradiograph the gel to give the results shown at bottom. Notice that three sites are protected against methylation by the protein, but one site is actually made more sensitive to methylation (darker band). This is because of the opening up of the double helix that occurs in this position when the protein binds. **(b)** Actual experimental results. Lanes 1 and 4 have no added protein, whereas lanes 2 and 3 have increasing concentrations of a protein that binds to this region of the DNA. The bracket indicates a pronounced footprint region. The asterisks denote bases made *more* susceptible to methylation by protein binding. (*Source:* (*b*) Learned et al., Human rRNA transcription is modulated by the coordinate binding of two factors to an upstream control element. *Cell* 45 (20 June 1986) p. 849, f. 2a. Reprinted by permission of Elsevier Science.)

DMS Footprinting and Other Footprinting Methods

DNase footprinting gives a good idea of the location of the binding site for the protein, but DNase is a macromolecule and is therefore a rather blunt instrument for probing the fine details of the binding site. That is, gaps may occur in the interaction between protein and DNA that DNase would not fit into and therefore would not detect. Moreover, DNA-binding proteins frequently perturb the DNA within the binding region, distorting the double helix. These perturbations are interesting, but are

not generally detected by DNase footprinting because the protein keeps the DNase away. More detailed footprinting requires a smaller molecule that can fit into the nooks and crannies of the DNA–protein complex and reveal more of the subtleties of the interaction. A favorite tool for this job is the methylating agent **dimethyl sulfate (DMS).**

Figure 5.38 illustrates DMS footprinting, which starts in the same way as DNase footprinting, with end-labeling the DNA and binding the protein. Then the DNA–protein complex is methylated with DMS, using a mild treatment so that on average only one methylation event occurs per DNA molecule. Next, the protein is dislodged, and the DNA is treated with piperidine, which removes methylated purines, creating apurinic sites (deoxyriboses without bases), then breaks the DNA at these apurinic sites. Finally, the DNA fragments are electrophoresed, and the gel is autoradiographed to detect the labeled DNA bands. Each band ends next to a nucleotide that was methylated and thus unprotected by the protein. In this example, three bands progressively disappear as more and more protein is added. But one band actually becomes more prominent at high protein concentration. This suggests that binding the protein distorts the DNA double helix such that it makes the base corresponding to this band more vulnerable to methylation.

In addition to DNase and DMS, other reagents are commonly used to footprint protein–DNA complexes by breaking DNA except where it is protected by bound proteins. For example, organometallic complexes containing copper or iron act by generating **hydroxyl radicals** that attack and break DNA strands.

SUMMARY Footprinting is a means of finding the target DNA sequence, or binding site, of a DNA-binding protein. DNase footprinting is performed by binding the protein to its end-labeled DNA target, then attacking the DNA–protein complex with DNase. When the resulting DNA fragments are electrophoresed, the protein binding site shows up as a gap, or "footprint" in the pattern where the protein protected the DNA from degradation. DMS footprinting follows a similar principle, except that the DNA methylating agent DMS, instead of DNase, is used to attack the DNA–protein complex. The DNA is then broken at the methylated sites. Unmethylated (or hypermethylated) sites show up on electrophoresis of the labeled DNA fragments and demonstrate where the protein bound to the DNA. Hydroxyl radical footprinting uses copper- or iron-containing organometallic complexes to generate hydroxyl radicals that break DNA strands.

5.7 Finding RNA Sequences That Interact with Other Molecules

SELEX

SELEX (systematic evolution of ligands by exponential enrichment) is a method that was originally developed to discover short RNA sequences (**aptamers**) that bind to particular molecules. Figure 5.39 illustrates the classical SELEX procedure. One starts with a pool of PCR-amplified synthetic DNAs that have constant end regions (red), but random central regions (blue) that can potentially encode over 10^{15} different RNA sequences. In the first step, these DNAs are transcribed in vitro, using the phage T7 RNA polymerase, which recognizes the T7 promoter in the upstream constant region of every DNA in the pool. In the next step, the aptamers are selected by affinity chromatography (this chapter), using a resin with the target molecule immobilized. The selected RNAs bind to the resin and then can be released with a solution containing the target molecule. These selected RNAs are then reverse-transcribed to yield double-stranded DNA, which is then subjected to PCR, using primers specific for the DNAs' constant ends.

One round of SELEX yields a population of molecules only partially enriched in aptamers, so the process is repeated several more times to produce a highly enriched population of aptamers. SELEX has been extensively exploited to find the RNA sequences that are contacted

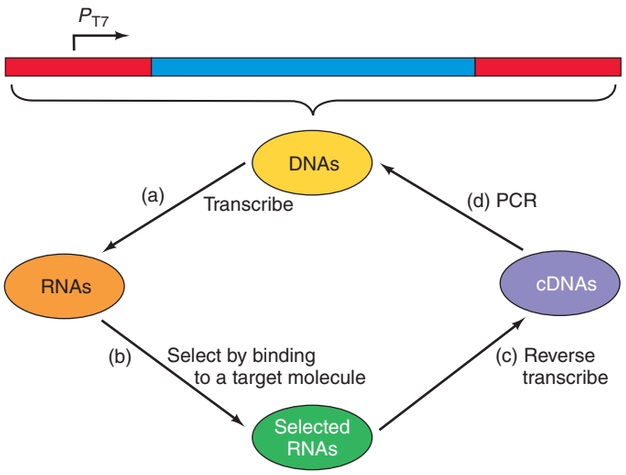


Figure 5.39 SELEX. Start with a large collection of DNAs (top) that have a random sequence (blue) flanked by constant sequences (red). **(a)** Transcribe the DNA pool to produce a pool of RNAs that also contain a random sequence flanked by constant sequences. **(b)** Select for aptamers by affinity chromatography with the target molecule. **(c)** Reverse-transcribe the selected RNAs to produce a pool of cDNAs. **(d)** Amplify the cDNAs by PCR, using primers complementary to the constant regions at the ends of the DNAs. This cycle is repeated several times to enrich the aptamers in the pool.

by proteins. It is extremely powerful in that it finds a few aptamers among an astronomically high number of starting RNA sequences.

Functional SELEX

Functional SELEX is similar to classical SELEX in that it finds a few "needles" (RNA sequences) in a "haystack" of starting sequences. But instead of finding aptamers that bind to other molecules, it finds RNA sequences that carry out, or make possible, some function. With simple binding, selection is easy; it just requires affinity chromatography. But selection based on function is trickier and requires creativity in designing the selection step. For instance, the first functional SELEX procedures detected a ribozyme (an RNA with enzymatic activity), and this ribozyme activity altered the RNA itself to allow it to be amplified. One simple example is a ribozyme that can add an olignucleotide to its own end. This activity allowed the investigators to supply an oligonucleotide of defined sequence to the ribozyme, which then added this tag to itself. Once tagged, the ribozyme becomes subject to amplification using a PCR primer complementary to the tag.

A pool of random RNA sequences may not contain any RNAs with high activity. But that problem can be overcome by carrying out the amplification step under mutagenizing conditions, such that many variants of the mildly active sequences are created. Some of these will probably have greater activity than the original. After several rounds of selection and mutagenesis, RNAs with very strong enzymatic activity can be produced.

SUMMARY SELEX is a method that allows one to find RNA sequences that interact with other molecules, including proteins. RNAs that interact with a target molecule are selected by affinity chromatography, then converted to double-stranded DNAs and amplified by PCR. After several rounds of this procedure, the RNAs are highly enriched for sequences that bind to the target molecule. Functional SELEX is a variation on this theme in which the desired function somehow alters the RNA so it can be amplified. If the desired function is enzymatic, mutagenesis can be introduced into the amplification step to produce variants with higher activity.

5.8 Knockouts

Most of the techniques we have discussed in Chapter 5 are designed to probe the structures and activities of genes. But these frequently leave a big question about the role of the gene being studied: What purpose does the gene play in the life of the organism? We can often answer this question best by seeing what happens when we create deliberate mutations in a particular gene in a living organism. We now have techniques for targeted disruption of genes in several organisms. For example, we can disrupt genes in mice, and when we do, we call the products **knockout mice.**

Figure 5.40 explains one way to begin the process of creating a knockout mouse. We start with cloned DNA containing the mouse gene we want to knock out. We interrupt this gene with another gene that confers resistance to the antibiotic neomycin. Elsewhere in the cloned DNA, outside the target gene, we introduce a thymidine kinase *(tk)* gene. Later, these extra genes will enable us to weed out those clones in which targeted disruption did not occur.

Next, we mix the engineered mouse DNA with stem cells from an embryonic brown mouse. In a small percentage of these cells, the interrupted gene will find its way into the nucleus and homologous recombination will occur between the altered gene and the resident, intact gene. This recombination places the altered gene into the mouse genome and removes the *tk* gene. Unfortunately, such recombination events are relatively rare, so many stem cells will experience no recombination and therefore will suffer no interruption of their resident gene. Still other cells will experience nonspecific recombination, in which the interrupted gene will insert randomly into the genome without replacing the intact gene.

The problem now is to eliminate the cells in which homologous recombination did not occur. This is where the extra genes we introduced earlier come in handy. Cells in which no recombination took place will have no neomycin-resistance gene. Thus, we can eliminate them by growing the cells in medium containing the neomycin derivative G418. Cells that experienced nonspecific recombination will have incorporated the *tk* gene, along with the interrupted gene, into their genome. We can kill these cells with gangcyclovir, a drug that is lethal to tk^+ cells. (The stem cells we used are tk^-.) Treatment with these two drugs leaves us with engineered cells that have undergone homologous recombination and are therefore heterozygous for an interruption in the target gene.

Our next task is to introduce this interrupted gene into a whole mouse (Figure 5.41). We do this by injecting our engineered cells into a mouse blastocyst that is destined to develop into a black mouse. Then we place this altered embryo into a surrogate mother, who eventually gives birth to a chimeric mouse. We can recognize this mouse as a chimera by its patchy coat; the black zones come from the original black embryo, and the brown zones result from the transplanted engineered cells.

To get a mouse that is a true heterozygote instead of a chimera, we allow the chimera to mature, then mate it with a black mouse. Because brown (agouti) is dominant, some of the progeny should be brown. In fact, all of the

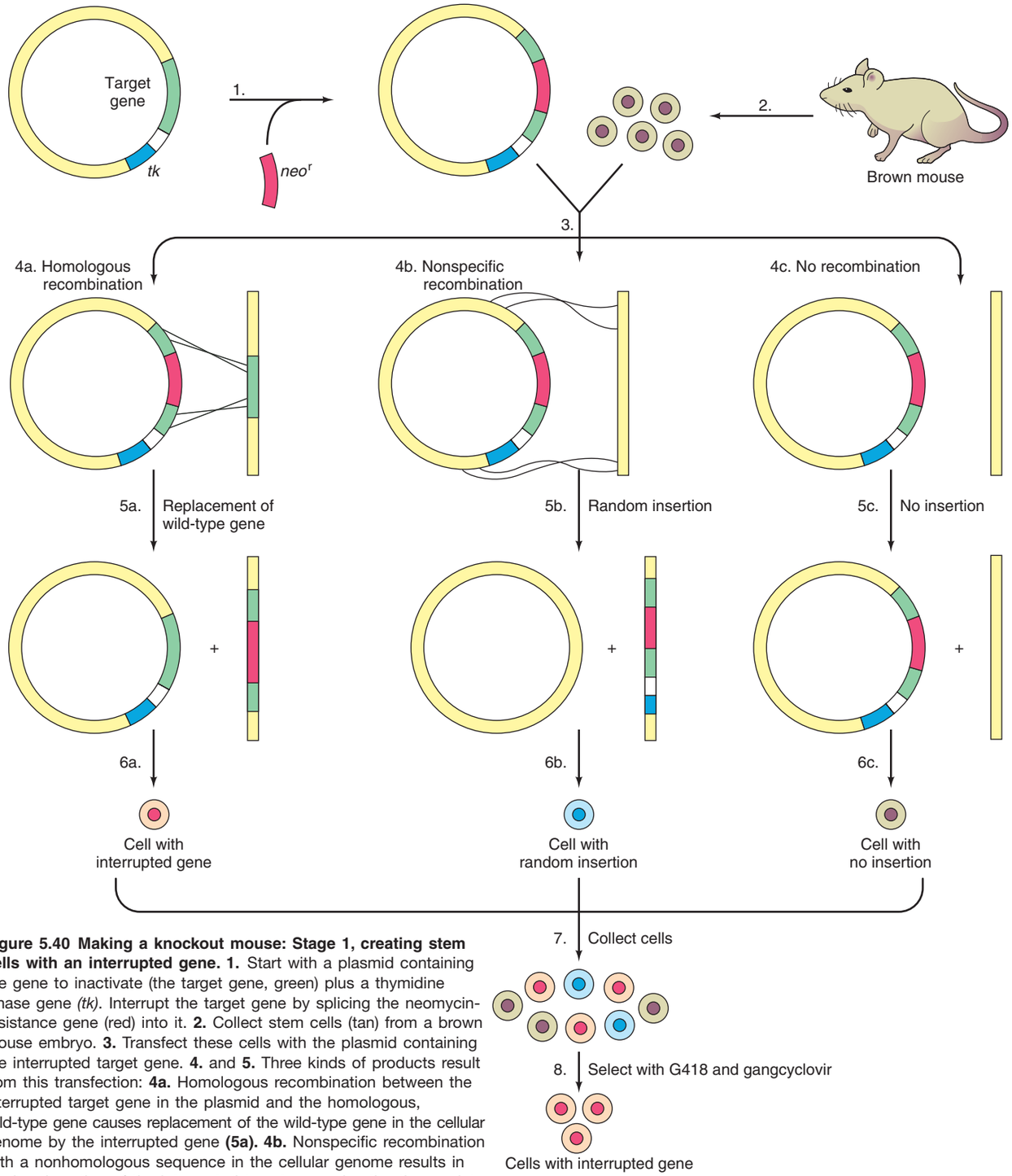


Figure 5.40 Making a knockout mouse: Stage 1, creating stem cells with an interrupted gene. 1. Start with a plasmid containing the gene to inactivate (the target gene, green) plus a thymidine kinase gene *(tk)*. Interrupt the target gene by splicing the neomycin-resistance gene (red) into it. **2.** Collect stem cells (tan) from a brown mouse embryo. **3.** Transfect these cells with the plasmid containing the interrupted target gene. **4.** and **5.** Three kinds of products result from this transfection: **4a.** Homologous recombination between the interrupted target gene in the plasmid and the homologous, wild-type gene causes replacement of the wild-type gene in the cellular genome by the interrupted gene **(5a). 4b.** Nonspecific recombination with a nonhomologous sequence in the cellular genome results in random insertion of the interrupted target gene plus the *tk* gene into the cellular genome **(5b). 4c.** When no recombination occurs, the interrupted target gene is not integrated into the cellular genome at all **(5c). 6.** The cells resulting from these three events are color-coded as indicated: Homologous recombination yields a cell (red) with an interrupted target gene **(6a);** nonspecific recombination yields a cell (blue) with the interrupted target gene and the *tk* gene inserted at random **(6b);** no recombination yields a cell (tan) with no integration of the interrupted gene **(6c). 7.** Collect the transfected cells, containing all three types (red, blue, and tan). **8.** Grow the cells in medium containing the neomycin analog G418 and the drug gangcyclovir. The G418 kills all cells without a neomycin-resistance gene, namely those cells (tan) that did not experience a recombination event. The gangcyclovir kills all cells that have a *tk* gene, namely those cells (blue) that experienced a nonspecific recombination. This leaves only the cells (red) that experienced homologous recombination and therefore have an interrupted target gene.

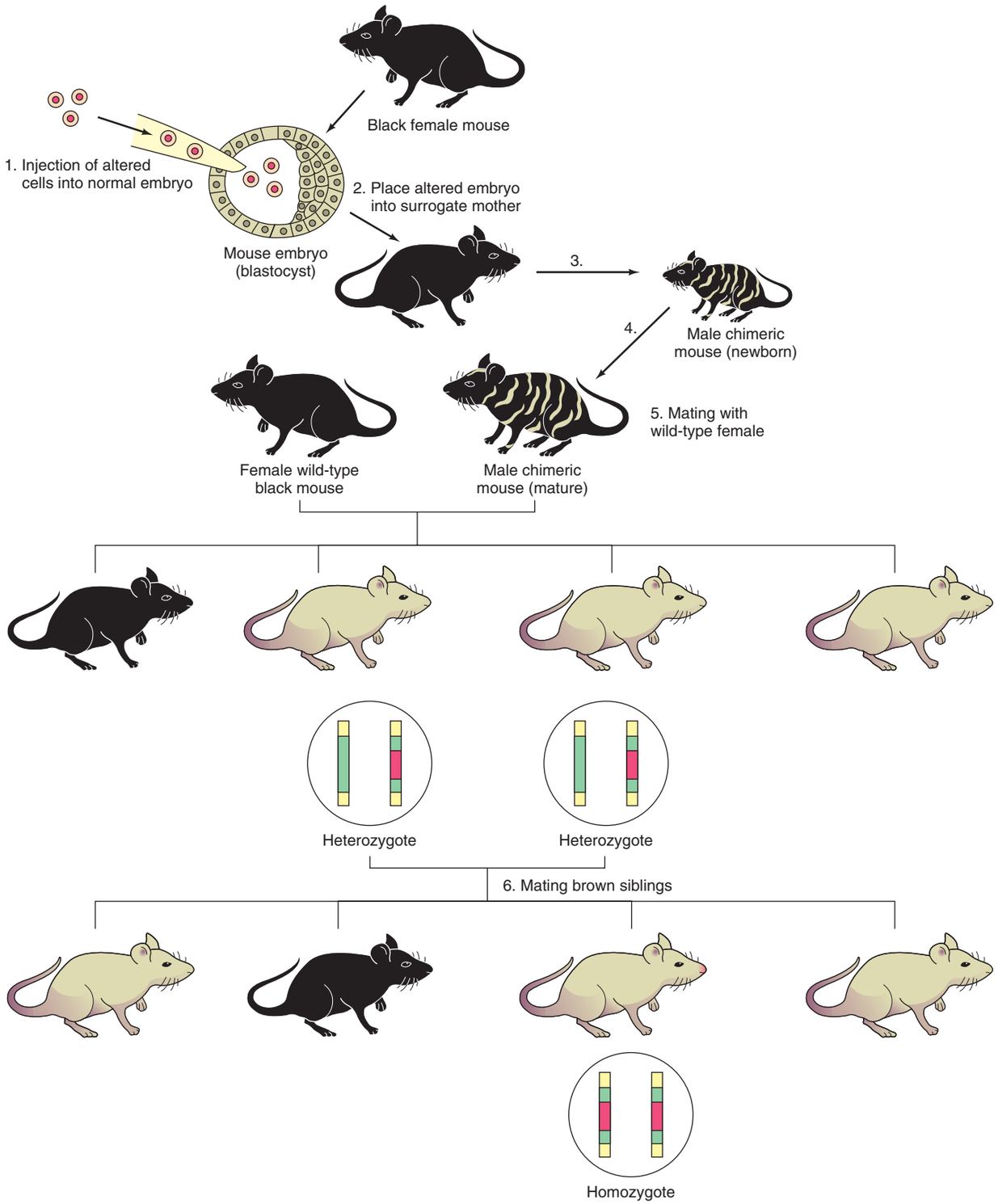


Figure 5.41 Making a knockout mouse: Stage 2, placing the interrupted gene in the animal. **(1)** Inject the cells with the interrupted gene (see Figure 5.40) into a blastocyst-stage embryo from black parent mice. **(2)** Transplant this mixed embryo to the uterus of a surrogate mother. **(3)** The surrogate mother gives birth to a chimeric mouse, which one can identify by its black and brown coat. (Recall that the altered cells came from an agouti [brown] mouse, and they were placed into an embryo from a black mouse.) **(4)** Allow the chimeric mouse (a male) to mature. **(5)** Mate it with a wild-type black female. Discard any black offspring, which must have derived from the wild-type blastocyst; only brown mice could have derived from the transplanted cells. **(6)** Select a brown brother and sister pair, both of which show evidence of an interrupted target gene (by Southern blot analysis), and mate them. Again, examine the DNA of the brown progeny by Southern blotting. This time, one animal that is homozygous for the interrupted target gene is found. This is the knockout mouse. Now observe this animal to determine the effects of knocking out the target gene.

offspring resulting from gametes derived from the engineered stem cells should be brown. But only half of these brown mice will carry the interrupted gene because the engineered stem cells were heterozygous for the knockout. Southern blots showed that two of the brown mice in our example carry the interrupted gene. We mate these and look for progeny that are homozygous for the knockout by examination of their DNA. In our example, one of the mice from this mating is a knockout, and now our job is to observe its phenotype. Frequently, as here, the phenotype is not obvious. (Can you see it?) But obvious or not, it can be very instructive.

In other cases, the knockout is lethal and the affected mouse fetuses die before birth. Still other knockouts have intermediate effects. For example, consider the tumor suppressor gene called *p53*. Humans with defects in this gene are highly susceptible to certain cancers. Mice with their *p53* gene knocked out develop normally but are afflicted with tumors at an early age.

SUMMARY To probe the role of a gene, molecular biologists can perform targeted disruption of the corresponding gene in a mouse, and then look for the effects of that mutation on the "knockout mouse."

SUMMARY

Methods of purifying proteins and nucleic acids are crucial in molecular biology. DNAs, RNAs, and proteins of various sizes can be separated by gel electrophoresis. The most common gel used in nucleic acid electrophoresis is agarose, and polyacrylamide is usually used in protein electrophoresis. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) is used to separate polypeptides according to their sizes. High-resolution separation of polypeptides can be achieved by two-dimensional gel electrophoresis, which uses isoelectric focusing in the first dimension and SDS-PAGE in the second.

Ion-exchange chromatography can be used to separate substances, including proteins, according to their charges. Positively charged resins like DEAE-Sephadex are used for anion-exchange chromatography, and negatively charged resins like phosphocellulose are used for cation-exchange chromatography. Gel filtration chromatography uses columns filled with porous resins that let smaller substances in, but exclude larger substances. Thus, the smaller substances are slowed, but larger substances travel relatively rapidly through the column. Affinity chromatography is a powerful purification technique that exploits an affinity reagent with strong and specific affinity for a molecule of interest. That molecules binds to a column coupled to the affinity reagent but all or most other molecules flow through without binding. Then the molecule of interest can be eluted from the column with a solution of a substance that disrupts the specific binding.

Detection of the tiny quantities of substances in molecular biology experiments generally requires labeled tracers. If the tracer is radioactive it can be detected by autoradiography, using x-ray film or a phosphorimager, or by liquid scintillation counting. Some very sensitive nonradioactive labeled tracers are also now available. These produce light (chemiluminescence) or colored spots.

Labeled DNA (or RNA) probes can be hybridized to DNAs of the same, or very similar, sequence on a Southern blot. Modern DNA typing uses Southern blots and a battery of DNA probes to detect variable sites in individual animals, including humans. As a forensic tool, DNA typing can be used to test parentage, to identify criminals, or to remove innocent people from suspicion.

Labeled probes can be hydridized to whole chromosomes to locate genes or other specific DNA sequences. This type of procedure is called in situ hybridization; if the probe is fluorescently labeled, the technique is called fluorescence in situ hybridization (FISH). Proteins can be detected and quantified in complex mixtures using immunoblots (or Western blots). Proteins are electrophoresed, then blotted to a membrane and the proteins on the blot are probed with specific antibodies that can be detected with labeled secondary antibodies or protein A.

The Sanger DNA sequencing method uses dideoxy nucleotides to terminate DNA synthesis, yielding a series of DNA fragments whose sizes can be measured by electrophoresis. The last base in each of these fragments is known because we know which dideoxy nucleotide was used to terminate each reaction. Therefore, ordering these fragments by size—each fragment one (known) base longer than the next—tells us the base sequence of the DNA.

A physical map depicts the spatial arrangement of physical "landmarks," such as restriction sites, on a DNA molecule. We can improve a restriction map considerably by Southern blotting some of the fragments and then hybridizing these fragments to labeled fragments generated by another restriction enzyme. This reveals overlaps between individual restriction fragments.

Using cloned genes, we can introduce changes conveniently by site-directed mutagenesis, thus altering the amino acid sequences of the protein products. The mutagenized DNA can be made with double-stranded DNA, two complementary mutagenic primers, and PCR. Wild-type DNA can be eliminated by cleavage with *Dpn*I, so clones are transformed primarily with mutagenized DNA, not with wild-type.

A Northern blot is similar to a Southern blot, but it contains electrophoretically separated RNAs instead of

DNAs. The RNAs on the blot can be detected by hybridizing them to a labeled probe. The intensities of the bands reveal the relative amounts of specific RNA in each, and the positions of the bands indicate the lengths of the respective RNAs.

In S1 mapping, a labeled DNA probe is used to detect the 5′- or 3′-end of a transcript. Hybridization of the probe to the transcript protects a portion of the probe from digestion by S1 nuclease, which specifically degrades single-stranded polynucleotides. The length of the section of probe protected by the transcript locates the end of the transcript, relative to the known location of an end of the probe. Because the amount of probe protected by the transcript is proportional to the concentration of transcript, S1 mapping can also be used as a quantitative method. RNase mapping is a variation on S1 mapping that uses an RNA probe and RNase instead of a DNA probe and S1 nuclease.

Using primer extension one can locate the 5′-end of a transcript by hybridizing an oligonucleotide primer to the RNA of interest, extending the primer with reverse transcriptase to the 5′-end of the transcript, and electrophoresing the reverse transcript to determine its size. The intensity of the signal obtained by this method is a measure of the concentration of the transcript.

Run-off transcription is a means of checking the efficiency and accuracy of in vitro transcription. One truncates a gene in the middle and transcribes it in vitro in the presence of labeled nucleotides. The RNA polymerase runs off the end and releases an incomplete transcript. The size of this run-off transcript locates the transcription start site, and the amount of this transcript reflects the efficiency of transcription. G-less cassette transcription also produces a shortened transcript of predictable size, but does so by placing a G-less cassette just downstream of a promoter and transcribing this construct in the absence of GTP.

Nuclear run-on transcription is a way of ascertaining which genes are active in a given cell by allowing transcription of these genes to continue in isolated nuclei. Specific transcripts can be identified by their hybridization to known DNAs on dot blots. One can also use the run-on assay to determine the effects of assay conditions on nuclear transcription.

To measure the activity of a promoter, one can link it to a reporter gene, such as the genes encoding β-galactosidase, CAT, or luciferase, and let the easily assayed reporter gene products tell us indirectly the activity of the promoter. One can also use reporter genes to detect changes in translational efficiency after altering regions of a gene that affect translation.

Gene expression can be quantified by measuring the accumulation of the protein products of genes by immunoblotting or immunoprecipitation.

Filter binding as a means of measuring DNA–protein interaction is based on the fact that double-stranded DNA will not bind by itself to a nitrocellulose filter, or similar medium, but a protein–DNA complex will. Thus, one can label a double-stranded DNA, mix it with a protein, and assay protein–DNA binding by measuring the amount of label retained by the filter. A gel mobility shift assay detects interaction between a protein and DNA by the reduction of the electrophoretic mobility of a small DNA that occurs when the DNA binds to a protein.

Footprinting is a means of finding the target DNA sequence, or binding site, of a DNA-binding protein. We perform DNase footprinting by binding the protein to its DNA target, then digesting the DNA–protein complex with DNase. When we electrophorese the resulting DNA fragments, the protein binding site shows up as a gap, or "footprint," in the pattern where the protein protected the DNA from degradation. DMS footprinting follows a similar principle, except that we use the DNA methylating agent DMS, instead of DNase, to attack the DNA–protein complex. Unmethylated (or hypermethylated) sites show up on electrophoresis and demonstrate where the protein is bound to the DNA. Hydroxyl radical footprinting uses organometallic complexes to generate hydroxyl radicals that break DNA strands.

SELEX is a method that allows one to find RNA sequences that interact with other molecules, including proteins. RNAs that interact with a target molecule are selected by affinity chromatography, then converted to double-stranded DNAs and amplified by PCR. After several rounds of this procedure, the RNAs are highly enriched for sequences that bind to the target molecule. Functional SELEX is a variation on this theme in which the desired function somehow alters the RNA so it can be amplified. If the desired function is enzymatic, mutagenesis can be introduced into the amplification step to produce variants with higher activity.

To probe the role of a gene, one can perform targeted disruption of the corresponding gene in a mouse, and then look for the effects of that mutation on the "knockout mouse."

REVIEW QUESTIONS

1. Use a drawing to illustrate the principle of DNA gel electrophoresis. Indicate roughly the comparative electrophoretic mobilities of DNAs with 150, 600, and 1200 bp.
2. What is SDS? What are its functions in SDS-PAGE?
3. Compare and contrast SDS-PAGE and modern two-dimensional gel electrophoresis of proteins.

4. Describe the principle of ion-exchange chromatography. Use a graph to illustrate the separation of three different proteins by this method.
5. Describe the principle of gel filtration chromatography. Use a graph to illustrate the separation of three different proteins by this method. Indicate on the graph the largest and smallest of these proteins.
6. Compare and contrast the principles of autoradiography and phosphorimaging. Which method provides more quantitative information?
7. Describe a nonradioactive method for detecting a particular nucleic acid fragment in an electrophoretic gel.
8. Diagram the process of Southern blotting and probing to detect a DNA of interest. Compare and contrast this procedure with Northern blotting.
9. Describe a DNA fingerprinting method using a minisatellite probe. Compare this method with a modern forensic DNA typing method using probes to detect single variable DNA loci.
10. What kinds of information can we obtain from a Northern blot?
11. Describe fluorescence in situ hybridization (FISH). When would you use this method, rather than Southern blotting?
12. Draw a diagram of an imaginary Sanger sequencing autoradiograph, and provide the corresponding DNA sequence.
13. Show how a manual DNA sequencing method can be automated.
14. Show how to use restriction mapping to determine the orientation of a restriction fragment ligated into a restriction site in a vector. Use fragment sizes different from those in the text.
15. Explain the principle of site-directed mutagenesis, then describe a method to carry out this process.
16. Compare and contrast the S1 mapping and primer extension methods for mapping the 5′-end of an mRNA. Which of these methods can be used to map the 3′-end of an mRNA. Why would the other method not work?
17. Describe the run-off transcription method. Why does this method not work with in vivo transcripts, as S1 mapping and primer extension do?
18. How would you label the 5′-ends of a double-stranded DNA? The 3′-ends?
19. Describe a nuclear run-on assay, and show how it differs from a run-off assay.
20. How does a dot blot differ from a Southern blot?
21. Describe the use of a reporter gene to measure the strength of a promoter.
22. Describe a filter-binding assay to measure binding between a DNA and a protein.
23. Compare and contrast the gel mobility shift and DNase footprinting methods of assaying specific DNA–protein interactions. What information does DNase footprinting provide that gel mobility shift does not?
24. Compare and contrast DMS and DNase footprinting. Why is the former method more precise than the latter?
25. Describe a method for creating a knockout mouse. Explain the importance of the thymidine kinase and neomycin-resistance genes in this procedure. What information can a knockout mouse provide?

ANALYTICAL QUESTIONS

1. You have electrophoresed some DNA fragments on an agarose gel and obtain the results shown in Figure 5.2. (a) What is the size of a fragment that migrated 25 mm? (b) How far did the 200 bp fragment migrate?
2. Design a Southern blot experiment to check a chimeric mouse's DNA for insertion of the neomycin-resistance gene. You may assume any array of restriction sites you wish in the target gene and in the *neo*^{r} gene. Show sample results for a successful and an unsuccessful insertion.
3. In a DNase footprinting experiment, either the template or nontemplate strand can be end-labeled. In Figure 5.37a, the template strand is labeled. Which strand is labeled in Figure 5.37b? How do you know?

SUGGESTED READINGS

Galas, D.J. and A. Schmitz. 1978. DNase footprinting: A simple method for the detection of protein–DNA binding specificity. *Nucleic Acids Research* 5:3157–70.

Lichter, P. 1990. High resolution mapping of human chromosome 11 by in situ hybridization with cosmid clones. *Science* 247:64–69.

Sambrook, J., et al. 2001. *Molecular Cloning: A Laboratory Manual,* 3rd ed. Plainview, NY: Cold Spring Harbor Laboratory Press.