SECTION
17

Experimental Biochemistry

Learning Objectives

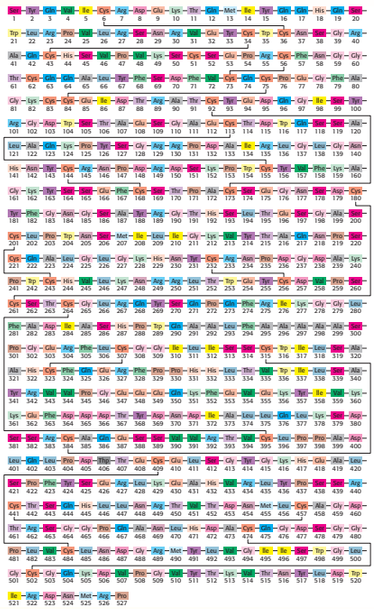
- ✓ How are proteins purified?
- ✓ How is the primary structure of a protein determined?
- ✓ What immunological techniques are used in biochemistry laboratories?
- ✓ How can a protein-encoding gene be cloned?
- ✓ How can a DNA molecule be sequenced and amplified?

You have learned much about how cells, tissues, and organisms function in your study of biochemistry. What you have learned has been presented as facts in a textbook. But everything that is now believed to be true was at some point simply an experimental observation and often a controversial one at that. For instance, recall the rejection letter that Hans Krebs received when he first submitted a paper describing the citric acid cycle (p. 290). This section describes some of the techniques used by researchers to tease information out of the cell.

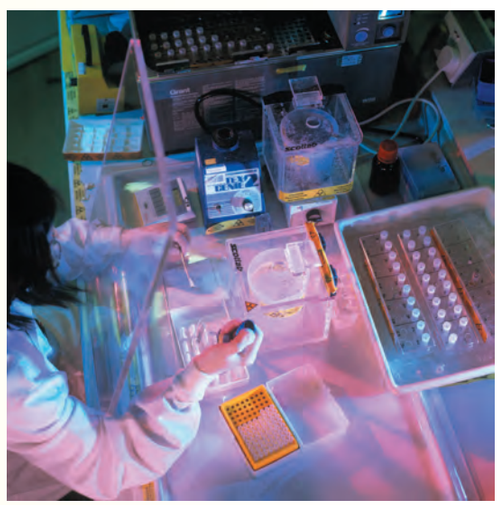
Although scientists are most interested in how biochemistry take place in an organism, this goal is technically very difficult to achieve. How can we learn about a particular biomolecule—a protein, for instance—when it is surrounded by thousands of other molecules and interacting with them? To circumvent this difficulty,

at least initially, the first approach to understanding how any biomolecule works is to isolate and purify the biomolecule and examine its biochemically properties in vitro. Toward this end, we will first examine how proteins, the workhorses of biochemistry, are purified. Protein purification is both a science and an art. Researchers take advantage of often slight differences in the physical characteristics of similar proteins to separate them from one another. After a protein has been purified, a key initial characterization is determination of its primary structure. Knowing the primary structure of a protein can be a source of insight into the structure and function of the protein and allows us to compare it with other similar proteins. Protein purification and primary structure determination are the subjects of Chapter 40.

In Chapter 41, we will examine additional techniques for the investigation of biomolecules. We will explore powerful immunological techniques that can be used to further investigate proteins as well as other biomolecules. These same techniques are also useful in clinical settings for diagnosis and treatment. Finally, we will investigate recombinant DNA technology—the tools and techniques that allow researchers to move and connect genes and large pieces of DNA. We will learn how genes are cloned and look into the variety of experimental and clinical opportunities that cloning provides.

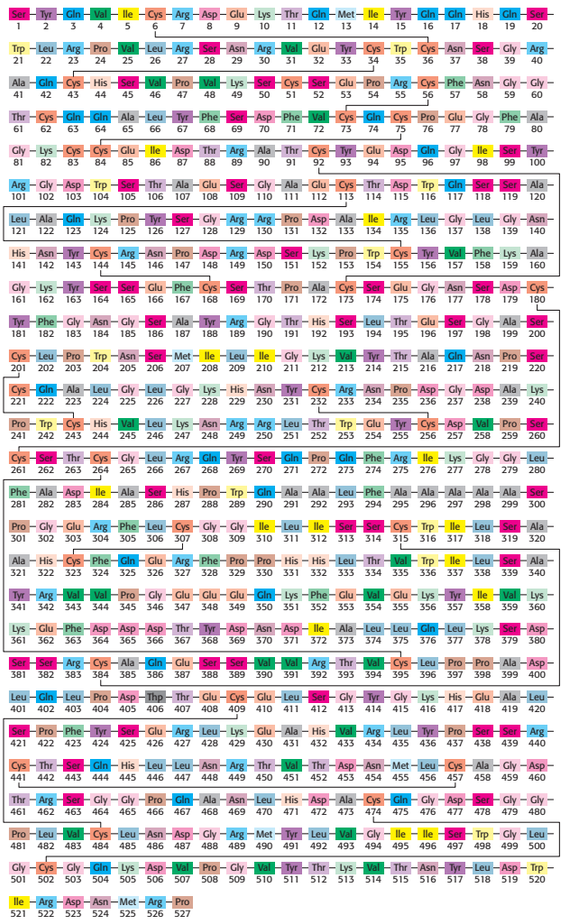


Chapter 40: Techniques in Protein Biochemistry



Chapter 41: Immunological and Recombinant DNA Techniques

CHAPTER
40 Techniques in Protein Biochemistry



The amino acid sequence of tenecteplase, a fibrinolytic for the acute treatment of myocardial infarction. [After X. Rabasseda, *Drugs Today* 37(11):749, 2001.]

Much of our study of biochemistry has focused on protein structure and function. We have observed that proteins are indeed the workhorses of the cell. All of the information that we have learned about proteins raises an interesting question: How do we know what we know about proteins? The first step toward learning how proteins work in the cell is to learn how they work outside the cell, *in vitro*. To do so, the proteins must be separated from all of the other constituents of the cell so that their biochemical properties can be identified and characterized. In other words, the protein must be purified.

In this chapter, we will examine some of the key techniques of protein purification. All of these techniques take advantage of biochemical properties unique to each protein. Then, we will learn how one crucial property of proteins—amino acid sequence, or primary structure—is elucidated.

40.1 The Proteome Is the Functional Representation of the Genome

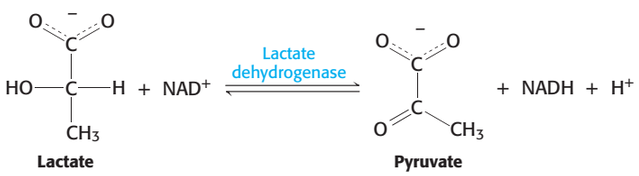
Every year, researchers are increasing their knowledge of the exact DNA base sequences and volume of information contained in the genomes of many organisms. For example, researchers recently concluded that the roundworm *Caenorhabditis elegans* has a genome of 97 million bases and about 19,000 protein-encoding genes, whereas that of the fruit fly *Drosophilia melanogaster* contains 180 million bases and about 14,000 genes. The completely sequenced human genome contains 3 billion bases and about 25,000 genes. But this genomic knowledge is analogous to a list of parts for a car: it does not explain which parts are present in different components or how the parts work together. A new word, the *proteome,* has been coined to signify a more complex level of information content—the level of *functional* information, which encompasses the types, functions, and interactions of proteins that yield a functional unit.

The term *proteome* is derived from *prote*ins expressed by the gen*ome*. The genome provides a list of gene products that *could* be present, but only a subset of these gene products will actually be expressed in a given biological context. The proteome tells us what is functionally present—for example, which proteins interact to form a signal-transduction pathway or an ion channel in a membrane. Unlike the genome, the proteome is not a fixed characteristic of the cell. Rather, because it represents the functional expression of information, it varies with cell type, developmental stage, and environmental conditions, such as the presence of hormones. Almost all gene products are proteins that can be chemically modified in a variety of ways. Furthermore, these proteins do not exist in isolation; they often interact with one another to form complexes with specific functional properties.

An understanding of the proteome is acquired by investigating, characterizing, and cataloging proteins. In some, but not all, cases, this process begins by separating a particular protein from all other biomolecules in the cell.

40.2 The Purification of Proteins Is the First Step in Understanding Their Function

To understand a protein—its amino acid sequence, its three-dimensional structure, and how it functions in normal and pathological states—we need to purify the protein. In other words, we need to isolate the protein of interest from the thousands of other proteins in the cell. This protein sample may be only a fraction of 1% of the starting material, whether that starting material consists of cells in culture or a particular organ from a plant or animal. This task is rather daunting and requires much ingenuity and patience, but, before we can even undertake the task, we need a test that identifies the protein in which we are interested. We will use this test after each stage of purification to see if the purification is working. Such a test is called an *assay,* and it is based on some unique identifying property of the protein. For enzymes, which are protein catalysts (Chapter 5), the assay is usually based on the reaction catalyzed by the enzyme in the cell. For instance, the enzyme lactate dehydrogenase, an important enzyme in glucose metabolism, carries out the following reaction:



The product, reduced nicotinamide adenine dinucleotide (NADH), in contrast with the other reaction components, absorbs light at 340 nm. Consequently, we can follow the progress of the reaction by measuring the light absorbance at 340 nm in unit time—for instance, within 1 minute after the addition of the sample that contains the enzyme. Our assay for enzyme activity during the purification of lactate dehydrogenase is thus the increase in absorbance of light at 340 nm observed in 1 minute. Note that the assay tells us how much enzyme *activity* is present, not how much enzyme protein is present.

To be certain that our purification scheme is working, we need one additional piece of information—the amount of total protein present in the mixture being assayed. This measurement of the total amount of protein includes the enzyme of interest as well as all the other proteins present, but it is *not* a measure of enzyme activity. After we know both how much enzyme activity is present and how much protein is present, we can assess the progress of our purification by measuring the *specific activity,* the ratio of enzyme activity to the amount of protein in the enzyme assay at each step of our purification. The specific activity will rise as the protein mixture used for the assay consists to a greater and greater extent of the protein of interest. In essence, the point of the purification is to remove all proteins except the protein in which we are interested. Quantitatively, it means that we want to maximize specific activity.

? QUICK QUIZ 1 Why is an assay required for protein purification?

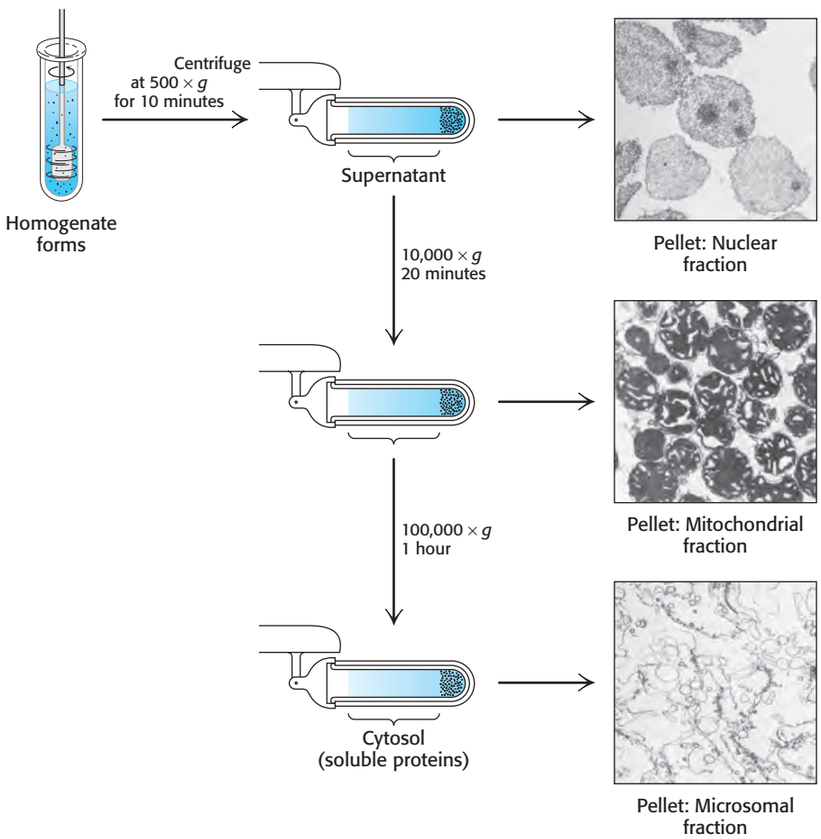


Figure 40.1 Differential centrifugation. Cells are disrupted in a homogenizer and the resulting mixture, called the homogenate, is centrifuged in a step-by-step fashion of increasing centrifugal force. The denser material will form a pellet at lower centrifugal force than will the less-dense material. The isolated fractions can be used for further purification. [Photographs courtesy of Dr. S. Fleischer and Dr. B. Fleischer.]

Proteins Must Be Removed from the Cell to Be Purified

Having found an assay, we must now break open the cells, releasing the cellular contents, so that we can gain access to our protein. The disruption of the cell membranes yields a *homogenate,* a mixture of all of the components of the cell but no intact cells. This mixture is centrifuged at low centrifugal force, yielding a pellet of heavy material at the bottom of the centrifuge tube and a lighter solution above, called supernatant (**Figure 40.1**). The pellet and supernatant are called *fractions* because we are fractionating the homogenate. The supernatant is again centrifuged at a greater force to yield yet another pellet and supernatant. The procedure, called *differential centrifugation,* yields several fractions of decreasing density, each still containing hundreds of different proteins, which are assayed for the activity being purified. Usually, one fraction will have more enzyme activity than any other fraction, and it then serves as the source of material to which more discriminating purification techniques are applied. The fraction that is used as a source for further purification is often called the crude extract.

Proteins Can Be Purified According to Solubility, Size, Charge, and Binding Affinity

Proteins are purified on the bases of differences in *solubility, size, charge,* and *specific binding affinity*. Usually, protein mixtures are subjected to a series of separations, each based on a different property.

Salting out. Most proteins require some salt to dissolve, a process called *salting in*. However, most proteins precipitate out of solution at high salt concentrations, an effect called *salting out* (**Figure 40.2**). Salting out is due to competition between the salt ions and the protein for water to keep the protein in solution (water of solvation). The salt concentration at which a protein precipitates differs from one protein to another. Hence, salting out can be used to fractionate a mixture of proteins. Unfortunately, many proteins lose their activity in the presence of such high concentrations of salt. However, the salt can be removed by the process of *dialysis*. The protein–salt solution is placed in a small bag made of a semipermeable membrane, such as a cellulose membrane, with pores (**Figure 40.3**). Proteins are too large to fit through the pores of the membrane, whereas smaller molecules and ions such as salts can escape through the pores and emerge in the medium outside the bag (the dialysate).

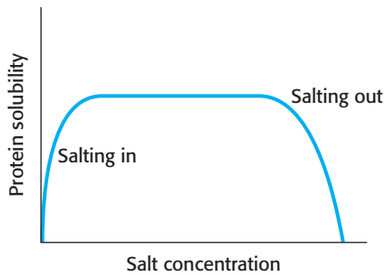


Figure 40.2 The dependency of protein solubility on salt concentration. The graph shows how altering the salt concentration affects the solubility of a hypothetical protein. Different proteins will display different curves.

Separation by size. *Gel-filtration chromatography,* also called molecular exclusion chromatography, separates proteins on the basis of size. The sample is applied to the top of a column consisting of porous beads made of an insoluble polymer such as dextran, agarose, or polyacrylamide (**Figure 40.4**). Small molecules can enter these beads, but large ones cannot, and so those larger molecules follow a shorter path to the bottom of the column and emerge first. Molecules that are of a size to occasionally enter a bead will flow from the column at an intermediate position, and small molecules, which take a longer, more circuitous path, will exit last.

Ion-exchange chromatography. Proteins can be separated on the basis of their net charge by *ion-exchange chromatography.* If a protein has a net positive charge at pH 7, it will usually bind to a column of beads containing negatively charged

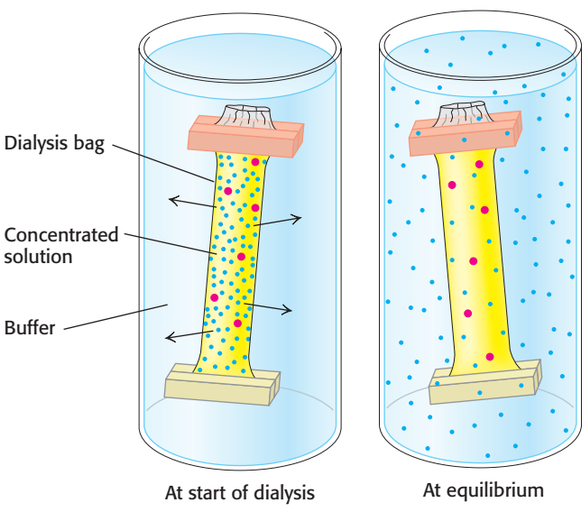


Figure 40.3 Dialysis. Protein molecules (red) are retained within the dialysis bag, whereas small molecules (blue) diffuse into the surrounding medium.

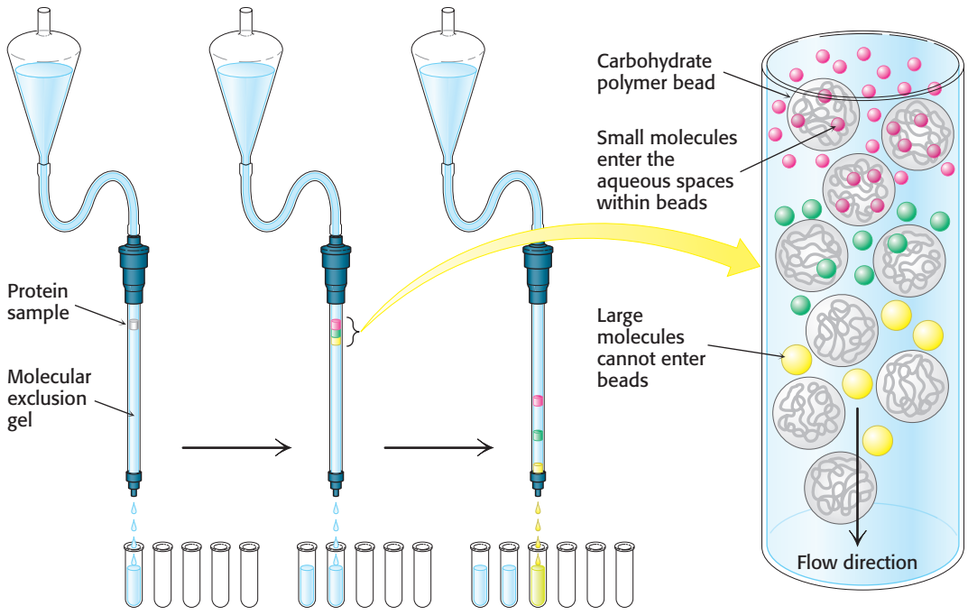


Figure 40.4 Gel-filtration chromatography. A mixture of proteins in a small volume is applied to a column filled with porous beads. Because large proteins cannot enter the internal volume of the beads, they emerge sooner than do small ones.

carboxylate groups, whereas a negatively charged protein will not bind to the column (**Figure 40.5**). A positively charged protein bound to such a column can then be released by increasing the concentration of salt in the buffer poured over the column. The positively charged ions of the salt compete with positively charged groups on the protein for binding to the column. Likewise, a protein with a net negative charge will be bound to ion-exchange beads carrying positive charges and can be eluted from the column with the use of a buffer containing salt.

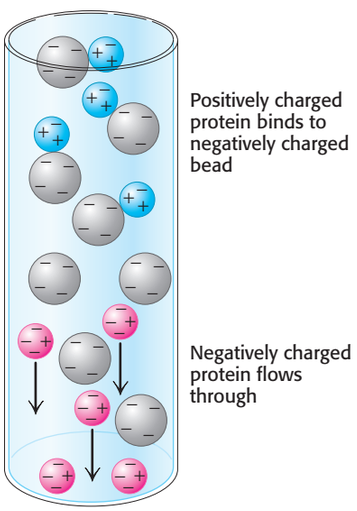


Figure 40.5 Ion-exchange chromatography. This technique separates proteins mainly according to their net charge.

Affinity chromatography. *Affinity chromatography* is another powerful means of purifying proteins. This technique takes advantage of the fact that some proteins have a high affinity for specific chemical groups or specific molecules. For example, the plant protein concanavalin A, which binds to glucose, can be purified by passing a crude extract through a column of beads containing covalently attached glucose residues. Concanavalin A binds to such beads, whereas most other proteins do not. The bound concanavalin A can then be released from the column by adding a concentrated solution of glucose. The glucose in solution displaces the column-attached glucose residues from binding sites on concanavalin A (**Figure 40.6**).

High-pressure liquid chromatography. The ability of column techniques to separate individual proteins, called the resolving power, can be improved substantially through the use of a technique called *high-pressure liquid chromatography* (HPLC), which is an enhanced version of the column techniques already discussed. The beads that make up the column material themselves are much more finely divided and, as a consequence, there are more interaction sites and thus greater resolving power. Because the column is made of finer material, pressure must be applied to the column to obtain adequate flow rates. The net result is high resolution as well as rapid separation (**Figure 40.7**).

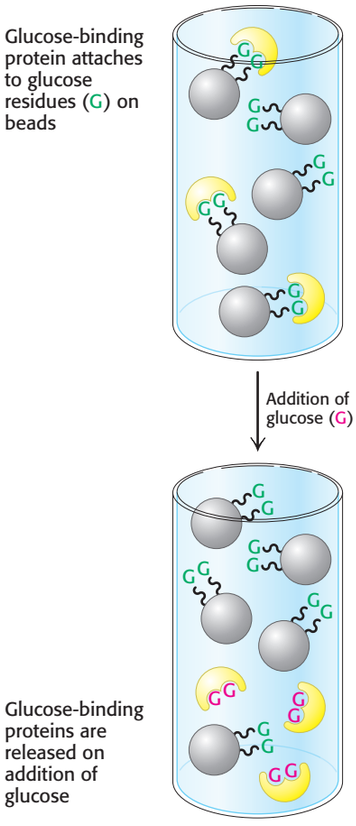


Figure 40.6 Affinity chromatography. Affinity chromatography of concanavalin A (shown in yellow) on a solid support containing covalently attached glucose residues (G).

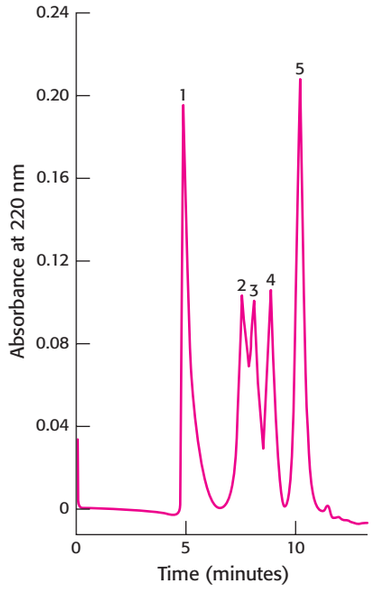


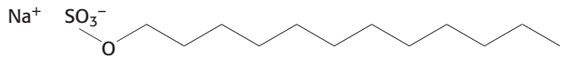
Figure 40.7 High-pressure liquid chromatography (HPLC). Gel filtration by HPLC clearly defines the individual proteins because of its greater resolving power. Proteins are detected by their absorbance of 220-nm light waves: (1) thyroglobulin (669 kd), (2) catalase (232 kd), (3) bovine serum albumin (67 kd), (4) ovalbumin (43 kd), and (5) ribonuclease (13.4 kd). [After K. J. Wilson and T. D. Schlabach. In *Current Protocols in Molecular Biology*, vol. 2, suppl. 41, F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl, Eds. (Wiley, 1998), p. 10.14.1.]

Proteins Can Be Separated by Gel Electrophoresis and Displayed

How can we tell whether a purification scheme is effective? One way is to demonstrate that the specific activity rises with each purification step. Another is to visualize the number of proteins present at each step. The technique of *gel electrophoresis* makes the latter method possible.

A molecule with a net charge will move in an electric field, a phenomenon termed *electrophoresis*. The distance and speed that a protein moves in electrophoresis depends on the electric-field strength, the net charge on the protein, which is a function of the pH of the electrophoretic solution, and the shape of the protein. Electrophoretic separations are nearly always carried out in gels, such as polyacrylamide, because the gel serves as a molecular sieve that enhances separation. Molecules that are small compared with the pores in the gel readily move through the gel, whereas molecules much larger than the pores are almost immobile. Intermediate-size molecules move through the gel with various degrees of ease. The electrophoresis of proteins is performed in a thin, vertical slab of polyacrylamide. The direction of flow is from top to bottom (**Figure 40.8**).

Proteins can be separated largely on the basis of mass by electrophoresis in a polyacrylamide gel in the presence of the detergent sodium dodecyl sulfate (SDS). The negatively charged SDS denatures proteins and binds to the denatured protein at a constant ratio of one SDS molecule for every two amino acids in the protein. The negative charges on the many SDS molecules bound to the protein "swamp" the normal charge on the protein and cause all proteins to have the same charge-to-mass ratio. Thus, proteins will differ only in their mass. Finally, a sulfhydryl agent such as mercaptoethanol is added to reduce disulfide bonds and completely linearize the proteins. The SDS–protein complexes are then subjected



Sodium dodecyl sulfate (SDS)

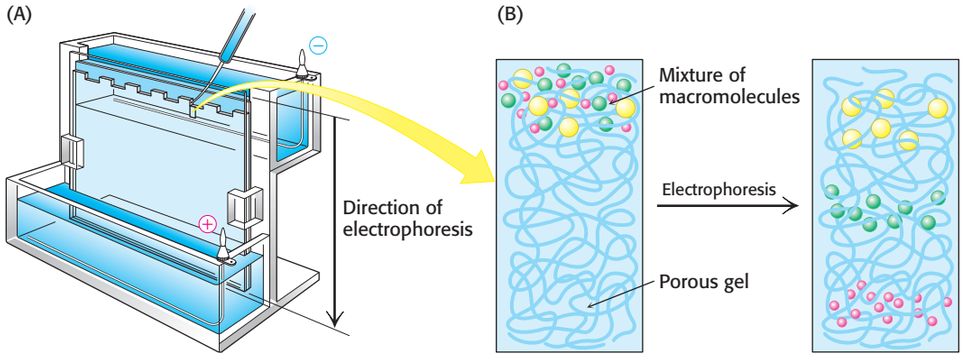


Figure 40.8 Polyacrylamide-gel electrophoresis. (A) Gel-electrophoresis apparatus. Typically, several samples undergo electrophoresis on one flat polyacrylamide gel. A microliter pipette is used to place solutions of proteins in the wells of the slab. A cover is then placed over the gel chamber and voltage is applied. The negatively charged SDS (sodium dodecyl sulfate)–protein complexes migrate in the direction of the anode, at the bottom of the gel. (B) The sieving action of a porous polyacrylamide gel separates proteins according to size, with the smallest moving most rapidly.

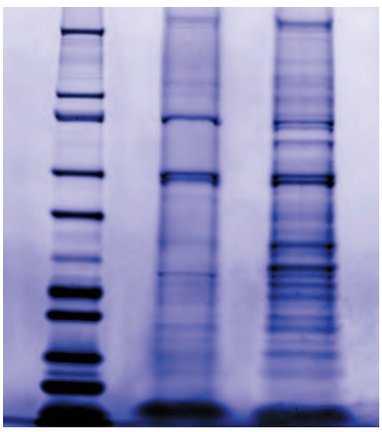


Figure 40.9 The staining of proteins after electrophoresis. Proteins subjected to electrophoresis on an SDS–polyacrylamide gel can be visualized by staining with Coomassie blue. The lane on the left is a set of marker proteins of known molecular weight. These marker proteins have been separated on the basis of size, with the smaller proteins moving farther into the gel than the larger proteins. Two different protein mixtures are in the remaining lanes. [Wellcome Photo Library.]

to electrophoresis. When the electrophoresis is complete, the proteins in the gel can be visualized by staining them with silver or a dye such as Coomassie blue, which reveals a series of bands (**Figure 40.9**). *Small proteins move rapidly through the gel, whereas large proteins stay at the top, near the point of application of the mixture.*

Isoelectric focusing. Proteins can also be separated electrophoretically on the basis of their relative contents of acidic and basic residues. The *isoelectric point* (pI) of a protein is the pH at which its net charge is zero. At this pH, the protein will not migrate in an electric field. If a mixture of proteins is subjected to electrophoresis in a pH gradient in a gel in the absence of SDS, each protein will move until it reaches a position in the gel at which the pH is equal to the pI of the protein. This method of separating proteins is called *isoelectric focusing*. Proteins differing by one net charge can be separated (**Figure 40.10**).

Two-dimensional electrophoresis. Isoelectric focusing can be combined with SDS–PAGE (SDS–polyacrylamide gel electrophoresis) to obtain very high resolution separations. A single sample is first subjected to isoelectric focusing. This single-lane gel is then placed horizontally on top of an SDS–polyacrylamide slab and subjected to electrophoresis again, in a direction perpendicular to the isoelectric focusing, to yield a two-dimensional pattern of spots. In such a gel, proteins have been separated in the horizontal direction on the basis of isoelectric point and in

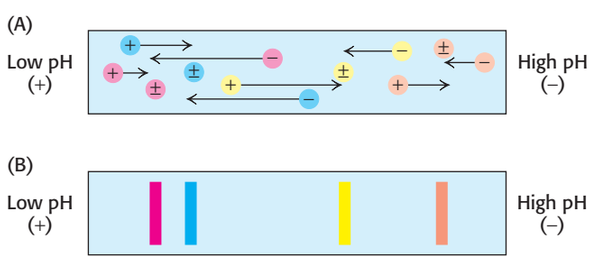


Figure 40.10 The principle of isoelectric focusing. A pH gradient is established in a gel before the sample has been loaded. (A) The sample is loaded and voltage is applied. The proteins will migrate to their isoelectric pH, the location at which they have no net charge. (B) The proteins form bands that can be excised and used for further experimentation.

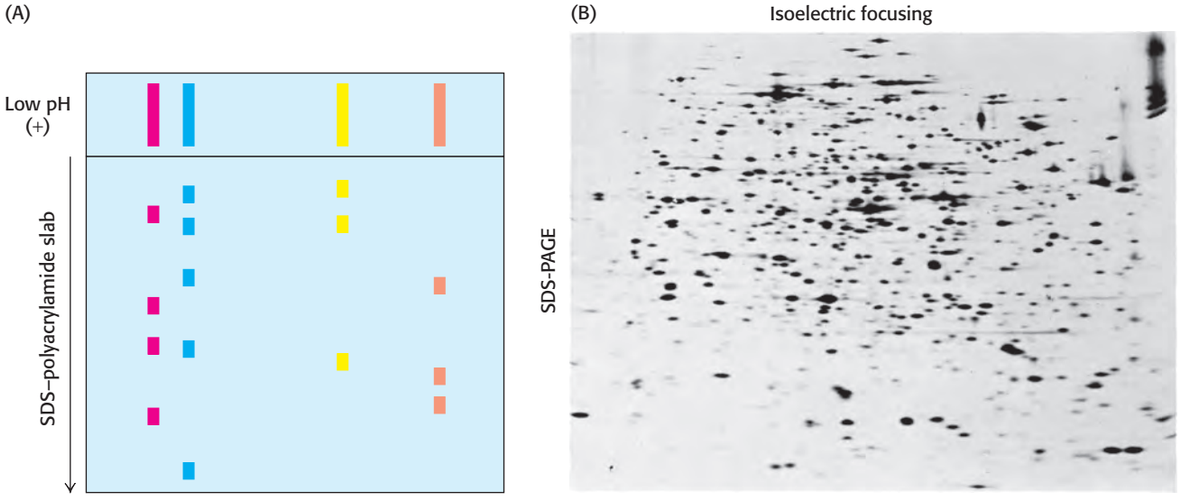


Figure 40.11 Two-dimensional gel electrophoresis. (A) A protein sample is initially fractionated in one direction by isoelectric focusing as described in Figure 40.10. The isoelectric focusing gel is then attached to an SDS–polyacrylamide gel, and electrophoresis is performed in the second direction, perpendicular to the original separation. Proteins with the same pI value are now separated on the basis of mass. (B) Proteins from *E. coli* were separated by two-dimensional gel electrophoresis, resolving more than a thousand different proteins. The proteins were first separated according to their isoelectric pH in the horizontal direction and then by their apparent mass in the vertical direction. [(B) Courtesy of Dr. Patrick H. O'Farrell.]

the vertical direction on the basis of mass. More than a thousand different proteins in the bacterium *Escherichia coli* can be resolved in a single experiment by *two-dimensional electrophoresis* (**Figure 40.11**).

A Purification Scheme Can Be Quantitatively Evaluated

Some combination of purification techniques will usually yield a pure protein. To determine the success of a protein-purification scheme, we monitor the procedure at each step by determining specific activity and by performing an SDS-PAGE analysis. Consider the results for the purification of a hypothetical protein, summarized in **Table 40.1** and **Figure 40.12**. At each step, the following parameters are measured:

- *Total Protein.* The quantity of protein present in a fraction is obtained by determining the protein concentration of a part of each fraction and multiplying by the fraction's total volume.

Table 40.1 Quantification of a purification protocol for a hypothetical protein

Step	Total protein (mg)	Total activity (units)	Specific activity (units mg^{-1})	Yield (%)	Purification level
Homogenization	15,000	150,000	10	100	1
Salt fractionation	4,600	138,000	30	92	3
Ion-exchange chromatography	1,278	115,500	90	77	9
Gel-filtration chromatography	68.8	75,000	1,100	50	110
Affinity chromatography	1.75	52,500	30,000	35	3,000

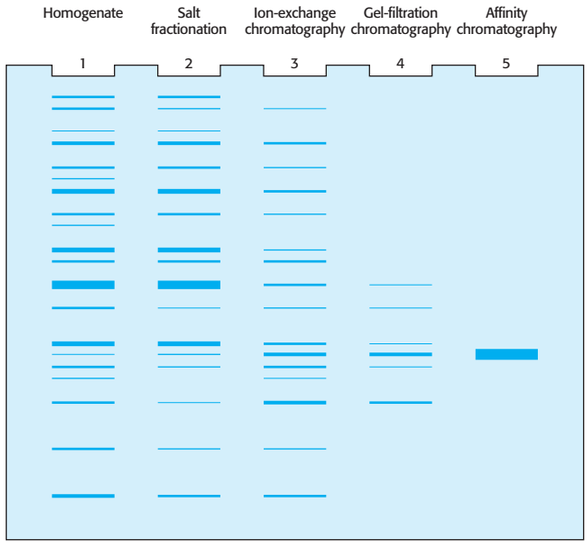


Figure 40.12 Electrophoretic analysis of a protein purification. The purification scheme in Table 40.1 was analyzed by SDS-PAGE. Each lane contained 50 μg of sample. The effectiveness of the purification can be seen as the band for the protein of interest becomes more prominent relative to other bands.

- *Total Activity.* The enzyme activity for the fraction is obtained by measuring the enzyme activity in the volume of fraction used in the assay and multiplying by the fraction's total volume.
- *Specific Activity.* This parameter, obtained by dividing total activity by total protein, enables us to measure the degree of purification by comparing specific activities after each purification step. Recall that the goal of a purification scheme is to maximize specific activity.
- *Yield.* This parameter is a measure of the total activity retained after each purification step as a percentage of the activity in the crude extract. The amount of activity in the initial extract is taken to be 100%.
- *Purification level.* This parameter is a measure of the increase in purity and is obtained by dividing the specific activity, calculated after each purification step, by the specific activity of the initial extract.

As we see in Table 40.1, several purification steps can lead to several thousand-fold purification. Inevitably, in each purification step, some of the protein of interest is lost, and so our overall yield is 35%. A good purification scheme takes into account purification levels as well as yield.

The SDS-PAGE depicted in Figure 40.12 shows that, if we load the same amount of protein onto each lane after each step, the number of bands decreases in proportion to the level of purification and the amount of protein of interest increases as a proportion of the total protein present.

? QUICK QUIZ 2 What physical differences among proteins allow for their purification?

40.3 Determining Primary Structure Facilitates an Understanding of Protein Function

An important means of characterizing a pure protein is to determine its primary structure, which can tell us much about the protein. Recall that the primary structure of a protein is the determinant of its three-dimensional structure, which ultimately determines the protein's function. Comparison of the sequence of normal

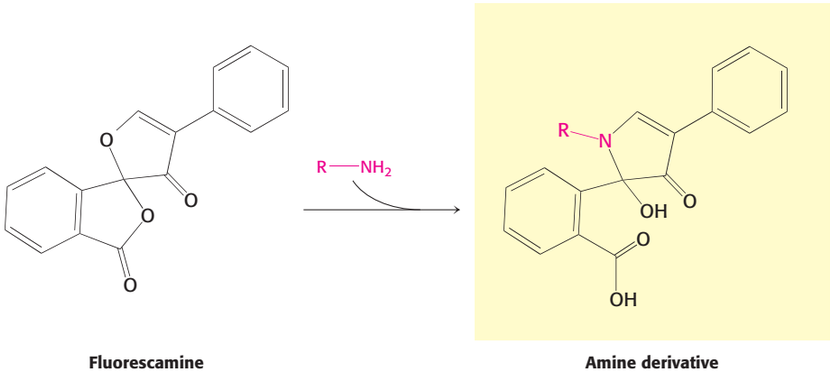


Figure 40.13 Fluorescent derivatives of amino acids. Fluorescamine reacts with the α-amino group of an amino acid to form a fluorescent derivative.

proteins with those isolated from patients with pathological conditions allows an understanding of the molecular basis of diseases.

Let us examine first how we can sequence a simple peptide, such as

Ala-Gly-Asp-Phe-Arg-Gly

The first step is to determine the *amino acid composition* of the peptide. The peptide is hydrolyzed into its constituent amino acids by heating it in strong acid. The individual amino acids can then be separated by ion-exchange chromatography and visualized by treatment with *fluorescamine,* which reacts with the α-amino group to form a highly fluorescent product (**Figure 40.13**).

The concentration of an amino acid in solution is proportional to the fluorescence of the solution. The solution is then run through a column. The amount of buffer required to remove the amino acid from the column is compared with the elution pattern of a standard mixture of amino acids, revealing the identity of the amino acid in the solution (**Figure 40.14**). The composition of our peptide is

(Ala, Arg, Asp, Gly_2, Phe)

The parentheses denote that this is the amino acid composition of the peptide, not its sequence.

The sequence of a protein can then be determined by a process called the Edman degradation. The *Edman degradation* sequentially removes one residue at

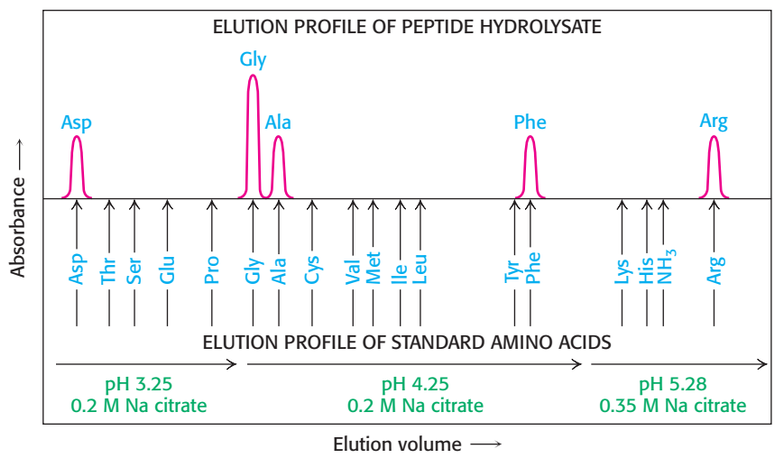


Figure 40.14 Determination of amino acid composition. Different amino acids in a peptide hydrolysate can be separated by ion-exchange chromatography on a sulfonated polystyrene resin (such as Dowex-50). Buffers (in this case, sodium citrate) of increasing pH are used to elute the amino acids from the column. The amount of each amino acid present is determined from the absorbance. Aspartate, which has an acidic side chain, is the first to emerge, whereas arginine, which has a basic side chain, is the last. The original peptide is revealed to be composed of one aspartate, one alanine, one phenylalanine, one arginine, and two glycine residues.

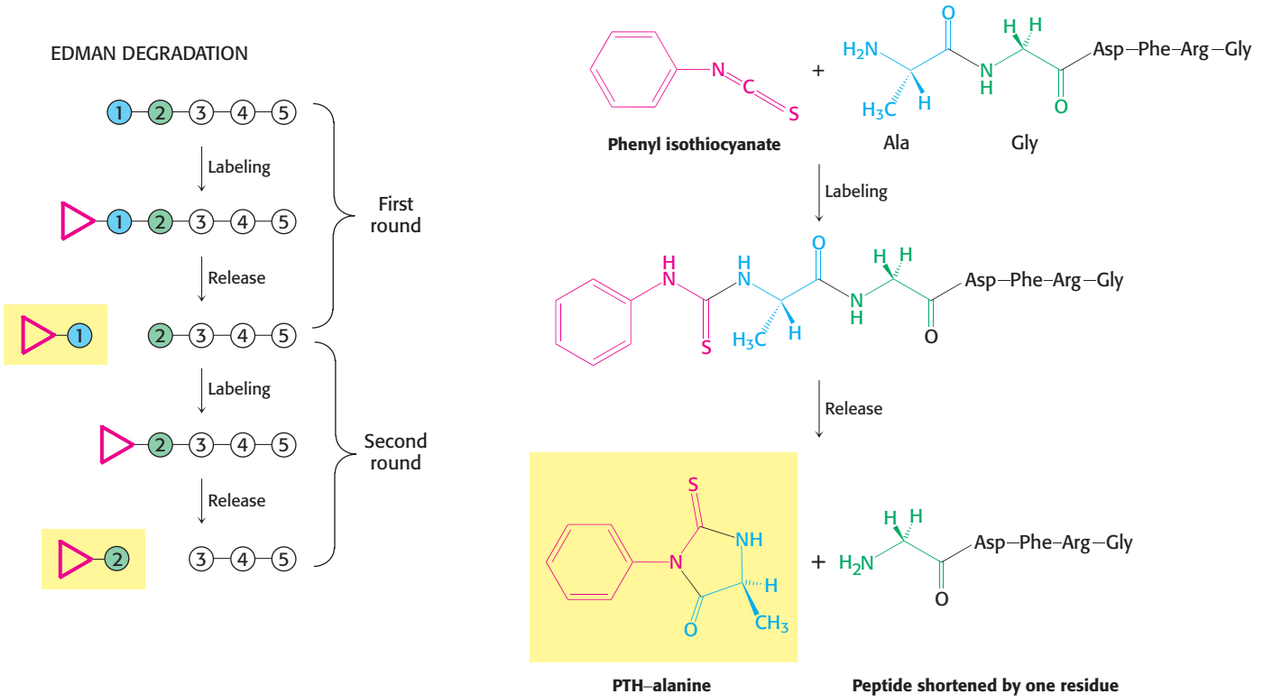


Figure 40.15 The Edman degradation. The labeled amino-terminal residue (PTH–alanine in the first round) can be released without hydrolyzing the rest of the peptide. Hence, the amino-terminal residue of the shortened peptide (Gly-Asp-Phe-Arg-Gly) can be determined in the second round. Three more rounds of the Edman degradation reveal the complete sequence of the original peptide.

a time from the amino end of a peptide (**Figure 40.15**). *Phenyl isothiocyanate* reacts with the terminal amino group of the peptide, which then cyclizes and breaks off the peptide, yielding an intact peptide shortened by one amino acid. The cyclic compound is a phenylthiohydantoin (PTH)–amino acid, which can be identified by chromatographical procedures. The Edman procedure can then be repeated sequentially to yield the amino acid sequence of the peptide.

Table 40.2 Specific cleavage of polypeptides

Reagent	Cleavage site
Chemical cleavage	
Cyanogen bromide	Carboxyl side of methionine residues
O-Iodosobenzoate	Carboxyl side of tryptophan residues
Hydroxylamine	Asparagine–glycine bonds
2-Nitro-5-thiocyanobenzoate	Amino side of cysteine residues
Enzymatic cleavage	
Trypsin	Carboxyl side of lysine and arginine residues
Clostripain	Carboxyl side of arginine residues
Staphylococcal protease	Carboxyl side of aspartate and glutamate residues (glutamate only under certain conditions)
Thrombin	Carboxyl side of arginine
Chymotrypsin	Carboxyl side of tyrosine, tryptophan, phenylalanine, leucine, and methionine
Carboxypeptidase A	Amino side of C-terminal amino acid (not arginine, lysine, or proline)

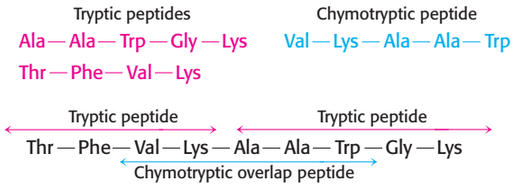


Figure 40.16 Overlap peptides. The peptide obtained by chymotryptic digestion overlaps two tryptic peptides, establishing their order.

In principle, we should be able to sequence an entire protein by using the Edman method. In practice, the peptides cannot be much longer than about 50 residues, because the reactions of the Edman method are not 100% efficient and, eventually, the sequencing reactions are out of order. We can circumvent this obstacle by cleaving the original protein at specific amino acids into smaller peptides that can be sequenced independently. In essence, the strategy is to *divide and conquer*.

Specific cleavage can be achieved by chemical or enzymatic methods. **Table 40.2** gives several ways of specifically cleaving polypeptide chains. The peptides obtained by specific chemical or enzymatic cleavage are separated, and the sequence of each purified peptide is then determined by the Edman method. At this point, the amino acid sequences of segments of the protein are known, but the order of these segments is not yet defined. How can we order the peptides to obtain the primary structure of the original protein? The necessary additional information is obtained from *overlap peptides* (**Figure 40.16**). A second cleavage technique is used to split the polypeptide chain at different sites. Some of the peptides from the second cleavage will overlap two or more peptides from the first cleavage, and they can be used to establish the order of the peptides. The entire amino acid sequence of the polypeptide chain is then known.

? QUICK QUIZ 3 Differentiate between amino acid composition and amino acid sequence.

Amino Acid Sequences Are Sources of Many Kinds of Insight

A protein's amino acid sequence is a valuable source of insight into the protein's function, structure, and history.

1. *The sequence of a protein of interest can be compared with all other known sequences to ascertain whether significant similarities exist. Does this protein belong to an established family?* A search for kinship between a newly sequenced protein and the millions of previously sequenced ones takes only a few seconds on a personal computer. If the newly isolated protein is a member of an established class of protein, we can begin to infer information about the protein's structure and function. For instance, chymotrypsin and trypsin are members of the serine protease family, a clan of proteolytic enzymes that have a common catalytic mechanism based on a reactive serine residue. If the sequence of the newly isolated protein shows sequence similarity with trypsin or chymotrypsin, the result suggests that it, too, may be a serine protease.

2. *Comparison of sequences of the same protein in different species yields a wealth of information about evolutionary pathways.* Genealogical relations between species can be inferred from sequence differences between their proteins. We can even estimate the time at which two evolutionary lines diverged, thanks to the clocklike nature of random mutations. For example, a comparison of serum albumins found in primates indicates that human beings and African apes diverged 5 million years ago, not 30 million years ago as was once thought. Sequence analyses have opened a new perspective on the fossil record and the pathway of human evolution.

3. *Amino acid sequences can be searched for the presence of internal repeats.* Such internal repeats can reveal the history of an individual protein itself. Many

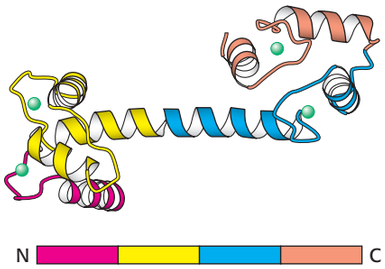


Figure 40.17 Repeating motifs in a protein chain. Calmodulin, a calcium sensor, contains four similar units (shown in red, yellow, blue, and orange) in a single polypeptide chain. *Notice* that each unit binds a calcium ion (shown in green). [Drawn from 1CLL.pdb.]

proteins apparently have arisen by the duplication of primordial genes. For example, calmodulin, a ubiquitous calcium sensor in eukaryotes, contains four similar calcium-binding modules that arose by gene duplication (**Figure 40.17**).

4. *Many proteins contain amino acid sequences that serve as signals designating their destinations or controlling their processing.* For example, a protein destined for export from a cell or for location in a membrane contains a *signal sequence,* a stretch of about 20 hydrophobic residues near the amino terminus that directs the protein to the appropriate membrane. Another protein may contain a stretch of amino acids that functions as a *nuclear localization signal,* directing the protein to the nucleus.

5. *Sequence data allow a molecular understanding of diseases.* Many diseases are caused by mutations in DNA that result in alterations in the amino acid sequence of a particular protein. These alterations often compromise the protein's function. For instance, sickle-cell anemia is caused by a change in a single amino acid in the primary structure of the β chain of hemoglobin. Approximately 70% of the cases of cystic fibrosis are caused by the deletion of one particular amino acid from the 1480-amino-acid-containing protein that controls chloride transport across cell membranes. Indeed, a major goal of biochemistry is to elucidate the molecular basis of disease with the hope that this understanding will lead to effective treatment.

Clinical Insight

Understanding Disease at the Molecular Level: Sickle-Cell Anemia Results from a Single Amino Acid Change

Studies with mutations of hemoglobin have provided many examples showing that alterations in the primary structure of a protein result in pathological conditions. The role of hemoglobin, which is contained in red blood cells, is to bind oxygen in the lungs and to transport and release oxygen to tissues that require oxygen for combusting fuels. If the ability of hemoglobin to carry and release oxygen is somehow compromised, anemia results. Anemia is characterized by a host of symptoms, most commonly fatigue. A well-studied example of anemia is sickle-cell anemia, which is most commonly found in people from sub-Saharan Africa or their descendants (p. 113). Recall from Chapter 8 that the defining feature of sickle-cell anemia is that the red blood cells adopt a sickle shape after the hemoglobin has released its bound oxygen. These sickled cells clog small capillaries and impair blood flow. The results may be painful swelling of the extremities and a higher risk of stroke or bacterial infection (owing to poor circulation). The sickled red cells also do not remain in circulation as long as normal cells do, leading to anemia.

What is the molecular defect associated with sickle-cell anemia? A single amino acid substitution in the β chain of hemoglobin is responsible—namely, the substitution of a valine residue for a glutamate residue in position 6. The mutated form is referred to as *hemoglobin S* (Hb S). In people with sickle-cell anemia, both copies of the hemoglobin β-chain (Hb B) gene are mutated. The Hb S substitution substantially decreases the solubility of deoxyhemoglobin, although it does not markedly alter the properties of oxyhemoglobin. Hence, sickling takes place in the small capillaries after the hemoglobin has released its oxygen to the tissues. The deoxyhemoglobin molecules associate with one another and form insoluble aggregates that deform the cell, leading to the characteristic sickle shape (**Figure 40.18**). Thus, sickle-cell anemia, like the spongiform encephalopathies discussed earlier (Chapter 4), is a pathological condition resulting from protein aggregation. ■

Figure 40.18 Sickled red blood cells trapped in capillaries. The micrograph shows sickled red blood cell trapped in tiny blood vessels called capillaries. The trapped cells impede the flow of blood through the tissues. Consequently, the cells are deprived of oxygen and the tissues are damaged. [Courtesy of National Heart, Lung and Blood Institute.]

SUMMARY

40.1 The Proteome Is the Functional Representation of the Genome

The rapid progress in gene sequencing has advanced another goal of biochemistry—the elucidation of the proteome. The proteome is the complete set of proteins expressed and includes information about how they are modified, how they function, and how they interact with other molecules. Unlike the genome, the proteome is not static and varies with cell type, developmental stage, and environmental conditions.

40.2 The Purification of Proteins Is the First Step in Understanding Their Function

Proteins can be separated from one another and from other molecules on the basis of such characteristics as solubility, size, charge, and binding affinity. SDS-PAGE separates the polypeptide chains of proteins under denaturing conditions largely according to mass. Proteins can also be separated electrophoretically on the basis of net charge by isoelectric focusing in a pH gradient.

40.3 Determining Primary Structure Facilitates an Understanding of Protein Function

The amino acid composition of a protein can be ascertained by hydrolyzing the protein into its constituent amino acids. The amino acids can be separated by ion-exchange chromatography and quantitated by their reaction with fluorescamine. Amino acid sequences can be determined by Edman degradation, which removes one amino acid at a time from the amino end of a peptide. Phenyl isothiocyanate reacts with the terminal amino group to form a phenylthiohydantoin–amino acid and a peptide shortened by one residue. Longer polypeptide chains are broken into shorter ones for analysis by specifically cleaving them with a reagent that breaks the peptide at specific sites. Amino acid sequences are rich in information concerning the kinship of proteins, their evolutionary relations, and diseases produced by mutations. Knowledge of a sequence provides valuable clues to conformation and function.

Key Terms

proteome (p. 623)
assay (p. 623)
homogenate (p. 625)
salting in (p. 625)
salting out (p. 625)
dialysis (p. 625)
gel-filtration chromatography (p. 625)
ion-exchange chromatography (p. 625)
affinity chromatography (p. 626)
high-pressure liquid chromatography (HPLC) (p. 626)
gel electrophoresis (p. 627)
isoelectric point (p. 628)
isoelectric focusing (p. 628)
two-dimensional electrophoresis (p. 629)
Edman degradation (p. 631)
phenyl isothiocyanate (p. 632)
overlap peptide (p. 633)

Answers to QUICK QUIZZES

1. An assay, which should be based on some unique biochemical property of the protein that is being purified, allows the detection of the protein of interest.

2. Differences in size, solubility, charge, and the specific binding of certain molecules.

3. Amino acid composition is simply the amino acids that are present in the protein. Many proteins can have the same amino acid composition. Amino acid sequence is the sequence of amino acids or the primary structure of the protein. Each protein has a unique amino acid sequence.

Problems

1. *Salting out.* Why do proteins precipitate at high salt concentrations?

2. *Salting in.* Although many proteins precipitate at high salt concentrations, some proteins require salt in order to dissolve in water. Explain why some proteins require salt to dissolve.

3. *Competition for water*. What types of R groups would compete with salt ions for water of solvation?

4. *Column choice.* (a) The octapeptide AVGWRVKS was digested with the enzyme trypsin. Would ion-exchange or gel-filtration chromatography be most appropriate for separating the products? Explain. (b) Suppose that the peptide had, instead, been digested with chymotrypsin. What would be the optimal separation technique? Explain.

5. *Frequently used in shampoos*. The detergent sodium dodecyl sulfate (SDS) denatures proteins. Suggest how SDS destroys protein structure.

6. *Making more enzyme?* In the course of purifying an enzyme, a researcher performs a purification step that results in an *increase* in the total activity to a value greater than that present in the original crude extract. Explain how the amount of total activity might increase.

7. *Protein purification problem.* Complete the following table.

Purification procedure	Total protein (mg)	Total activity (units)	Specific activity (units mg^{-1})	Purification level	Yield (%)
Crude extract	20,000	4,000,000		1	100
$(NH_4)_2SO_4$ precipitation	5,000	3,000,000			
DEAE-cellulose chromatography	1,500	1,000,000			
Gel-filtration chromatography	500	750,000			
Affinity chromatography	45	675,000			

8. *Dialysis*. Suppose that you precipitate a protein with 1 M $(NH_4)_2SO_4$, and you wish to reduce the concentration of the $(NH_4)_2SO_4$. You take 1 ml of your sample and dialyze it in 1000 ml of buffer. At the end of dialysis, what is the concentration of $(NH_4)_2SO_4$ in your sample? How could you further lower the $(NH_4)_2SO_4$ concentration?

9. *Charge to mass*. (a) Proteins treated with a sulfhydryl reagent such as β-mercaptoethanol and dissolved in sodium dodecyl sulfate have the same charge-to-mass ratio. Explain. (b) Under what conditions might the statement in part *a* be incorrect?

(c) Some proteins migrate anomalously in SDS-PAGE gels. For instance, the molecular weight determined from an SDS-PAGE gel is sometimes very different from the molecular weight determined from the amino acid sequence. Suggest an explanation for this discrepancy.

10. *A question of efficiency*. The Edman method of protein sequencing can be used to determine the sequence of proteins no longer than approximately 50 amino acids. Why is this length limitation the case?

Chapter Integration Problem

11. *Quaternary structure.* A protein was purified to homogeneity. Determination of the mass by gel-filtration chromatography yields 60 kd. Chromatography in the presence of urea yields a 30-kd species. When the chromatography is repeated in the presence of urea and β-mercaptoethanol, a single molecular species of 15 kd results. Describe the structure of the molecule.

Data Interpretation Problems

12. *Protein sequencing 1.* Determine the sequence of hexapeptide on the basis of the following data. Note: When the sequence is not known, a comma separates the amino acids. (See Table 40.2.)

Amino acid composition: (2R,A,S,V,Y)
N-terminal analysis of the hexapeptide: A
Trypsin digestion: (R,A,V) and (R,S,Y)
Carboxypeptidase digestion: no digestion
Chymotrypsin digestion: (A,R,V,Y) and (R,S)

13. *Protein sequencing 2.* Determine the sequence of a peptide consisting of 14 amino acids on the basis of the following data.

Amino acid composition: (4S,2L,F,G,I,K,M,T,W,Y)
N-terminal analysis: S
Carboxypeptidase digestion: L
Trypsin digestion: (3S,2L,F,I,M,T,W) (G,K,S,Y)
Chymotrypsin digestion: (F,I,S) (G,K,L) (L,S) (M,T) (S,W) (S,Y)
N-terminal analysis of (F,I,S) peptide: S
Cyanogen bromide treatment: (2S,F,G,I,K,L,M*,T,Y) (2S,L,W)
M*, methionine detected as homoserine

Answers to Problems

1. If the salt concentration becomes too high, the salt ions interact with the water molecules. Eventually, there are not enough water molecules to interact with the protein, and the protein precipitates.

2. If there is lack of salt in a protein solution, the proteins may interact with one another—the positive charges on one protein with the negative charges on another or several others. Such an aggregate becomes too large to be solublized by water alone. If salt is added, the salt neutralizes the charges on the proteins, preventing protein–protein interactions.

3. Charged and polar R groups on the surface of the enzyme.

4. (a) Trypsin cleaves after arginine (R) and lysine (K), generating AVGWR, VK, and S. Because they differ in size, these products could be separated by molecular exclusion chromatography.

(b) Chymotrypsin, which cleaves after large aliphatic or aromatic R groups, generates two peptides of equal size (AVGW) and (RVKS). Separation based on size would not be effective. The peptide RVKS has two positive charges (R and K), whereas the other peptide is neutral. Therefore, the two products could be separated by ion-exchange chromatography.

5. The long hydrophobic tail on the SDS molecule (p. 627) disrupts the hydrophobic interactions in the interior of the protein. The protein unfolds, with the hydrophobic R groups now interacting with the SDS rather than with one another.

6. An inhibitor of the enzyme being purified might have been present and subsequently removed by a purification step. This removal would lead to an apparent increase in the total amount of enzyme present.

7.

Purification procedure	Total protein (mg)	Total activity (units)	Specific activity (units mg^{-1})	Purification level	Yield (%)
Crude extract	20,000	4,000,000	200	1	100
$(NH_4)_2SO_4$ precipitation	5,000	3,000,000	600	3	75
DEAE–cellulose chromatography	1,500	1,000,000	667	3.3	25
Size-exclusion chromatography	500	750,000	1,500	7.5	19
Affinity chromatography	45	675,000	15,000	75	17

8. The sample was diluted 1000-fold. The concentration after dialysis is thus 0.001 M or 1 mM. You could reduce the salt concentration by dialyzing your sample, now 1 mM, in more buffer free of $(NH_4)_2SO_4$.

9. (a) Because one SDS molecule binds to a protein for every two amino acids in the proteins, in principle, all proteins will have the same charge-to-mass ratio. For instance, a protein consisting of 200 amino acids will bind 100 SDS molecules, whereas a protein consisting of 400 amino acids will bind 200 SDS molecules. The average mass of an amino acid is 110, and there is one negative charge per SDS molecule. Thus, the charge-to-mass ratio of both proteins is the same—0.0045. (b) The statement might be incorrect if the protein contains many charged amino acids. (c) The protein may be modified. For instance, serine, threonine, and tyrosine may have phosphoryl groups attached.

10. Because the cleavage does not occur every time for each peptide being sequenced. Consequently, after many repetitions (approximately 50), many different peptides are releasing different amino acids at the same time. To illustrate this point, assume that each sequencing step is 98% efficient. The proportion of correct amino acids released after 50 rounds is 0.98^{50}, or 0.4—a hopelessly impure mix.

11. Treatment with urea disrupts noncovalent bonds. Thus, the original 60-kd protein must be made of two 30-kd subunits. When these subunits are treated with urea and β-mercaptoethanol, a single 15-kd species results, suggesting that disulfide bonds link the 30-kd subunits.

12. N terminal: A
Trypsin digestion: Cleaves at R. Only two peptides are produced. Therefore, one R must be internal and the other must be the C-terminal amino acid. Because A is N terminal, the sequence of one of the peptides is AVR.
Carboxypeptidase digestion: No digestion confirms that R is the C-terminal amino acid.
Chymotrypsin digestion: Cleaves only at Y. Combined with the preceding information, chymotrypsin digestion tells us that the sequences of the two peptides are AVRY and SR.
Thus the complete peptide is AVRYSR.

13. First amino acid: S
Last amino acid: L
Cyanogen bromide cleavage: M is 10th position, C-terminal residues are: (2S,L,W)
N-terminal residues: (G,K,S,Y), tryptic peptide, ends in K
N-terminal sequence: SYGK
Chymotryptic peptide order: (S,Y), (G,K,L), (F,I,S), (M,T), (S,W), (S,L)
Sequence: SYGKLSIFTMSWSL

Selected readings for this chapter can be found online at www.whfreeman.com/Tymoczko