Protein Purification and Characterization Techniques

Column chromatography is widely used to purify proteins.

5.1 Extracting Pure Proteins from Cells

Many different proteins exist in a single cell. A detailed study of the properties of any one protein requires a homogeneous sample consisting of only one kind of molecule. The separation and isolation, or purification, of proteins constitutes an essential first step to further experimentation. In general, separation techniques focus on size, charge, and polarity—the sources of differences between molecules. Many techniques are performed to eliminate contaminants and to arrive at a pure sample of the protein of interest. As the purification steps are followed, we make a table of the recovery and purity of the protein to gauge our success. Table 5.1 shows a typical purification for an enzyme. The **percent recovery** column tracks how much of the protein of interest has been retained at each step. This number usually drops steadily during the purification, and we hope that by the time the protein is pure, sufficient product will be left for study and characterization. The specific activity column compares the purity of the protein at each step, and this value should go up if the purification is successful.

How do we get the proteins out of the cells?

Before the real purification steps can begin, the protein must be released from the cells and subcellular organelles. The first step, called **homogenization,** involves breaking open the cells. This can be done with a wide variety of techniques. The simplest approach is grinding the tissue in a blender with a suitable buffer. The cells are broken open, releasing soluble proteins. This process also breaks many of the subcellular organelles, such as mitochondria, peroxisomes, and endoplasmic reticulum. A gentler technique is to use a Potter–Elvehjem homogenizer, a thick-walled test tube through which a tight-fitting plunger is passed. The squeezing of the homogenate around the plunger breaks open cells, but it leaves many of the organelles intact. Another technique, called sonication, involves using sound waves to break

TABLE 5.1

Example of a Protein Purification Scheme: Purification of the Enzyme Xanthine Dehydrogenase from a Fungus

Fraction	Volume (mL)	Total Protein (mg)	Total Activity	Specific Activity	Percent Recovery
1. Crude extract	3,800	22,800	2,460	0.108	100
2. Salt precipitate	165	2,800	1,190	0.425	48
3. Ion-exchange chromatography	65	100	720	7.2	29
4. Molecular-sieve chromatography	40	14.5	555	38.3	23
5. Immunoaffinity chromatography	6	1.8	275	152.108	11

Chapter Outline

OWL Online homework for this chapter may be assigned in OWL.

open the cells. Cells can also be ruptured by cycles of freezing and thawing. If the protein of interest is solidly attached to a membrane, detergents may have to be added to detach the proteins. After the cells are homogenized, they are subjected to **differential centrifugation.**

Spinning the sample at 600 times the force of gravity ($600 \times g$) results in a pellet of unbroken cells and nuclei. If the protein of interest is not found in the nuclei, this precipitate is discarded. The supernatant can then be centrifuged at higher speed, such as $15{,}000 \times g$, to bring down the mitochondria. Further centrifugation at $100{,}000 \times g$ brings down the microsomal fraction, consisting of ribosomes and membrane fragments. If the protein of interest is soluble, the supernatant from this spin will be collected and will already be partially purified because the nuclei and mitochondria will have been removed. Figure 5.1 shows a typical separation via differential centrifugation.

After the proteins are solubilized, they are often subjected to a crude purification based on solubility. Ammonium sulfate is the most common reagent to use at this step, and this procedure is referred to as **salting out.** Proteins have varying solubilities in polar and ionic compounds. Proteins remain soluble because of their interactions with water. When ammonium sulfate is added to a protein solution, some of the water is taken away from the protein to make ion–dipole bonds with the salts. With less water available to hydrate the proteins, they begin to interact with each other through hydrophobic bonds. At a defined amount of ammonium sulfate, a precipitate that contains contaminating proteins forms. These proteins are centrifuged down and discarded. Then more salt is added, and a different set of proteins, which usually contains the protein of interest, precipitates. This precipitate is collected by centrifugation and saved. The quantity of ammonium sulfate is usually measured in comparison with a 100% saturated solution. A common procedure involves bringing the solution to around 40% saturation and then spinning down the precipitate that forms. Next, more ammonium sulfate is added to the supernatant, often to a level of 60%–70% saturation. The precipitate that forms often contains the protein of interest. These preliminary techniques do not generally give a sample that is very pure, but they serve the important task of preparing the crude homogenate for the more effective procedures that follow.

5.2 Column Chromatography

The word *chromatography* comes from the Greek *chroma,* "color," and *graphein,* "to write"; the technique was first used around the beginning of the 20th century to separate plant pigments with easily visible colors. It has long since been possible to separate colorless compounds, as long as methods exist for detecting them. Chromatography is based on the fact that different compounds can distribute themselves to varying extents between different phases, or separable portions of matter. One phase is the **stationary phase,** and the other is the **mobile phase.** The mobile phase flows over the stationary material and carries the sample to be separated along with it. The components of the sample interact with the stationary phase to different extents. Some components interact relatively strongly with the stationary phase and are therefore carried along more slowly by the mobile phase than are those that interact less strongly. The differing mobilities of the components are the basis of the separation.

Many chromatographic techniques used for research on proteins are forms of **column chromatography,** in which the material that makes up the stationary phase is packed in a column. The sample is a small volume of concentrated solution that is applied to the top of the column; the mobile phase, called the *eluent,* is passed through the column. The sample is diluted by the eluent, and the separation process also increases the volume occupied by the sample.

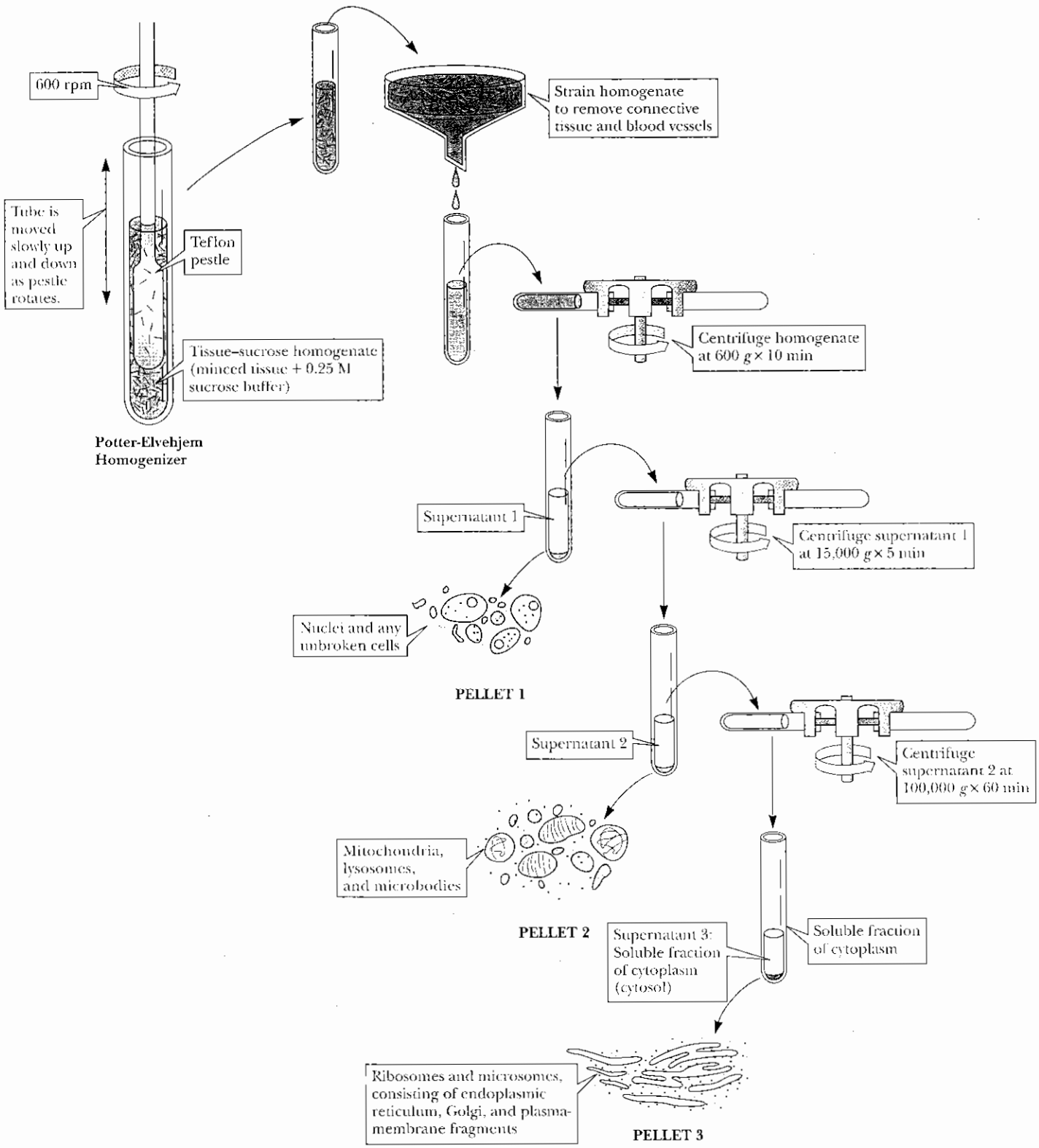
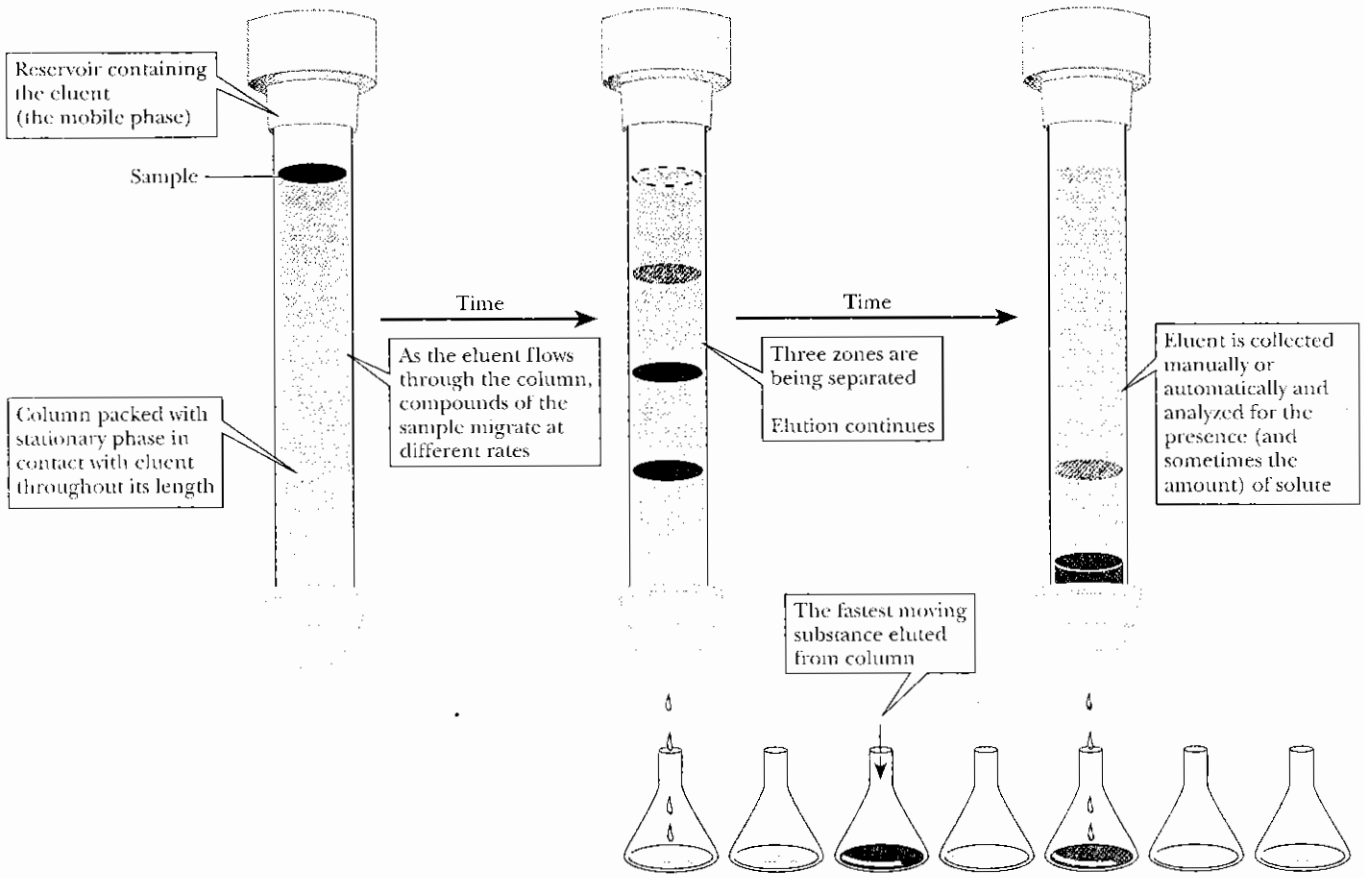


FIGURE 5.1 Differential centrifugation. Differential centrifugation is used to separate cell components. As a cell homogenate is subjected to increasing *g* forces, smaller and smaller cell components end up in the pellet.

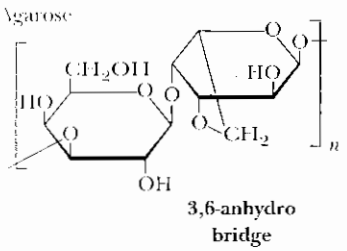


■ FIGURE 5.2 **Column chromatography.** A sample containing several components is applied to the column. The various components travel at different rates and can be collected individually.

In a successful experiment, the entire sample eventually comes off the column. Figure 5.2 diagrams an example of column chromatography.

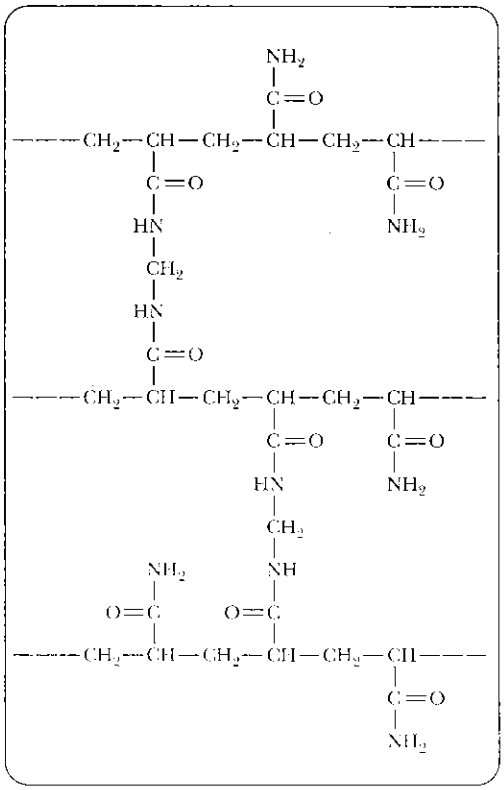
What are the different types of chromatography?

Size-exclusion chromatography, also called **gel-filtration chromatography,** separates molecules on the basis of size, making it a useful way to sort proteins of varied molecular weights. It is a form of column chromatography in which the stationary phase consists of cross-linked gel particles. The gel particles are usually in bead form and consist of one of two kinds of polymers. The first is a carbohydrate polymer, such as **dextran** or **agarose;** these two polymers are often referred to by the trade names Sephadex and Sepharose, respectively (Figure 5.3). The second is based on **polyacrylamide** (Figure 5.4), which is sold under the trade name Bio-Gel. The cross-linked structure of these polymers produces pores in the material. The extent of cross-linking can be controlled to select a desired pore size. When a sample is applied to the column, smaller molecules, which are able to enter the pores, tend to be delayed in their progress down the column, unlike the larger molecules. As a result, the larger molecules are eluted first, followed later by the smaller ones, after escaping from the pores. Molecular-sieve chromatography is represented schematically in Figure 5.5. The advantages of this type of chromatography are (1) its convenience as a way to separate molecules on the basis of size and (2) the fact that it can be used to estimate molecular weight by comparing the sample with a set of standards. Each type of gel used has a specific range of sizes that separate linearly with the log of the molecular weight. Each gel also has an exclusion limit, a size of protein that is too large to fit inside the pores. All proteins that size or larger elute first and simultaneously.

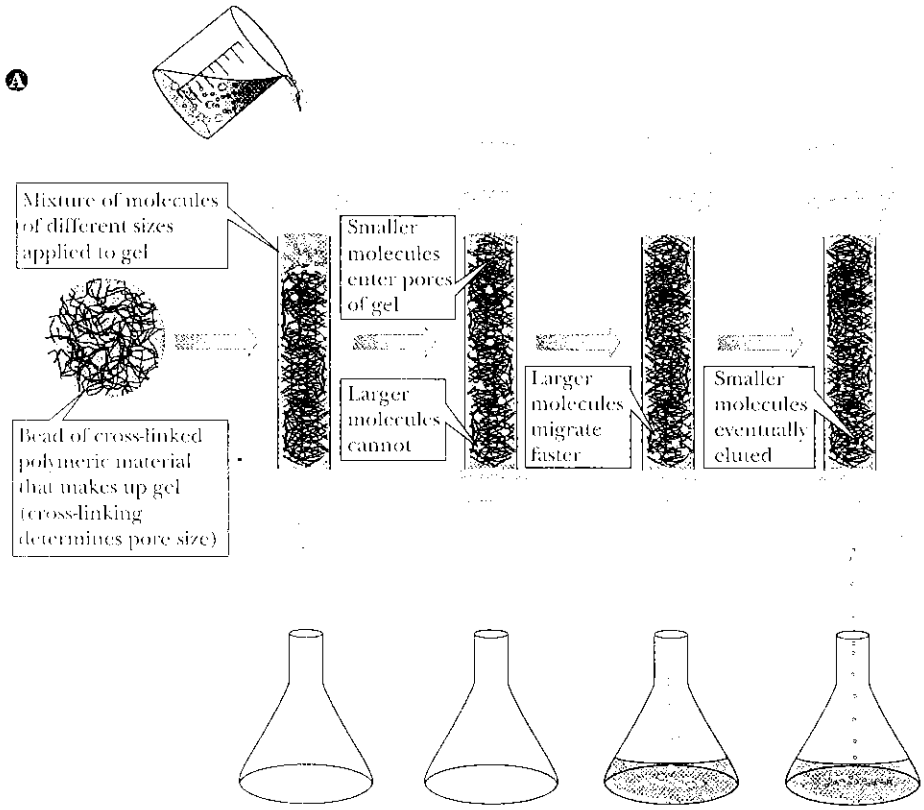


RE 5.3 **The repeating disaccharide unit of se that is used for column chromatography.**

Affinity chromatography uses the specific binding properties of many proteins. It is another form of column chromatography with a polymeric material used as the stationary phase. The distinguishing feature of affinity chromatography is that the polymer is covalently linked to some compound, called a *ligand,* that binds specifically to the desired protein (Figure 5.6). The other proteins in the sample do not bind to the column and can easily be eluted with buffer, while the bound protein remains on the column. The bound protein can then be eluted from the column by adding high concentrations of the ligand in soluble form, thus competing for the binding of the protein with the stationary phase. The protein binds to the ligand in the mobile phase and is recovered from the column. This protein–ligand interaction can also be disrupted with a change in pH or ionic strength. Affinity chromatography is a convenient separation method and has the advantage of producing very pure proteins. Some affinity ligands are designed to be completely specific



■ FIGURE 5.4 The structure of cross-linked polyacrylamide.



■ FIGURE 5.5 Gel-filtration chromatography. Larger molecules are excluded from the gel and move more quickly through the column. Small molecules have access to the interior of the gel beads, so they take a longer time to elute. In column chromatography procedures such as this, the protein concentration is usually measured by UV absorption as the samples elute from the column.

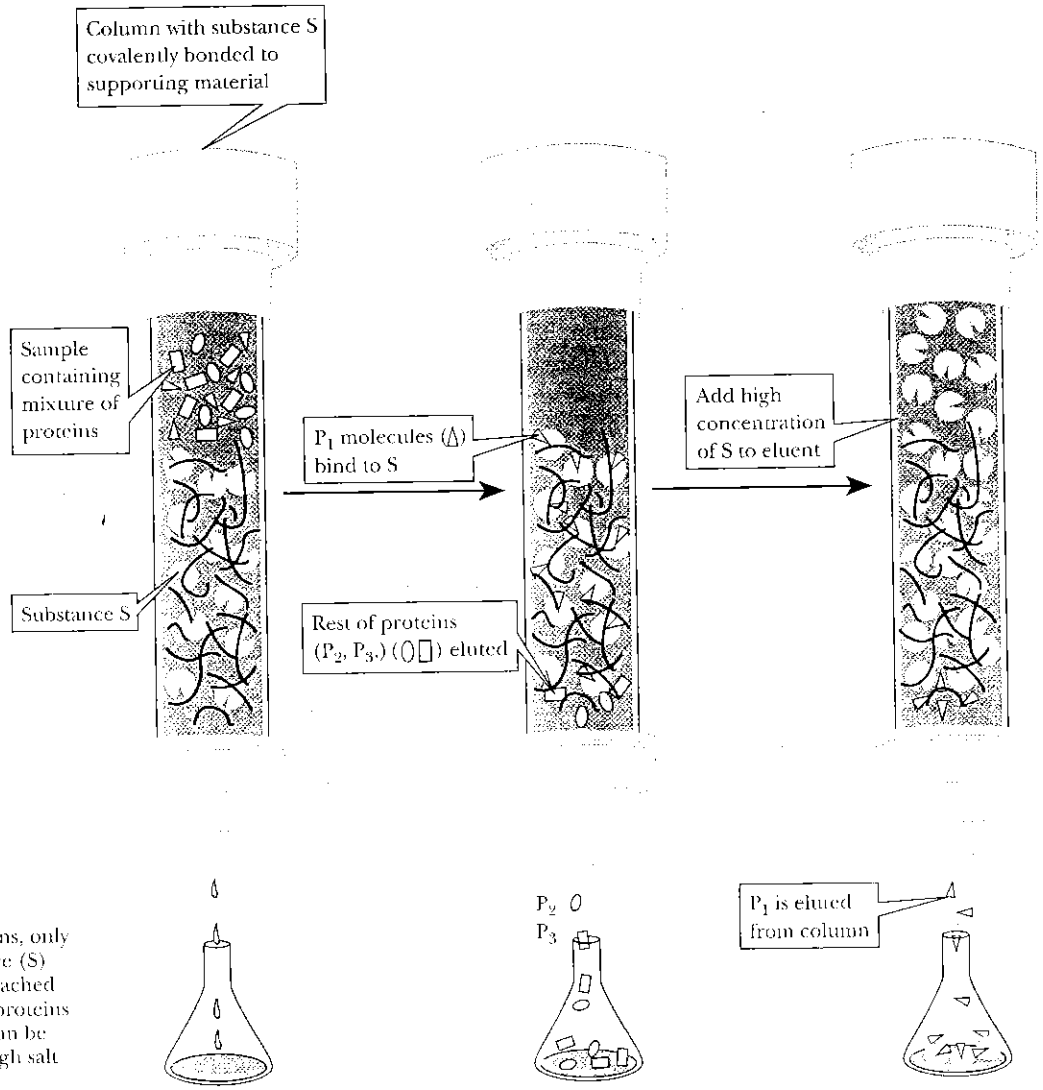


FIGURE 5.6 The principle of affinity chromatography. In a mixture of proteins, only one (designated P_1) binds to a substance (S) that is the substrate. The substrate is attached to the column matrix. Once the other proteins (designated P_2) have been washed out, P_1 can be eluted either by adding a solution of high salt concentration or by adding free S.

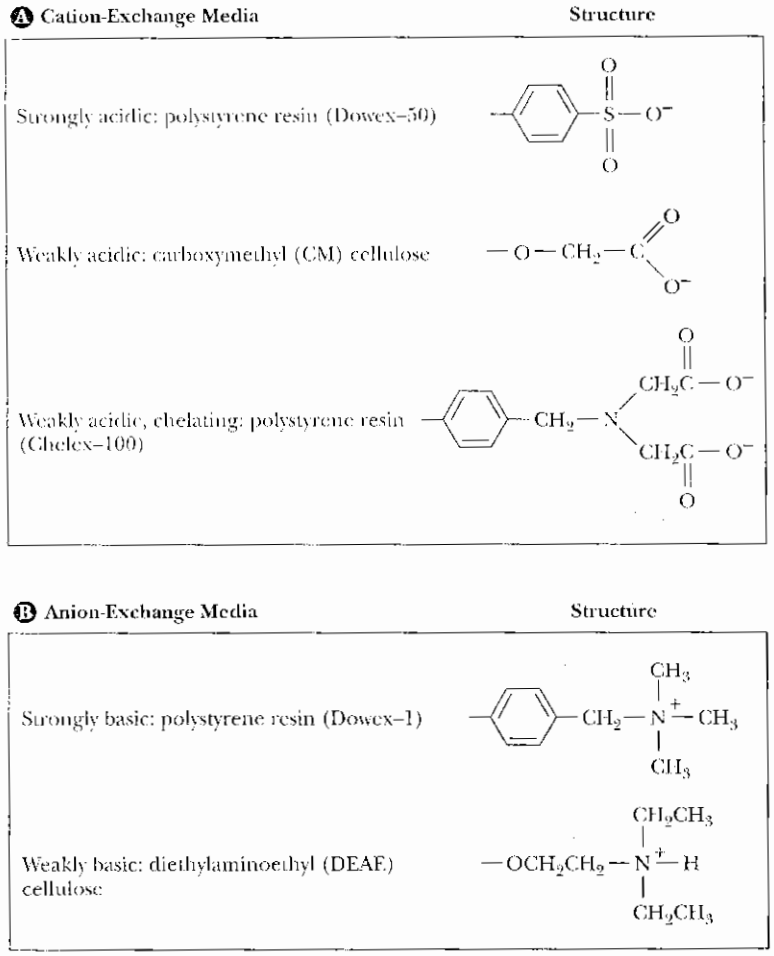
TABLE 5.2

Group-Specific Affinity Resins

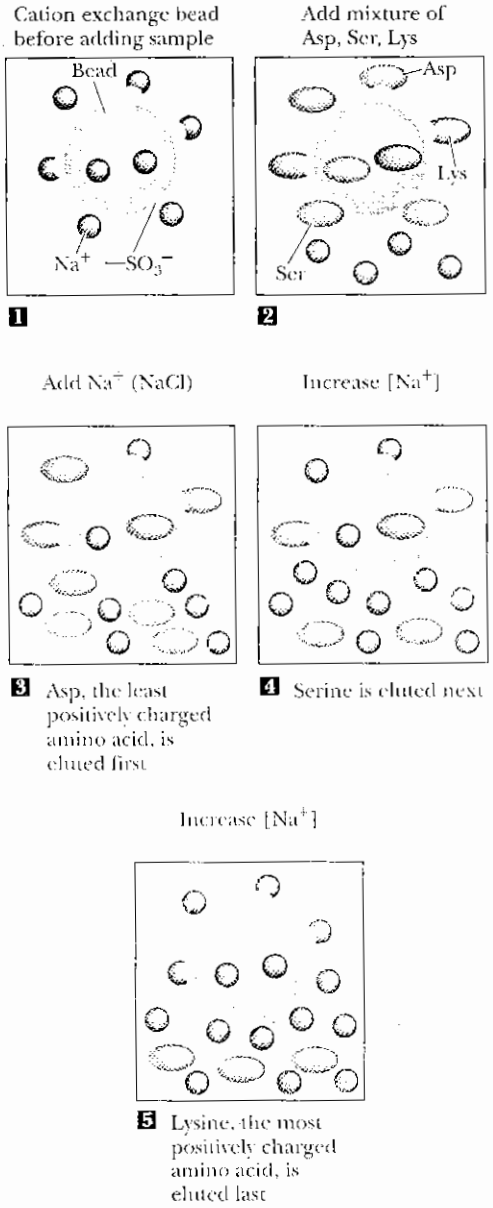
Group-Specific Adsorbent	Group Specificity
Concanavalin A–agarose	Glycoproteins and glycolipids
Cibacron Blue–agarose	Enzymes with nucleotide cofactors
Boronic acid–agarose	Compounds with cis-diol groups
Protein A–agarose	IgG-type antibodies
Poly(U)–agarose	Nucleic acids containing poly(A) sequences
Poly(A)–agarose	Nucleic acids containing poly(U) sequences
Iminodiacetate–agarose	Proteins with heavy metal affinity
AMP–agarose	Enzymes with NAD^+ cofactors, ATP-dependent kinases

for a molecule the scientist is interested in purifying. However, this is often very expensive. There are other ligands that are specific for groups of compounds. Table 5.2 lists some group-specific affinity resins. The Biochemical Connections box in Chapter 13 describes an interesting way in which affinity chromatography can be combined with molecular biological techniques to offer a one-step purification of a protein.

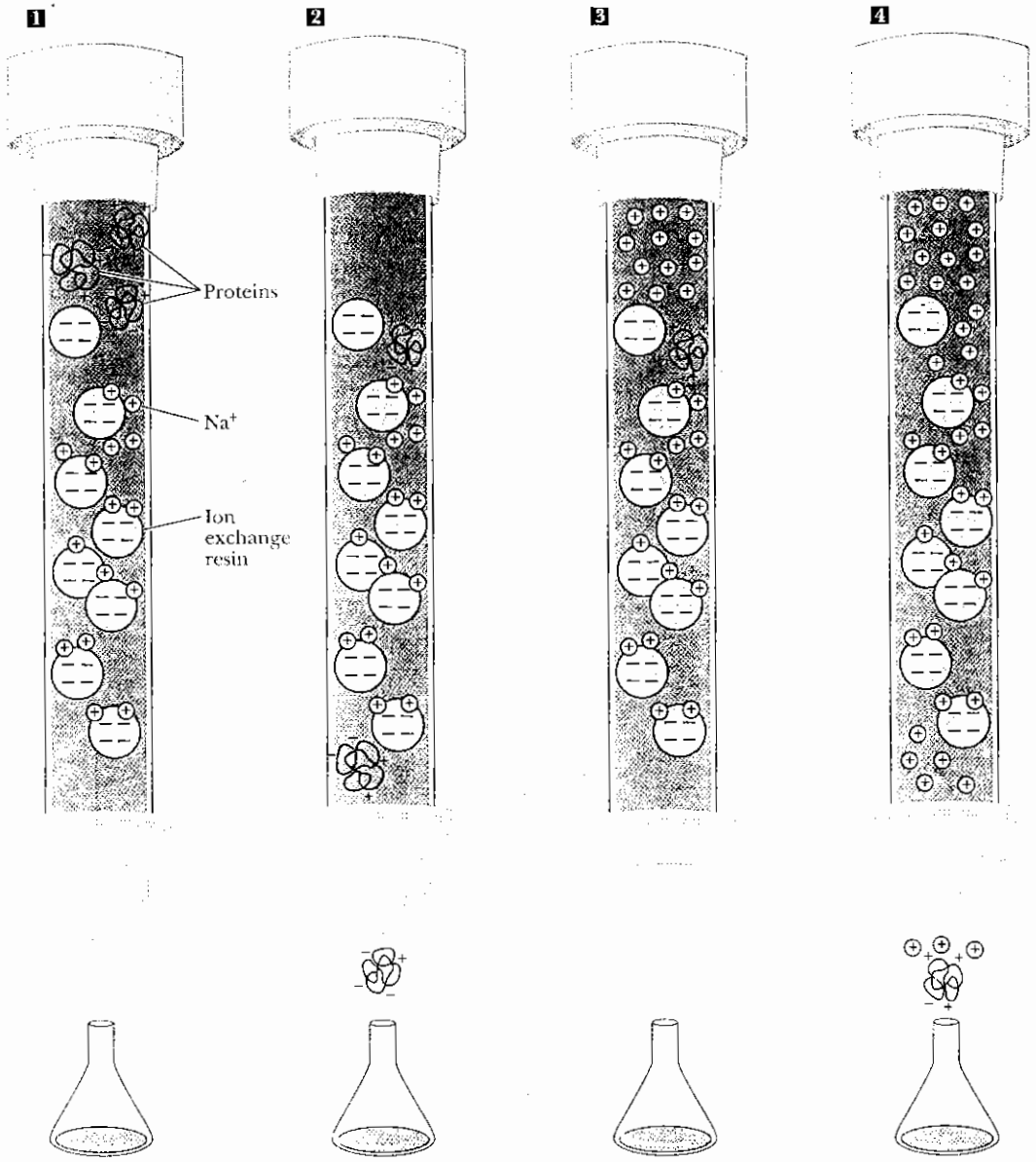
Ion-exchange chromatography is logistically similar to affinity chromatography. Both use a column resin that binds the protein of interest. With ion-exchange chromatography, however, the interaction is less specific and is based on net charge. An ion-exchange resin has a ligand with a positive charge or a negative charge. A negatively charged resin is a **cation exchanger,** and a positively charged one is an **anion exchanger.** Figure 5.7 shows some typical ion-exchange ligands. Figure 5.8 illustrates their principle of operation with three amino acids of different charge. Figure 5.9 shows how cation exchange chromatography would separate proteins. The column is initially equilibrated with a buffer of suitable pH and ionic strength. The exchange resin is bound to counterions. A cation-exchange resin is usually bound to Na^+ or K^+ ions, and an anion exchanger is usually bound to Cl^- ions. A mixture of proteins is loaded on the column and allowed to flow through it. Proteins that have a net charge opposite to that of the exchanger stick to the column, *exchanging* places with the bound counterions. Proteins that have no net charge or have the same charge as the exchanger elute. After all the nonbinding proteins are eluted, the eluent is changed either to a buffer that has a pH that removes the charge on the bound proteins or to one with a higher salt concentration. The latter outcompetes the bound proteins for the limited binding space on the column. The once-bound molecules then elute, having been separated from many of the contaminating ones.



■ FIGURE 5.7 **Resins used in ion-exchange chromatography.** (a) Cation-exchange resins and (b) anion-exchange resins commonly used for biochemical separations.



■ FIGURE 5.8 **Operation of a cation-exchange column to separate a mixture of aspartate, serine, and lysine.** (1) The cation-exchange resin in the initial Na^+ form. (2) A mixture of aspartate, serine, and lysine is added to the column containing the resin. (3) A gradient of the eluting salt (for example, NaCl) is added to the column. Aspartate, the least positively charged amino acid, is eluted first. (4) As the salt concentration increases, serine is eluted. (5) As the salt concentration is increased further, lysine, the most positively charged of the three amino acids, is eluted last.



■ FIGURE 5.9 **Ion-exchange chromatography using a cation exchanger.** (1) At the beginning of the separation, various proteins are applied to the column. The column resin is bound to Na^+ counterions (small red spheres). (2) Proteins that have no net charge or a net negative charge pass through the column. Proteins that have a net positive charge stick to the column, displacing the Na^+. (3) An excess of Na^+ ion is then added to the column. (4) The Na^+ ions outcompete the bound proteins for the binding sites on the resin, and the proteins elute.

High Performance Liquid Chromatography (HPLC) exploits the same principles seen with other chromatographic techniques, but very high resolution columns that can be run under high pressures are used. High resolution separations can be effected very quickly using automated instrumentation. A separation that might take hours on a standard column can be done in minutes with HPLC. **Reverse Phase HPLC** is a widely used technique for the separation of nonpolar molecules. In reverse phase HPLC, a solution of nonpolar compounds is put through a column that has a nonpolar liquid immobilized on an inert matrix. A more polar liquid serves as the mobile phase and is passed over the matrix. The solute molecules are eluted in proportion to their solubility in the more polar liquid.

Apply Your Knowledge

Protein Purification

The table below shows some typical results for a protein purification. The protein being purified is the enzyme Lactate Dehydrogenase, which catalyzes a reaction between lactic acid and NAD^+ to give NADH and pyruvate. The purification of an enzyme is monitored by comparing the specific activity of the enzyme at various points in the purification. The specific activity is a measure of the activity of the enzyme divided by the weight of protein in the sample. The higher the number, the more pure the sample.

Purification Step	Total Activity of Enzyme (μmol product/min)	Specific Activity of Enzyme (μmol product/min/mg protein)
Crude Homogenate	100,000	0.15
20,000 × *g* supernatant	75,000	0.24
Salt precipitation	36,000	0.75
Ion Exchange Chromatography	12,000	3.4
Cibacron Blue Agarose Affinity Chromatography	6,000	42
Sephadex Gel Filtration Chromatography	500	90

Which step was the most effective at purifying the enzyme? Which step was the most costly in terms of enzyme recovery?

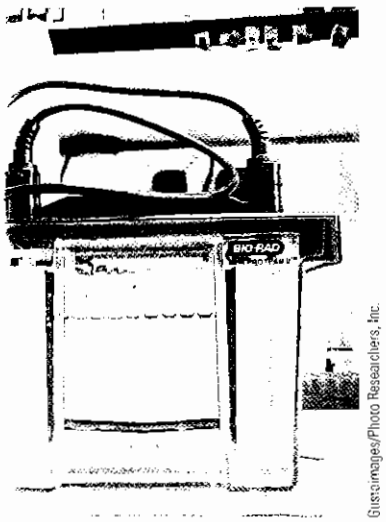
Solution

If we divide the specific activity of any purification step (fraction) by the one before it, we will get what is called the "fold purification" for that step. The larger the number, the more effective that step was. Adding that calculation to the table gives the following results:

Purification Step	Total Activity of Enzyme (μmol product/min)	Specific Activity of Enzyme (μmol product/min/mg protein)	Fold Purification
Crude Homogenate	100,000	0.15	n/a
20,000 × *g* supernatant	75,000	0.24	1.6
Salt precipitation	36,000	0.75	3.1
Ion Exchange Chromatography	12,000	3.4	4.5
Cibacron Blue Agarose Affinity Chromatography	6,000	42	12.4
Sephadex Gel Filtration Chromatography	500	90	2.1

The results show that the affinity chromatography step gave the highest single purification of 12.4 for a single step. This is often the case due to the power of this technique.

Similar calculations can show which step was the most costly by comparing the total activity of each fraction. The last step of the purification, the gel filtration chromatography, caused the loss of over 90% of the activity that was applied to it. This was the most costly single step. Scientists doing a purification must weigh the benefits and costs of each step in terms of purification and loss of product.



Gusto images/Photo Researchers, Inc.

The experimental setup for gel electrophoresis. The samples are placed on the top of the gel. When the current is applied, the negatively charged proteins migrate towards the positive electrode at the bottom.

Michael Gabridge/Visuals Unlimited

Separation of proteins by gel electrophoresis. Each band seen in the gel contains a different protein. In the SDS–PAGE (see the sample is treated with detergent before being applied to the gel. In isoelectric focusing, a pH gradient runs the length of the gel. The proteins in the gel have been stained with Coomassie Blue.

5.3 Electrophoresis

Electrophoresis is based on the motion of charged particles in an electric field toward an electrode of opposite charge. Macromolecules have differing mobilities based on their charge, shape, and size. Although many supporting media have been used for electrophoresis, including paper and liquid, the most common support is a polymer of agarose or acrylamide that is similar to those used for column chromatography. A sample is applied to wells that are formed in the supporting medium. An electric current is passed through the medium at a controlled voltage to achieve the desired separation (Figure 5.10). After the proteins are separated on the gel, the gel is stained to reveal the protein locations, as shown in Figure 5.11.

What is the difference between agarose gels and polyacrylamide gels?

Agarose-based gels are most often used to separate nucleic acids and will be discussed in Chapter 13. For proteins, the most common electrophoretic support is polyacrylamide (Figure 5.4), although sometimes agarose is used. A polyacrylamide gel is prepared and cast as a continuous cross-linked matrix, rather than being produced in the bead form employed in column chromatography. In one variation of polyacrylamide-gel electrophoresis, the protein sample is treated with the detergent sodium dodecyl sulfate (SDS) before it is applied to the gel. The structure of SDS is $CH_3(CH_2)_{10}CH_2OSO_3Na^+$. The anion binds strongly to proteins via nonspecific adsorption. The larger the protein, the more of the anion it adsorbs. SDS completely denatures proteins, breaking all the noncovalent interactions that determine tertiary and quaternary structure. This means that multisubunit proteins can be analyzed as the component polypeptide chains. All the proteins in a sample have a negative charge as a result of adsorption of the anionic SO_3^-. The proteins also have roughly the same shape, which is a random coil. In **SDS–polyacrylamide-gel electrophoresis (SDS–PAGE),** the acrylamide offers more resistance to large molecules than to small molecules. Because the shape and charge are approximately the same for all the proteins in the sample, the size of the protein becomes the determining factor in the separation: small proteins move faster than large ones. Like molecular-sieve chromatography, SDS–PAGE can be used to estimate the molecular weights of proteins by comparing the sample with standard samples. For most proteins, the log of the molecular weight is linearly related to its mobility on SDS–PAGE, as shown in Figure 5.12. Proteins can also be separated on acrylamide without the SDS, in which case the gel is called a **native gel.** This is useful for times when the study calls for a protein in its native conformation. In this case, however, the mobility is not correlated with size specifically, as three variables control the movement down the gel—size, shape, and charge.

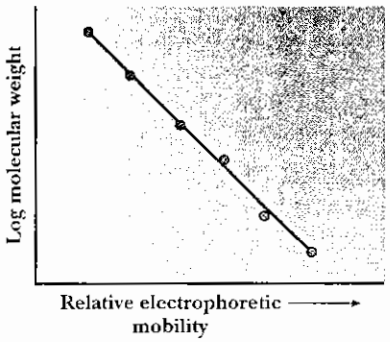
Isoelectric focusing is another variation of gel electrophoresis. Because different proteins have different titratable groups, they also have different isoelectric points. Recall (Section 3.3) that the isoelectric pH (pI) is the pH at which a protein (or amino acid or peptide) has no net charge. At the pI, the number of positive charges exactly balances the number of negative charges. In an isoelectric focusing experiment, the gel is prepared with a pH gradient that parallels the electric-field gradient. As proteins migrate through the gel under the influence of the electric field, they encounter regions of different pH, so the charge on the protein changes. Eventually each protein reaches the point at which it has no net charge—its isoelectric point—and no longer migrates. Each protein remains at the position on the gel corresponding to its pI, allowing for an effective method of separation.

An ingenious combination, known as two-dimensional gel electrophoresis (2-D gels), allows for enhanced separation by using isoelectric focusing in one dimension and SDS–PAGE run at 90° to the first (Figure 5.13).

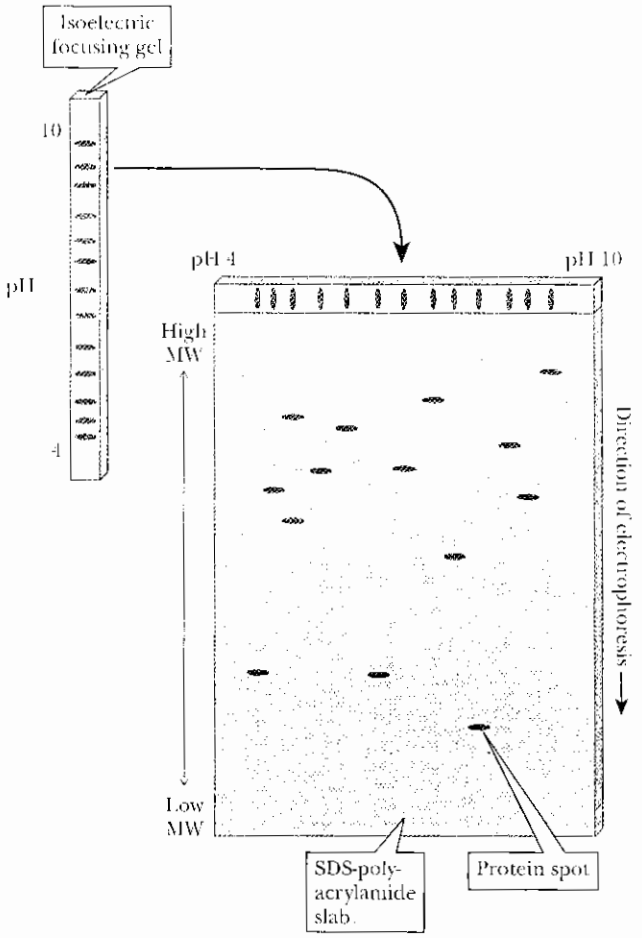
5.4 Determining the Primary Structure of a Protein

Determining the sequence of amino acids in a protein is a routine, but not trivial, operation in classical biochemistry. Its several parts must be carried out carefully to obtain accurate results (Figure 5.14).

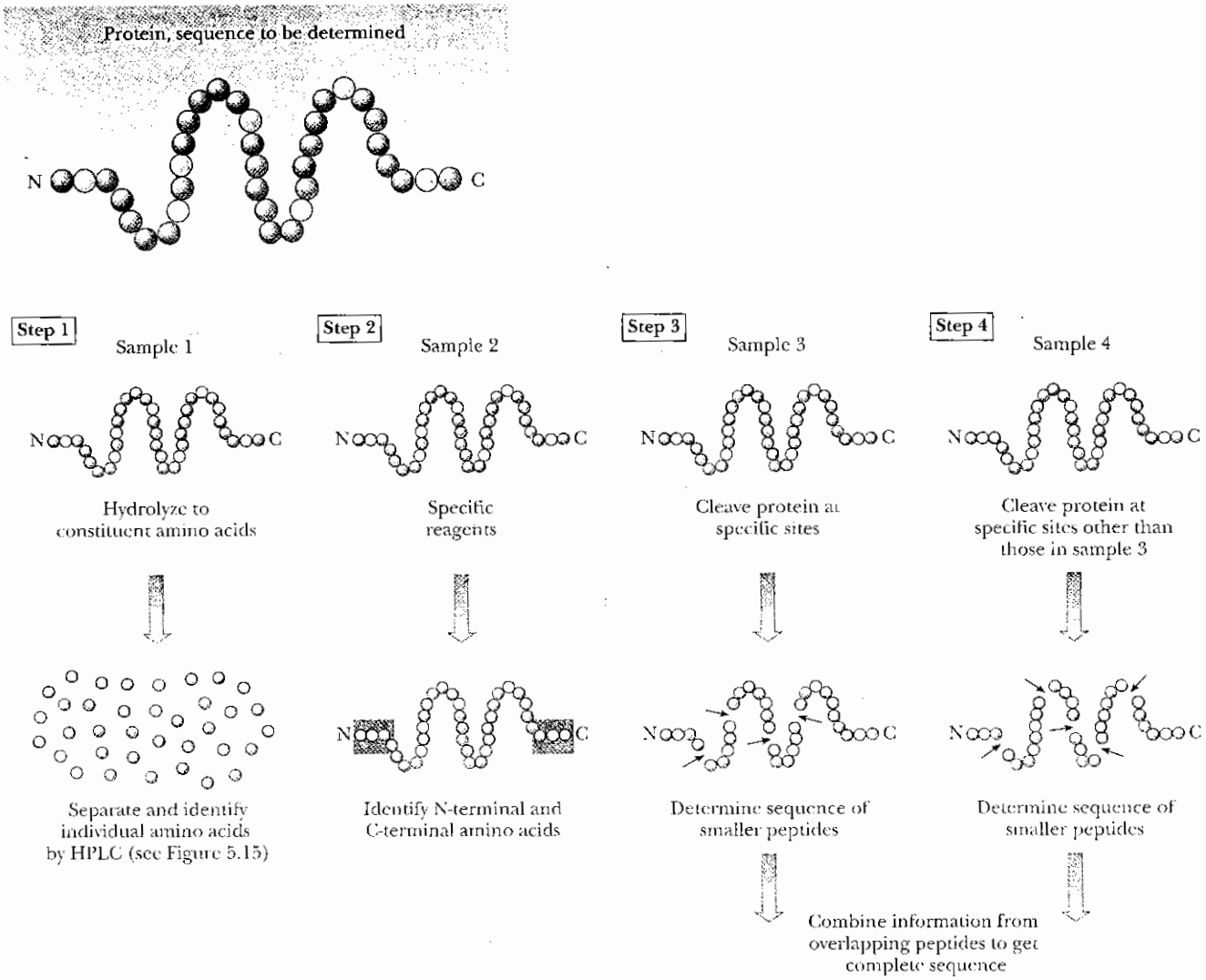
Step 1 in determining the primary structure of a protein is to establish which amino acids are present and in what proportions. Breaking a protein down to its component amino acids is relatively easy: heat a solution of the protein in acid, usually 6 *M* HCl, at 100°C to 110°C for 12 to 36 hours to hydrolyze the peptide bonds. Separation and identification of the products are somewhat more difficult and are best done by an amino acid analyzer. This automated instrument gives both qualitative information about the identities of the amino acids present and quantitative information about the relative amounts of those amino acids. Not only does it analyze amino acids, but it also allows informed decisions to be made about which procedures to choose later in the sequencing (see Steps 3 and 4 in Figure 5.14). An amino acid analyzer separates the mixture of amino acids either by ion-exchange chromatography or by **high-performance liquid chromatography (HPLC).** Figure 5.15 shows a typical result of amino acid separation with this technique.



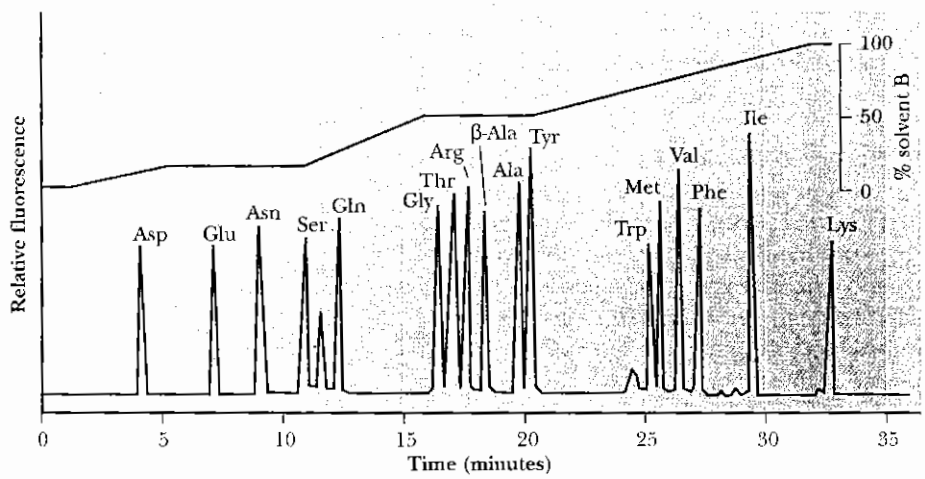
■ **FIGURE 5.12 Relationship between molecular weight and mobility.** A plot of the relative electrophoretic mobility of proteins in SDS–PAGE versus the log of the molecular weights of the individual polypeptides approximates a straight line.



■ **FIGURE 5.13 Two-dimensional electrophoresis.** A mixture of proteins is separated by isoelectric focusing in one direction. The focused proteins are then run using SDS–PAGE perpendicular to the direction of the isoelectric focusing. Thus the bands that appear on the gel have been separated first by their isoelectric points and then by size.



■ FIGURE 5.14 **The strategy for determining the primary structure of a given protein.** The amino acid can be determined by four different analyses performed on four separate samples of the same protein.



IRE 5.15 **HPLC chromatogram of amino separation.**

In Step 2, the identities of the N-terminal and C-terminal amino acids in a protein sequence are determined. This procedure is becoming less and less necessary as the sequencing of individual peptides improves, but it can be used to check whether a protein consists of one or two polypeptide chains.

In Steps 3 and 4, the protein is cleaved into smaller fragments, and the amino acid sequence is determined. Automated instruments can perform a stepwise modification starting from the N-terminal end, followed by cleavage of each amino acid in the sequence and the subsequent identification of each modified amino acid as it is removed. This process is called the **Edman degradation.**

Why are the proteins cleaved into small fragments for protein sequencing?

The Edman degradation method becomes more difficult as the number of amino acids increases. In most proteins, the chain is more than 100 residues long. For sequencing, it is usually necessary to break a long polypeptide chain into fragments, ranging from 20 to 50 residues for reasons that will be explained later.

Cleavage of the Protein into Peptides

Proteins can be cleaved at specific sites by enzymes or by chemical reagents. The enzyme **trypsin** cleaves peptide bonds preferentially at amino acids that have positively charged R groups, such as lysine and arginine. The cleavage takes place in such a way that the amino acid with the charged side chain ends up at the C-terminal end of one of the peptides produced by the reaction (Figure 5.16). The C-terminal amino acid of the original protein can be any one of the 20 amino acids and is not necessarily one at which cleavage takes place. A peptide can be automatically identified as the C-terminal end of the original chain if its C-terminal amino acid is not a site of cleavage.

Another enzyme, **chymotrypsin,** cleaves peptide bonds preferentially at the aromatic amino acids: tyrosine, tryptophan, and phenylalanine. The aromatic amino acid ends up at the C-terminal ends of the peptides produced by the reaction (Figure 5.17).

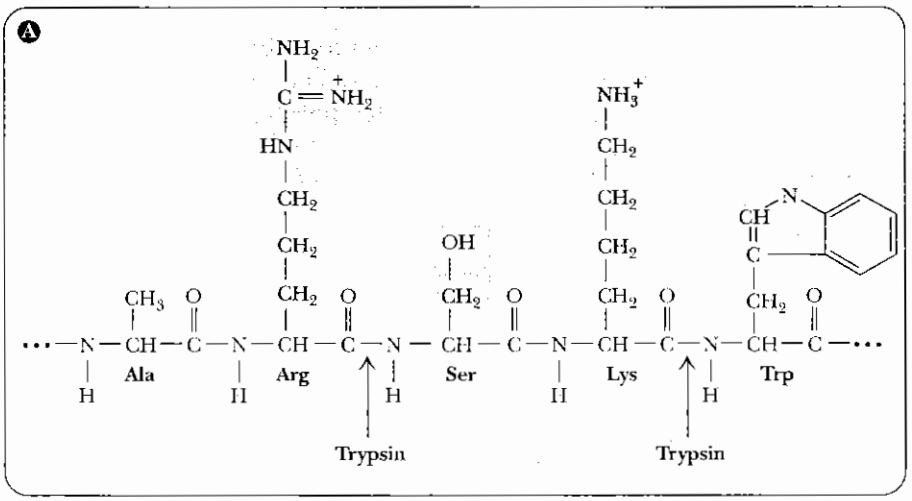
In the case of the chemical reagent **cyanogen bromide** (CNBr), the sites of cleavage are at internal methionine residues. The sulfur of the methionine reacts with the carbon of the cyanogen bromide to produce a homoserine lactone at the C-terminal end of the fragment (Figure 5.18).

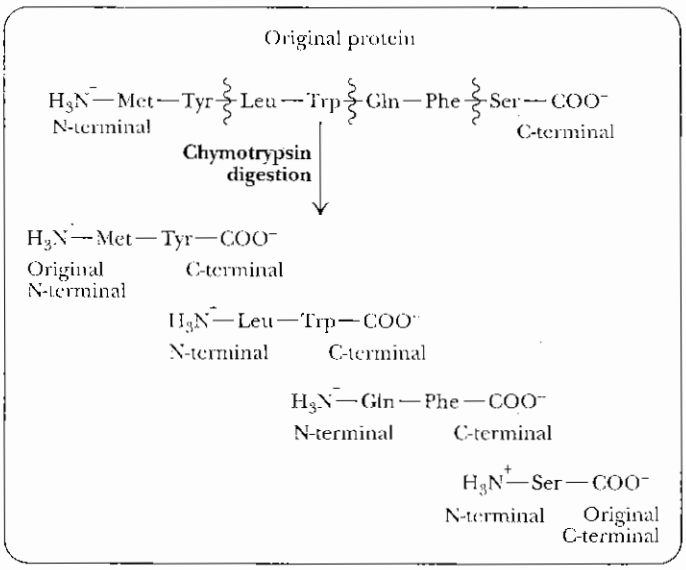
The cleavage of a protein by any of these reagents produces a mixture of peptides, which are then separated by high-performance liquid chromatography. The use of several such reagents on different samples of a protein to be sequenced produces different mixtures. The sequences of a set of peptides produced by one reagent overlap the sequences produced by another reagent (Figure 5.19). As a result, the peptides can be arranged in the proper order after their own sequences have been determined.

Sequencing of Peptides: The Edman Method

The actual sequencing of each peptide produced by specific cleavage of a protein is accomplished by repeated application of the Edman degradation. The sequence of a peptide containing 10 to 40 residues can be determined by this method in about 30 minutes using as little as 10 picomoles of material, with the range being based on the amount of purified fragment and the complexity of the sequence. For example, proline is more difficult to sequence than serine because of its chemical reactivity. (The amino acid sequences of the

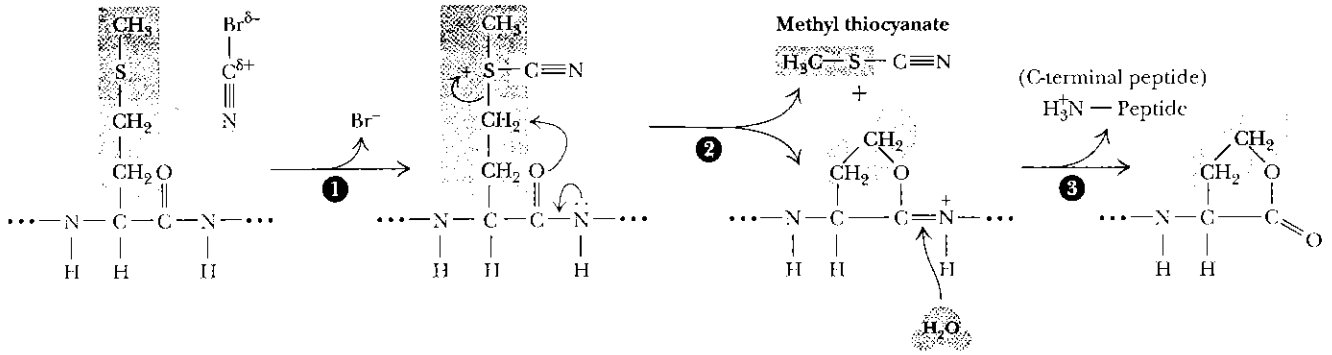
E 5.16 Peptide digestion with trypsin. psin is a proteolytic enzyme, or protease, ecifically cleaves only those peptide in which arginine or lysine contributes bonyl function. (b) The products of the n are a mixture of peptide fragments terminal Arg or Lys residues and a peptide derived from the polypeptide's inal end.



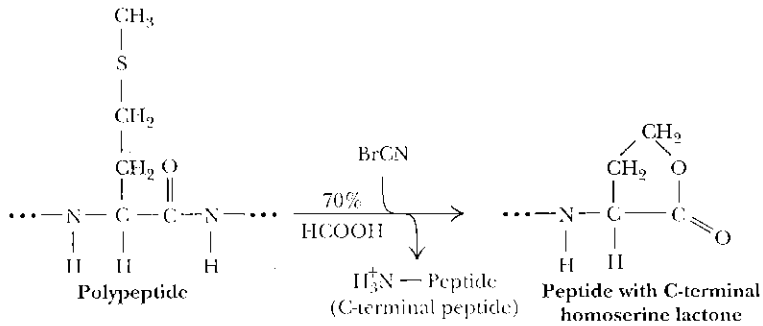


5.17 **Cleavage of proteins by rypsin.** Chymotrypsin hydrolyzes proteins atic amino acids.

individual peptides in Figure 5.19 are determined by the Edman method after the peptides are separated from one another.) The overlapping sequences of peptides produced by different reagents provide the key to solving the puzzle.



OVERALL REACTION:

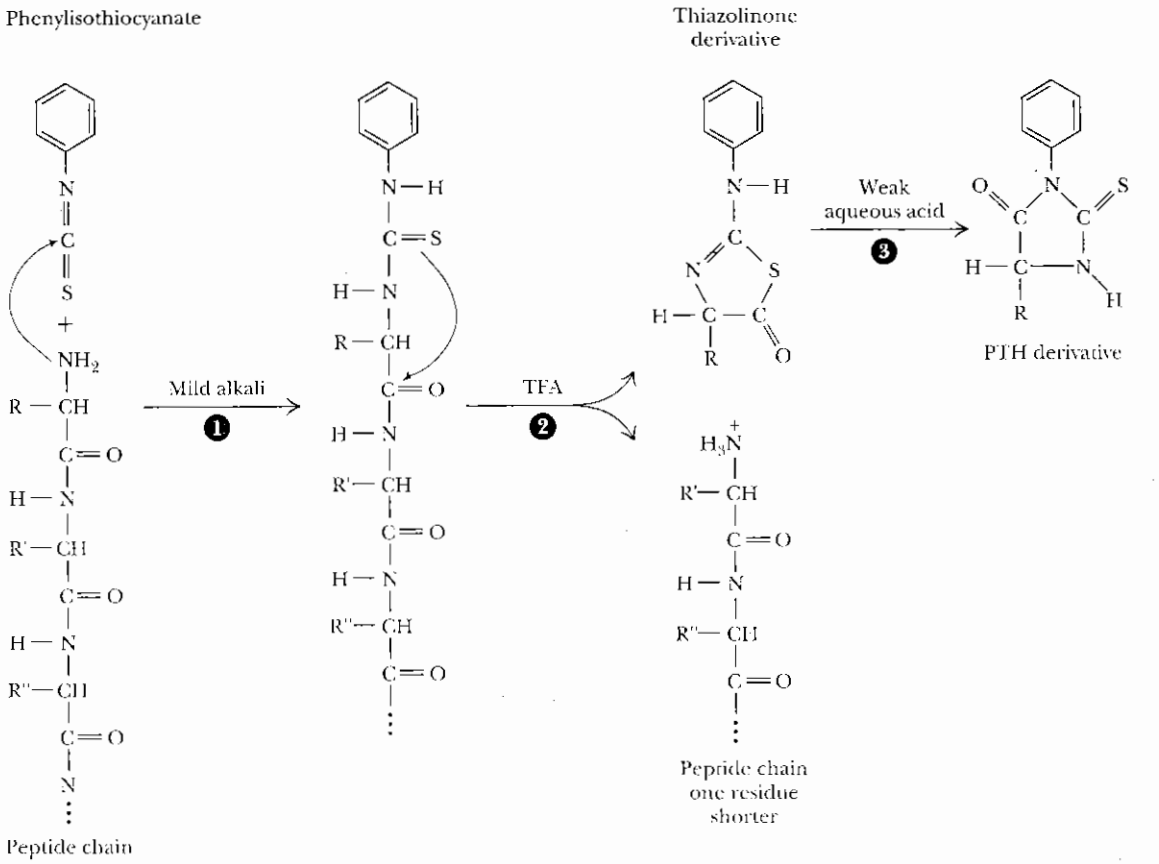


■ FIGURE 5.18 **Cleavage of proteins at internal methionine residues by cyanogen bromide.**

Chymotrypsin	H_3^+N—Leu—Asn—Asp—Phe
Cyanogen bromide	H_3^+N—Leu—Asn—Asp—Phe—His—Met
Chymotrypsin	His—Met—Thr—Met—Ala—Trp
Cyanogen bromide	Thr—Met
Cyanogen bromide	Ala—Trp—Val—Lys—COO^-
Chymotrypsin	Val—Lys—COO^-
Overall sequence	H_3^+N—Leu—Asn—Asp—Phe—His—Met—Thr—Met—Ala—Trp—Val—Lys—COO^-

■ FIGURE 5.19 **Use of overlapping sequences to determine protein sequence.** Partial digestion was effected using chymotrypsin and cyanogen bromide. For clarity, only the original N-terminus and C-terminus of the complete peptide are shown.

The alignment of like sequences on different peptides makes deducing the overall sequence possible. The Edman method has become so efficient that it is no longer considered necessary to identify the N-terminal and C-terminal ends of a protein by chemical or enzymatic methods. While interpreting results, however, it is necessary to keep in mind that a protein may consist of more than one polypeptide chain.



■ FIGURE 5.20 **Sequencing of peptides by the Edman method.** (1) Phenylisothiocyanate combines with the N-terminus of a peptide under mildly alkaline conditions to form a phenylthiocarbamoyl substitution. (2) Upon treatment with TFA (trifluoroacetic acid), this cyclizes to release the N-terminal amino acid residue as a thiazolinone derivative, but the other peptide bonds are not hydrolyzed. (3) Organic extraction and treatment with aqueous acid yield the N-terminal amino acid as a phenylthiohydantoin (PTH) derivative. The process is repeated with the remainder of the peptide chain to determine the N-terminus exposed at each stage until the entire peptide is sequenced.

In the sequencing of a peptide, the Edman reagent, *phenyl isothiocyanate,* reacts with the peptide's N-terminal residue. The modified amino acid can be cleaved off, *leaving the rest of the peptide intact,* and can be detected as the phenylthiohydantoin derivative of the amino acid. The second amino acid of the original peptide can then be treated in the same way, as can the third. With an automated instrument called a **sequencer** (Figure 5.20), the process is repeated until the whole peptide is sequenced.

Another sequencing method uses the fact that the amino acid sequence of a protein reflects the base sequence of the DNA in the gene that coded for that protein. Using currently available methods, it is sometimes easier to obtain the sequence of the DNA than that of the protein. (See Section 13.11 for a discussion of sequencing methods for nucleic acids.) Using the genetic code (Section 12.2), one can immediately determine the amino acid sequence of the protein. Convenient though this method may be, it does not determine the positions of disulfide bonds or detect amino acids, such as hydroxyproline, that are modified after translation, nor does it take into account the extensive processing that occurs with eukaryotic genomes before the final protein is synthesized (Chapters 11 and 12).

Apply Your Knowledge

Peptide Sequencing

A solution of a peptide of unknown sequence was divided into two samples. One sample was treated with trypsin, and the other was treated with chymotrypsin. The smaller peptides obtained by trypsin treatment had the following sequences:

Leu—Ser—Tyr—Ala—Ile—Arg
LSYAIR

and

Asp—Gly—Met—Phe—Val—Lys
DGMFVK

The smaller peptides obtained by chymotrypsin treatment had the following sequences:

Val—Lys—Leu—Ser—Tyr
VKLSY

Ala—Ile—Arg
AIR

and

Asp—Gly—Met—Phe
DGMF

Deduce the sequence of the original peptide.

Solution:

The key point here is that the fragments produced by treatment with the two different enzymes have overlapping sequences. These overlapping sequences can be compared to give the complete sequence. The results of the trypsin treatment indicate that there are two basic amino acids in the peptide, arginine and lysine. One of them must be the C-terminal amino acid, because no fragment was generated with a C-terminal amino acid other than these two. If there had been an amino acid other than a basic residue at the C-terminal position, trypsin treatment alone would have provided the sequence. Treatment with chymotrypsin gives the information needed. The sequence of the peptide Val—Lys—Leu—Ser—Tyr (VKLSY) indicates that lysine is an internal residue. The complete sequence is Asp—Gly—Met—Phe—Val—Lys—Leu—Ser—Tyr—Ala—Ile—Arg (DGMFVKLSYAIR).

To finish this section, let's go back to why we needed to cut the protein into pieces. Because the amino acid analyzer is giving us the sequence, it is easy to think that we could analyze a 100-amino-acid protein in one step with the analyzer and get the sequence without having to digest the protein with trypsin, chymotrypsin, or other chemicals. However, we must consider the logistical reality of doing the Edman degradation. As shown in step 1 of Figure 5.20, we react the peptide with the Edman reagent, phenylisothiocyanate (PITC). The stoichiometry of this reaction is that one molecule of the peptide reacts with one molecule of PITC. This yields one molecule of the PTH derivative

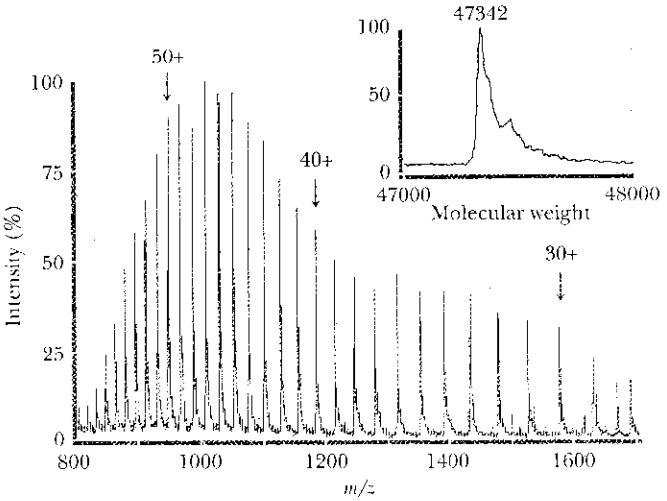
Biochemical Connections INSTRUMENTATION

The Power of Mass Spectrometry

While there are many techniques that allow a subtle approach to determining protein content and structure, none of them has the raw power of **mass spectrometry (MS).** A mass spectrometer exploits the difference in the mass-to-charge ratio (m/z) of ionized atoms or molecules in order to separate them from each other. The m/z ratio is such a characteristic property that it can be used to get structural and chemical information about the molecules and identify them.

When the charged particles are separated on the basis of their m/z ratio, they arrive at the detector at different times. The original application of MS in the early 20th century led to the discovery of isotopes. The isotopes of the noble gas argon were detected using apparatus that would seem quite simple now. For many years, the detection methods were based on having the substance to be analyzed in gaseous form or one that was easy to volatilize. As time went on, MS methods were developed that allow molecules as large as proteins, which are usually thought of as nonvolatile, to be analyzed.

One common type of MS is **electrospray Ionization (ESI-MS).** A solution of macromolecules is sprayed in the form of droplets from a capillary under a strong electric field. The droplets pick up positive charges as they leave the capillary. Evaporation of the solvent leaves multiply charged molecules. A typical 20 k-Da protein will pick up 10 to 30 positive charges. The MS spectrum of this protein reveals all of the differently charged species as a series of sharp peaks whose consecutive m/z values differ by the charge and mass of a single proton, as shown to the right: Decreasing m/z values indicate an increasing number of charges per molecule. **Tandem Mass Spectrometry** uses another spectrometer downstream from the ESI source that can analyze complex protein mixtures, such as tryptic digests or proteins emerging from an HPLC column.



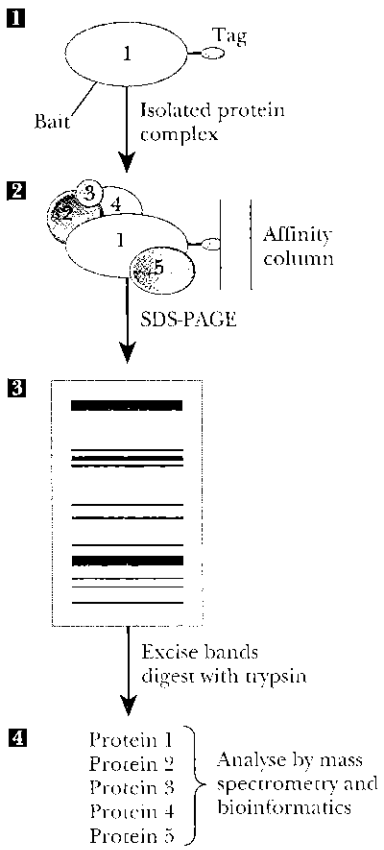
Another type of MS is called **Matrix-Assisted Laser Desorption Ionization-Time of Flight (MALDI-TOF MS).** A protein sample is mixed with a chemical matrix that includes a light-absorbing substance. A laser pulse is used to excite the chemical matrix, creating a microplasma that transfers the energy to protein molecules in the sample, ionizing them and ejecting them into the gas phase. Among the products are protein molecules that have picked up a single proton. These positively charged species can be selected by the MS for mass analysis. MALDI-TOF MS is very sensitive and very accurate. Attomole (10^{-18}) quantities of a molecule can be detected.

in step 3 that is then analyzed. Unfortunately, it is very difficult to get an exact stoichiometric match. For example, let's say we are analyzing a peptide with the sequence Asp—Leu—Tyr, etc. For simplicity, assume we add 100 molecules of the peptide to 98 molecules of the PITC because we cannot measure the quantities perfectly accurately. What happens then? In step 1, the PITC is limiting, so we eventually end up with 98 PTH derivatives of aspartate, which are analyzed correctly, and we know the N-terminus is aspartate. In the second round of the reaction, we add more PITC, but now there are two peptides; 98 of them begin with leucine and 2 of them begin with aspartate. When we analyze the PTH derivatives of round 2, we get two signals, one saying the derivative is leucine and the other saying aspartate. In round 2, the small amount of PTH derivative of aspartate does not interfere with our ability to recognize the true second amino acid. However, with every round, this situation gets worse and worse as more of the by-products show up. At some point, we get an analysis of the PTH derivatives that cannot be identified. For this reason, we have to start with smaller fragments so that we can analyze their sequences before the signal degrades.

Biochemical Connections PROTEOMICS

Pulling It All Together

The techniques introduced in this chapter are the backbone of the modern biologically based sciences, and they will be seen often throughout this book along with the information gleaned from their use. This could not be any more true than for the current trend known as **proteomics.** Proteomics is the systematic analysis of an organism's complete complement of proteins, its **proteome,** and it is one of the fastest-growing fields. Kumar et al. describe an elegant system involving three of the techniques we have seen to determine interactions between proteins in a cellular system. They created proteins they called "the bait," shown as protein 1 in the figure. These were tagged with an affinity label and allowed to react with the other cell components. The tagged bait proteins were then allowed to bind to an affinity column. In binding to the column, they took any other bound proteins with them. The bound complex was eluted from the column, then purified with SDS–PAGE. The bands were excised and digested with trypsin. After digestion, the pieces were identified with mass spectrometry. In this way the identities of the proteins associated with the bait protein were established. In the course of this book you will see many examples of interactions of proteins. This example demonstrates one of the ways that such information is gathered.



Analyzing protein interactions. In the method shown, an affinity tag is first attached to a target protein (the bait) in step (1) and allowed to react with other cell proteins. In step (2) the bait protein is bound to an affinity column. Any proteins that were interacting with the bait protein bind as well. In step (3), these proteins are purified using SDS–PAGE. In step (4), the proteins are excised from the gel and digested with trypsin, and the pieces are identified by mass spectrometry. *Redrawn by permission from Kumar, Anuj & Snyder, Michael, Proteomics: Protein complexes take the bait.* Nature *415, fig 1, p123–124 (10 Jan 2002).*

SUMMARY

How do we get the proteins out of the cells? Disruption of cells is the first step in protein purification. The various parts of cells can be separated by centrifugation. This is a useful step because proteins tend to occur in given organelles. High salt concentrations precipitate groups of proteins, which are then further separated by chromatography and electrophoresis.

What are the different types of chromatography? Gel-filtration chromatography separates proteins based on size. Ion-exchange chromatography separates proteins based on net charge. Affinity chromatography separates proteins based on their affinity for specific ligands. To purify a protein, many techniques are used and often several different chromatography steps are used.

What is the difference between agarose gels and polyacrylamide gels? Agarose gel electrophoresis is mainly used for separating nucleic acids, although it can also be used for native gel separation of proteins. Acrylamide is the usual medium for protein separation. When acrylamide gels are run with the chemical SDS, then the proteins separate based on size alone.

Why are the proteins cleaved into small fragments for protein sequencing? The Edman degradation has practical limits to how many amino acids can be cleaved from a protein and analyzed before the resulting data become confusing. To avoid this problem, the proteins are cut into small fragments using enzymes and chemicals, and these fragments are sequenced by the Edman degradation.

VIEW EXERCISES

/L Interactive versions of these problems are assignable in OWL

/hat is meant by "salting out"? How does it work?

low could you isolate mitochondria from liver cells using differen-al centrifugation?

ive an example of a scenario in which you could partially isolate a rotein with differential centrifugation using only one spin.

you had a protein X, which is a soluble enzyme found inside the eroxisome, and you wished to separate it from a similar protein Y, hich is an enzyme found embedded in the mitochondrial mem-rane, what would be your initial techniques for isolating those roteins?

/hat types of homogenization techniques are available for solubi-zing a protein?

escribe a procedure for isolating a protein that is strongly embed-ed in the mitochondrial membrane.

/hat differences between proteins are responsible for their differ-ntial solubility in ammonium sulfate?

an you separate mitochondria from peroxisomes using only dif-rential centrifugation?

hen would you choose to use a Potter–Elvehjem homogenizer in-ead of a blender?

ou are purifying a protein for the first time. You have solubilized with homogenization in a blender followed by differential cen-ifugation. You wish to try ammonium sulfate precipitation as the ext step. Knowing nothing beforehand about the amount of am-onium sulfate to add, design an experiment to find the proper oncentration (% saturation) of ammonium sulfate to use.

/hy do most people elute bound proteins from an ion-exchange olumn by raising the salt concentration instead of changing ne pH?

/hat is the basis for the separation of proteins by the following chniques?

(a) gel-filtration chromatography

(b) affinity chromatography

(c) ion-exchange chromatography

(d) reverse phase HPLC

low can gel-filtration chromatography be used to arrive at an esti-ate of the molecular weight of a protein?

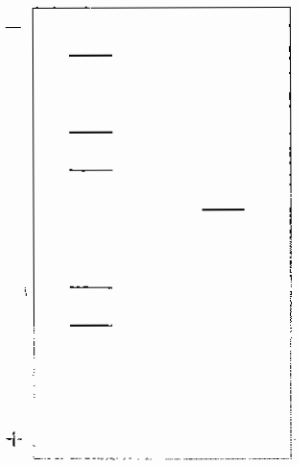
/hat is the order of elution of proteins on a gel-filtration column? /hy is this so?

n amino acid mixture consisting of lysine, leucine, and glutamic cid is to be separated by ion-exchange chromatography, using a ation-exchange resin at pH 3.5, with the eluting buffer at the same H. Which of these amino acids will be eluted from the column irst? Will any other treatment be needed to elute one of these mino acids from the column?

/hat are two ways that a compound can be eluted from an affinity olumn? What could be the advantages or disadvantages of each?

/hat are two ways that a compound can be eluted from an ion-xchange column? What could be the advantages or disadvantages f each?

18. Gel-filtration chromatography is a useful method for removing salts, such as ammonium sulfate, from protein solutions. Describe how such a separation is accomplished.
19. Sephadex G-75 has an exclusion limit of 80,000 molecular weight for globular proteins. If you tried to use this column material to separate alcohol dehydrogenase (MW 150,000) from β-amylase (MW 200,000), what would happen?
20. Referring to Question 19, could you separate β-amylase from bovine serum albumin (MW 66,000) using this column?
21. What is the main difference between reverse phase HPLC and standard ion-exchange or gel filtration chromatography?
22. What are two types of compounds that make up the resin for column chromatography?
23. You wish to separate and purify enzyme A from contaminating enzymes B and C. Enzyme A is found in the matrix of the mitochondria. Enzyme B is embedded in the mitochondrial membrane, and enzyme C is found in the peroxisome. Enzymes A and B have molecular weights of 60,000 Da. Enzyme C has a molecular weight of 100,000 Da. Enzyme A has a pI of 6.5. Enzymes B and C have pI values of 7.5. Design an experiment to separate enzyme A from the other two enzymes.
24. Referring to Question 21, how would you purify protein X using ion-exchange chromatography if it turns out the protein is only stable at a pH between 6 and 6.5?
25. Draw an example of a compound that would serve as a cation exchanger. Draw one for an anion exchanger.
26. How does HPLC differ from ion-exchange chromatography?
27. Design an experiment to purify protein X on an anion-exchange column. Protein X has an isoelectric point of 7.0.
28. An amino acid mixture consisting of phenylalanine, glycine, and glutamic acid is to be separated by HPLC. The stationary phase is aqueous and the mobile phase is a solvent less polar than water. Which of these amino acids will move the fastest? Which one will move the slowest?
29. In reverse-phase HPLC, the stationary phase is nonpolar and the mobile phase is a polar solvent at neutral pH. Which of the three amino acids in Question 28 will move fastest on a reverse-phase HPLC column? Which one will move the slowest?
30. What could be an advantage of using an anion exchange column based on a quaternary amine [i.e., resin–$N^+(CH_2CH_3)_3$] as opposed to a tertiary amine [resin–$NH^+(CH_2CH_3)_2$]?
31. What does SDS–PAGE stand for? What is the benefit of doing SDS–PAGE?
32. If you had a mixture of proteins with different sizes, shapes, and charges and you separated them with electrophoresis, which proteins would move fastest toward the anode (positive electrode)?
33. What types of macromolecules are usually separated on agarose electrophoresis gels?
34. The accompanying figure is from an electrophoresis experiment using SDS–PAGE. The left lane has the following standards: bovine serum albumin (MW 66,000), ovalbumin (MW 45,000), glyceraldehyde 3-phosphate dehydrogenase (MW 36,000), carbonic anhydrase (MW 24,000), and trypsinogen (MW 20,000). The right lane is an unknown. Calculate the MW of the unknown.



35. Why is the order of separation based on size opposite for gel filtration and gel electrophoresis, even though they often use the same compound to form the matrix?
36. What physical parameters of a protein control its migration on electrophoresis?
37. What types of compounds make up the gels used in electrophoresis?
38. How does the addition of sodium dodecylsulfate to proteins affect the basis of separation on electrophoresis?
39. Of the two principal polymers used in column chromatography and electrophoresis, which one would be most immune to contamination by bacteria and other organisms?
40. Amino acid compositions can be determined by heating a protein in 6 *M* HCl and running the hydrolysate through an ion-exchange column. If you were going to do an amino acid sequencing experiment, why would you want to get an amino acid composition first?
41. Why is it no longer considered necessary to determine the N-terminal amino acid of a protein as a separate step?
42. Why can the Edman degradation not be used effectively with very long peptides? *Hint:* Think about the stoichiometry of the peptides and the Edman reagent and the percent yield of the organic reactions involving them.
43. What are the two principal types of mass spectrometry?
44. What useful information might you get if you did determine the N-terminal amino acid as a separate step?
45. Show by a series of equations (with structures) the first stage of the Edman method applied to a peptide that has leucine as its N-terminal residue.
46. What are some of the assumptions behind the logic of the experiment described in the Biochemical Connections box on page 135?
47. What is the advantage of MALDI-TOF MS?
48. What would happen during an amino acid sequencing experiment using the Edman degradation if you accidentally added twice as much Edman reagent (on a per-mole basis) as the peptide you were sequencing?
49. A sample of an unknown peptide was divided into two aliquots. One aliquot was treated with trypsin; the other was treated with cyanogen bromide. Given the following sequences (N-terminal to C-terminal) of the resulting fragments, deduce the sequence of the original peptide.

Trypsin treatment

Asn—Thr—Trp—Met—Ile—Lys

Gly—Tyr—Met—Gln—Phe

Val—Leu—Gly—Met—Ser—Arg

Cyanogen bromide treatment

Gln—Phe

Val—Leu—Gly—Met

Ile—Lys—Gly—Tyr—Met

Ser—Arg—Asn—Thr—Trp—Met

50. Assume that you are getting ready to do an amino acid sequencing experiment on a protein containing 100 amino acids, and amino acid analysis shows the following data:

Amino Acid	Number of Residues
Ala	7
Arg	23.7
Asn	5.6
Asp	4.1
Cys	4.7
Gln	4.5
Glu	2.2
Gly	3.7
His	3.7
Ile	1.1
Leu	1.7
Lys	11.4
Met	0
Phe	2.4
Pro	4.5
Ser	8.2
Thr	4.7
Trp	0
Tyr	2.0
Val	5.1

Which of the chemicals or enzymes normally used for cutting proteins into fragments would be the least useful to you?

51. Which enzymes or chemicals would you choose to use to cut the protein from Question 50? Why?
52. With which amino acid sequences would chymotrypsin be an effective reagent for sequencing the protein from Question 50? Why?
53. What is proteomics?
54. A sample of a peptide of unknown sequence was treated with trypsin; another sample of the same peptide was treated with chymotrypsin. The sequences (N-terminal to C-terminal) of the smaller peptides produced by trypsin digestion were as follows:

Met—Val—Ser—Thr—Lys

Val—Ile—Trp—Thr—Leu—Met—Ile

Leu—Phe—Asn—Glu—Ser—Arg

The sequences of the smaller peptides produced by chymotrypsin digestion were as follows:

Asn—Glu—Ser—Arg—Val—Ile—Trp

Thr—Leu—Met—Ile

Met—Val—Ser—Thr—Lys—Leu—Phe

Deduce the sequence of the original peptide.

55. You are in the process of determining the amino acid sequence of a protein and must reconcile contradictory results. In one trial, you determine a sequence with glycine as the N-terminal amino acid and asparagine as the C-terminal amino acid. In another trial, your results indicate phenylalanine as the N-terminal amino acid and alanine as the C-terminal amino acid. How do you reconcile this apparent contradiction?

56. You are in the process of determining the amino acid sequence of a peptide. After trypsin digestion followed by the Edman degradation, you see the following peptide fragments:

Leu—Gly—Arg

Gly—Ser—Phe—Tyr—Asn—His

Ser—Glu—Asp—Met—Cys—Lys

Thr—Tyr—Glu—Val—Cys—Met—His

What is abnormal concerning these results? What might have been the problem that caused it?

57. What is the purpose of the tag on the bait protein described in the Biochemical Connections box on page 135?

ANNOTATED BIBLIOGRAPHY

Explore the annotated bibliography for this chapter online at www.cengage.com/international.