LODISH DECLARATION
EXHIBIT E
PART 1 OF 2

Second Edition

BIOCHEMISTRY

Lubert Stryer

STANFORD UNIVERSITY

W. H. FREEMAN AND COMPANY
New York

DESIGNER: *Robert Ishi*
ILLUSTRATOR: *Donna Salmon*
ILLUSTRATION COORDINATOR: *Audre W. Loverde*
PRODUCTION COORDINATOR: *William Murdock*
COMPOSITOR: *York Graphic Services*
PRINTER AND BINDER: *Arcata Book Group*

Library of Congress Cataloging in Publication Data

Stryer, Lubert.
Biochemistry.

Includes bibliographies and index.
1. Biological chemistry. I. Title. [DNLM: 1. Biochemistry. QU4 S928b]
QP514.2.S66 1981 574.19'2 80-24699
ISBN 0-7167-1226-1

Printed in the United States of America

10 11 12 13 14 15 KP 1 0 8 9 8 7 6 5

CONTENTS

CHAPTER 2

INTRODUCTION TO PROTEIN STRUCTURE AND FUNCTION

Proteins play crucial roles in virtually all biological processes. The significance and remarkable scope of their functions are exemplified in:

1. *Enzymatic catalysis.* Nearly all chemical reactions in biological systems are catalyzed by specific macromolecules called enzymes. Some of these reactions, such as the hydration of carbon dioxide, are quite simple. Others, such as the replication of an entire chromosome, are highly intricate. Nearly all enzymes exhibit enormous catalytic power. They usually enhance reaction rates by at least a millionfold. Indeed, chemical transformations rarely occur at perceptible rates in vivo in the absence of enzymes. Several thousand enzymes have been characterized, and many of them have been crystallized. The striking fact is that all known enzymes are proteins. Thus, proteins play the unique role of determining the pattern of chemical transformations in biological systems.

2. *Transport and storage.* Many small molecules and ions are transported by specific proteins. For example, hemoglobin transports oxygen in erythrocytes, whereas myoglobin, a related protein, transports oxygen in muscle. Iron is carried in the plasma of blood

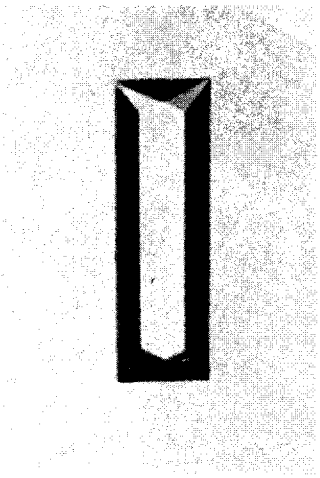


Figure 2-1
Photomicrograph of a crystal of hexokinase, a key enzyme in the utilization of glucose. [Courtesy of Dr. Thomas Steitz and Dr. Mark Yeager.]

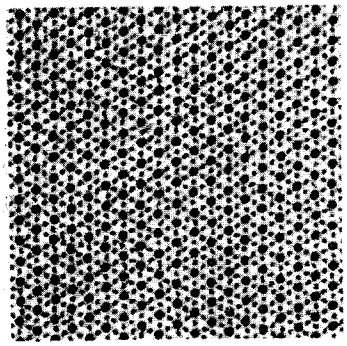


Figure 2-2
Electron micrograph of a cross section of insect flight muscle showing a hexagonal array of two kinds of protein filaments. [Courtesy of Dr. Michael Reedy.]

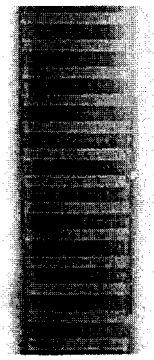


Figure 2-3
Electron micrograph of a fiber of collagen. [Courtesy of Dr. Jerome Gross and Dr. Romaine Bruns.]

by transferrin and is stored in the liver as a complex with ferritin, a different protein.

3. *Coordinated motion.* Proteins are the major component of muscle. Muscle contraction is accomplished by the sliding motion of two kinds of protein filaments. On the microscopic scale, such coordinated motions as the movement of chromosomes in mitosis and the propulsion of sperm by their flagella also are produced by contractile assemblies consisting of proteins.

4. *Mechanical support.* The high tensile strength of skin and bone is due to the presence of collagen, a fibrous protein.

5. *Immune protection.* Antibodies are highly specific proteins that recognize and combine with such foreign substances as viruses, bacteria, and cells from other organisms. Proteins thus play a vital role in distinguishing between self and nonself.

6. *Generation and transmission of nerve impulses.* The response of nerve cells to specific stimuli is mediated by receptor proteins. For example, rhodopsin is the photoreceptor protein in retinal rod cells. Receptor molecules that can be triggered by specific small molecules, such as acetylcholine, are responsible for transmitting nerve impulses at synapses—that is, at junctions between nerve cells.

7. *Control of growth and differentiation.* Controlled sequential expression of genetic information is essential for the orderly growth and differentiation of cells. Only a small fraction of the genome of a cell is expressed at any one time. In bacteria, repressor proteins are important control elements that silence specific segments of the DNA of a cell. A quite different way in which proteins act in differentiation is exemplified by nerve growth factor, a protein complex that guides the formation of neural networks in higher organisms.

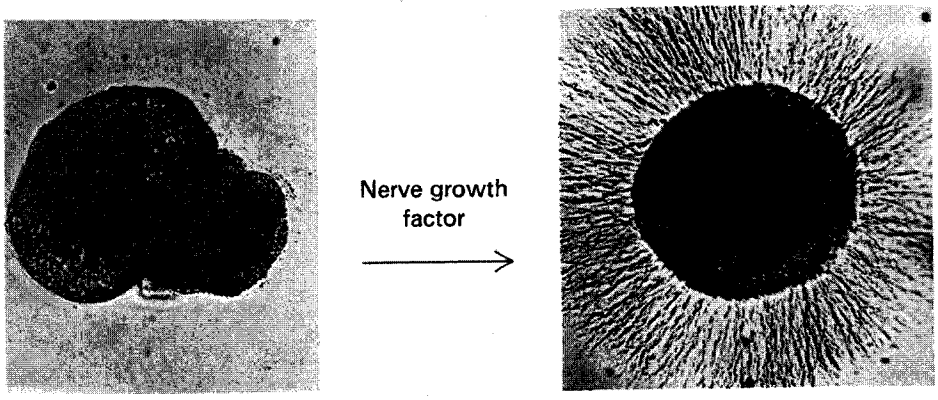
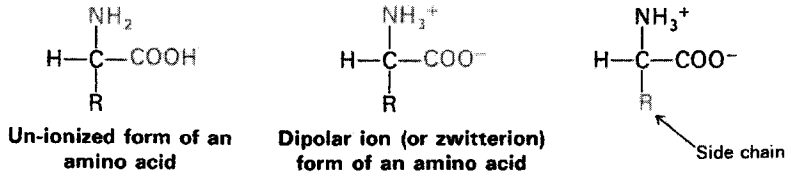


Figure 2-4
Photomicrograph of a ganglion showing the proliferation of nerves after addition of nerve growth factor, a complex of proteins. [Courtesy of Dr. Eric Shooter.]

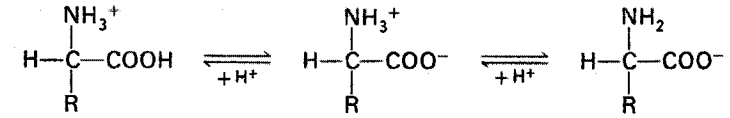
Amino acids are the basic structural units of proteins. An amino acid consists of an amino group, a carboxyl group, a hydrogen atom, and a distinctive R group bonded to a carbon atom, which is called the α-carbon (Figure 2-5). An R group is referred to as a *side*



Un-ionized form of an amino acid **Dipolar ion (or zwitterion) form of an amino acid**

Figure 2-5
Structure of the un-ionized and zwitterion forms of an amino acid.

chain for reasons that will be evident shortly. Amino acids in solution at neutral pH are predominantly *dipolar ions* (or *zwitterions*) rather than un-ionized molecules. In the dipolar form of an amino acid, the amino group is protonated ($—NH_3^+$) and the carboxyl group is dissociated ($—COO^-$). The ionization state of an amino acid varies with pH (Figure 2-6). In acid solution (e.g., pH 1), the



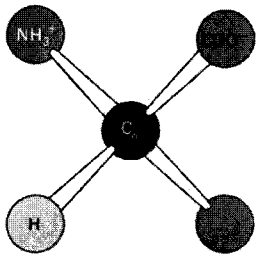
Predominant form at pH 1 **Predominant form at pH 7** **Predominant form at pH 11**

Figure 2-6
Ionization states of an amino acid as a function of pH.

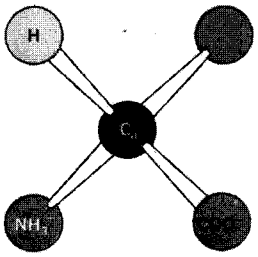
carboxyl group is un-ionized ($—COOH$) and the amino group is ionized ($—NH_3^+$). In alkaline solution (e.g., pH 11), the carboxyl group is ionized ($—COO^-$) and the amino group is un-ionized ($—NH_2$). The concept of pH and the acid-base properties of amino acids are discussed further in the Appendix to this chapter.

The tetrahedral array of four different groups about the α-carbon atom confers optical activity on amino acids. The two mirror-image forms are called the L-isomer and the D-isomer (Figure 2-7). *Only L-amino acids are constituents of proteins.* Hence, the designation of the optical isomer will be omitted and the L-isomer implied in discussions of proteins herein, unless otherwise noted.

Twenty kinds of side chains varying in *size, shape, charge, hydrogen-bonding capacity, and chemical reactivity* are commonly found in proteins. Indeed, all proteins in all species, from bacteria to humans, are constructed from the same set of twenty amino acids. This fun-



L-Isomer



D-Isomer

Figure 2-7
Absolute configurations of the L- and D-isomers of amino acids.

damental alphabet of proteins is at least two billion years old. The remarkable range of functions mediated by proteins results from the diversity and versatility of these twenty kinds of building blocks. In subsequent chapters, we will explore ways in which this alphabet is used to create the intricate three-dimensional structures that enable proteins to participate in so many biological processes.

Let us look at this repertoire of amino acids. The simplest one is glycine, which contains a hydrogen atom as its side chain (Figure 2-8). Alanine has a methyl group as its side chain. The other amino

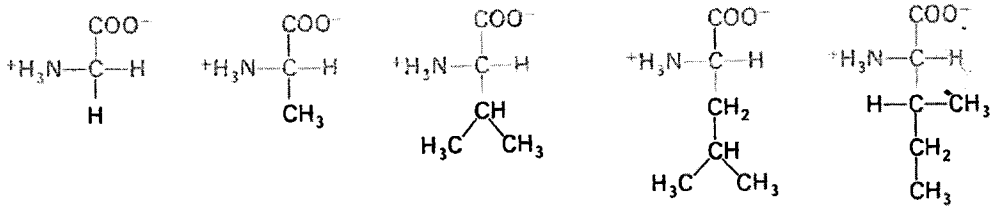


Figure 2-8
Amino acids having aliphatic side chains.

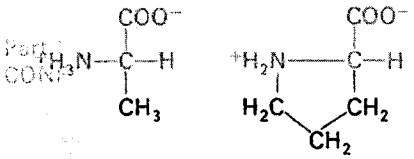


Figure 2-9
Proline differs from the other common amino acids in that it has a secondary amino group.

acids that have hydrocarbon side chains are valine, leucine, isoleucine, and proline. However, proline differs from the other amino acids in the basic set of twenty in that it contains a secondary rather than a primary amino group (Figure 2-9). Strictly speaking, proline is an imino acid rather than an amino acid. The side chain of proline is bonded to both the amino group and the α-carbon, which results in a cyclic structure.

Two amino acids, serine and threonine, contain aliphatic hydroxyl groups (Figure 2-10).

There are three common aromatic amino acids: phenylalanine, tyrosine, and tryptophan (Figure 2-11).

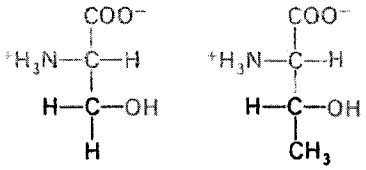


Figure 2-10
Serine and threonine have aliphatic hydroxyl side chains.

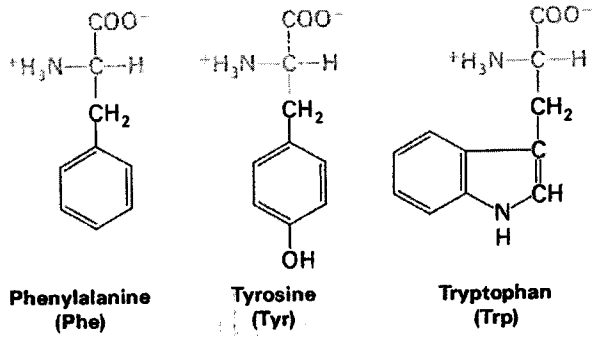
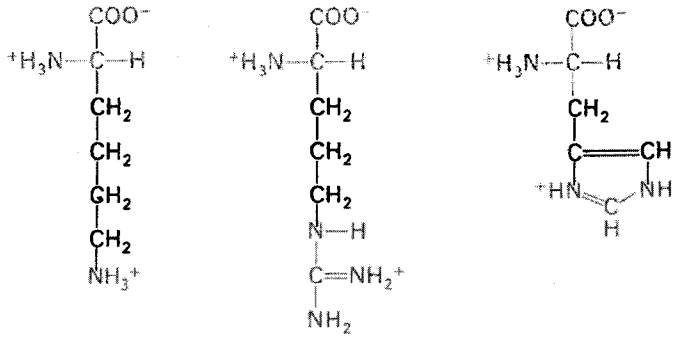


Figure 2-11
Phenylalanine, tyrosine, and tryptophan have aromatic side chains.

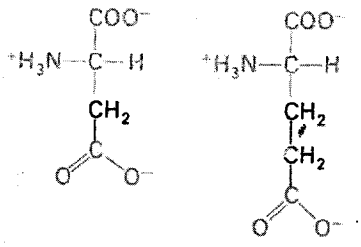
The side chains of the amino acids mentioned so far are uncharged at physiological pH. We turn now to some charged side chains. Lysine and arginine are positively charged at neutral pH, whereas whether histidine is positively charged or neutral depends on its local environment. These basic amino acids are shown in Figure 2-12. The negatively charged side chains are those of glu-



Lysine (Lys) **Arginine (Arg)** **Histidine (His)**

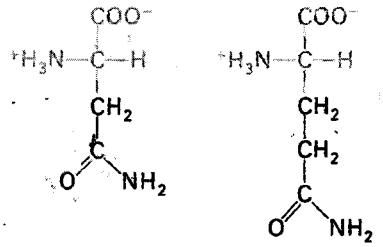
Figure 2-12
Lysine, arginine, and histidine have basic side chains.

tamic acid and aspartic acid (Figure 2-13). These amino acids will be called glutamate and aspartate to emphasize the fact that they are negatively charged at physiological pH. The uncharged derivatives of glutamate and aspartate are glutamine and asparagine (Figure 2-14), each of which contains a terminal amide group rather than a carboxylate. Finally, there are two amino acids whose side chains contain a sulfur atom: methionine and cysteine (Figure 2-15). As will be discussed shortly, cysteine plays a special role in some proteins by forming disulfide cross-links.



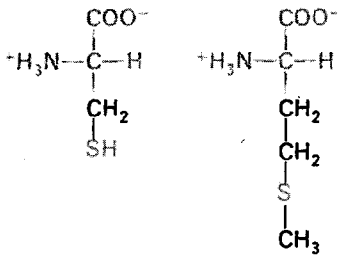
Aspartate (Asp) **Glutamate (Glu)**

Figure 2-13
Aspartate and glutamate have acidic side chains.



Asparagine (Asn) **Glutamine (Gln)**

Figure 2-14
Asparagine and glutamine have amide side chains.



Cysteine (Cys) **Methionine (Met)**

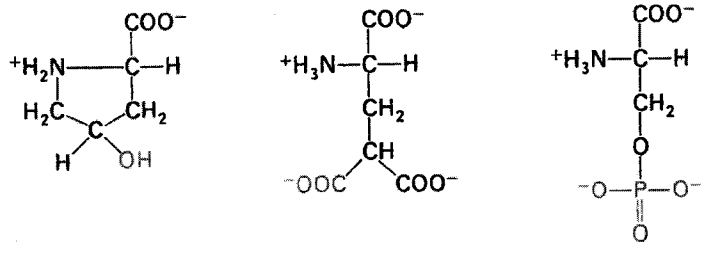
Figure 2-15
Cysteine and methionine have sulfur-containing side chains.

Table 2-1
Abbreviations for amino acids

Amino acid	*Three-letter abbreviation*	*One-letter symbol*
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Asparagine or aspartic acid	Asx	B
Cysteine	Cys	C
Glutamine	Gln	Q
Glutamic acid	Glu	E
Glutamine or glutamic acid	Glx	Z
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

SPECIAL AMINO ACIDS SUPPLEMENT THE BASIC SET OF TWENTY

Some proteins contain special amino acids that are formed by modification of a common amino acid following its incorporation into the polypeptide chain. For example, collagen contains hydroxyproline, a hydroxylated derivative of proline (Figure 2-16). The added



Hydroxyproline **γ-Carboxyglutamate** ***O*-Phosphoserine**

Figure 2-16
Some modified amino acid residues in proteins: hydroxyproline, γ-carboxyglutamate, and phosphoserine. Groups added after the polypeptide chain is synthesized are shown in red.

hydroxyl group stabilizes the collagen fiber, as will be discussed later (p. 192). The biological importance of this modification is evident in scurvy, which results from insufficient hydroxylation of collagen. Another special amino acid is γ-carboxyglutamate. Defective carboxylation of glutamate in prothrombin, a clotting protein, can lead to hemorrhage (p. 176). The most ubiquitous modified amino acid in proteins is phosphoserine. The action of some hormones is mediated by the phosphorylation and dephosphorylation of specific serine residues in a variety of proteins (p. 368).

AMINO ACIDS ARE LINKED BY PEPTIDE BONDS TO FORM POLYPEPTIDE CHAINS

In proteins, the α-carboxyl group of one amino acid is joined to the α-amino group of another amino acid by a *peptide bond* (also called an amide bond). The formation of a dipeptide from two amino acids by loss of a water molecule is shown in Figure 2-17. The equilibrium of this reaction lies far on the side of hydrolysis rather than synthesis. Hence, the biosynthesis of peptide bonds requires an input of free energy, whereas their hydrolysis is thermodynamically downhill.

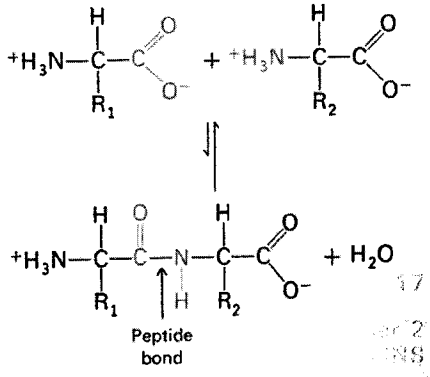


Figure 2-17
Formation of a peptide bond.

Many amino acids, usually more than a hundred, are joined by peptide bonds to form a *polypeptide chain*, which is an unbranched structure (Figure 2-18). An amino acid unit in a polypeptide is

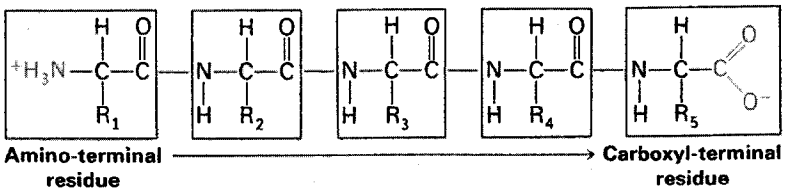


Figure 2-18
A pentapeptide. The constituent amino acid residues are outlined. The chain starts at the amino end.

called a *residue*. A polypeptide chain has direction because its building blocks have different ends—namely, the α-amino and the α-carboxyl groups. By convention, *the amino end is taken to be the beginning of a polypeptide chain.* The sequence of amino acids in a polypeptide chain is written starting with the amino-terminal residue. Thus, in the tripeptide alanine-glycine-tryptophan, alanine is the amino-terminal residue and tryptophan is the carboxyl-terminal residue. Note that tryptophan-glycine-alanine is a different tripeptide.

A polypeptide chain consists of a regularly repeating part, called the *main chain*, and a variable part, comprising the distinctive *side chains* (Figure 2-19). The main chain is sometimes termed the backbone.

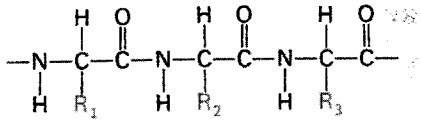


Figure 2-19
A polypeptide chain is made up of a regularly repeating *backbone* and distinctive *side chains* (R_1, R_2, R_3, shown in green).

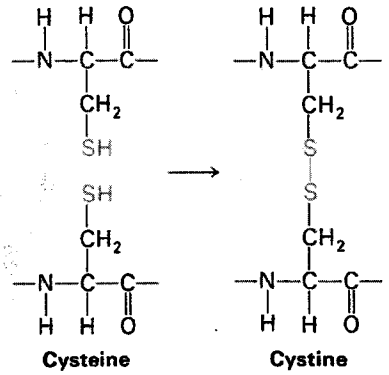


Figure 2-20
A disulfide bridge (—S—S—) is formed from the sulfhydryl groups (—SH) of two cysteine residues. The product is a *cystine* residue.

In some proteins, a few side chains are cross-linked by *disulfide bonds*. These cross-links are formed by the oxidation of cysteine residues. The resulting disulfide is called cystine (Figure 2-20). No other covalent cross-links are generally found in proteins.

PROTEINS CONSIST OF ONE OR MORE POLYPEPTIDE CHAINS

Many proteins, such as myoglobin, consist of a single polypeptide chain. Others contain two or more chains, which may be either identical or different. For example, hemoglobin is made up of two chains of one kind and two of another kind. These four chains are held together by noncovalent forces. Alternatively, the polypeptide chains of some multichain proteins are linked by disulfide bonds. The two chains of insulin, for example, are joined by two disulfide bonds.

PROTEINS CAN BE PURIFIED BY A VARIETY OF TECHNIQUES

The purification of a protein is an indispensable step toward the elucidation of its mechanism of action. Several thousand proteins have been isolated in pure form. Proteins can be separated from each other and from other kinds of molecules on the basis of such characteristics as *size, solubility, charge, and specific binding affinity*. In purifying a protein, various separation methods are tried and their efficiency is evaluated by assaying for a distinctive property of the protein of interest. The assay for an enzyme, for example, is usually based on its specific catalytic activity. The total amount of protein is also measured so that the degree of purification obtained in a particular step can be determined.

Proteins can be separated from small molecules by dialysis through a semipermeable membrane (Figure 2-21). Molecules

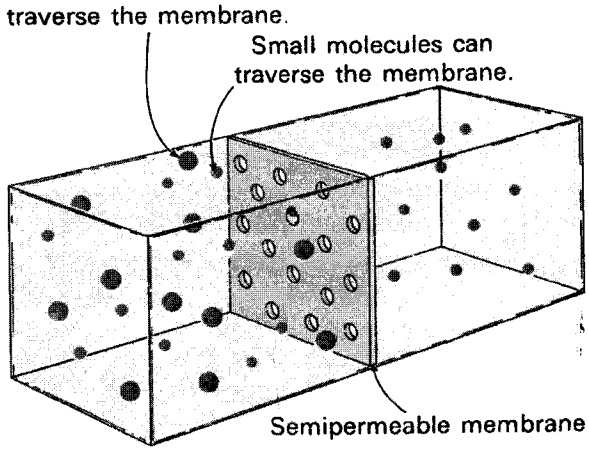


Figure 2-21
Separation of molecules on the basis of size by dialysis.

whose mass is greater than about 15 kilodaltons (kdal) are retained inside a typical dialysis bag, whereas smaller molecules and ions traverse the pores of such a dialysis membrane and emerge in the dialysate outside the bag. Separations on the basis of size can also be achieved by the technique of *gel-filtration chromatography* (Figure 2-22). The sample is applied to the top of a column consisting of an

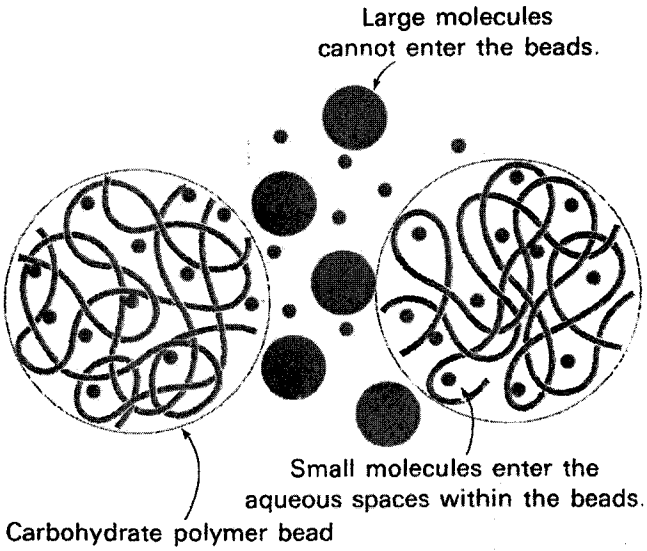


Figure 2-22
Separation of molecules on the basis of size by gel-filtration chromatography.

insoluble but highly hydrated carbohydrate polymer in the form of beads, which are typically 0.1 mm in diameter. Sephadex is a commonly used commercial preparation. Small molecules can enter these beads, but large ones cannot. The result is that small molecules are distributed both in the aqueous solution inside the beads and between them, whereas large molecules are located only in the solution between the beads. Large molecules flow more rapidly through this column and emerge first because a smaller volume is accessible to them.

Proteins can also 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 an ion-exchange column containing carboxylate groups, whereas a negatively charged protein will not. Such a positively charged protein can be released from the column by adding sodium chloride or another salt to the eluting buffer. Sodium ions compete with positively charged groups on the protein for binding to the column. Proteins that have a low density of net positive charge will tend to emerge first, followed by those having a higher charge density. Factors other than net charge also influence the behavior of proteins on ion-exchange columns. The net charge of a protein also influences its rate of migration in an electric field. This principle is exploited in *electrophoresis*, a technique that will be discussed in more detail in a subsequent chapter (p. 90). The very high resolving power of electrophoresis is worth noting here. It is

Dalton—
A unit of mass very nearly equal to that of a hydrogen atom (precisely equal to 1.0000 on the atomic mass scale).
The terms "dalton" and "molecular weight" are used interchangeably; for example, a 20,000-dalton protein has a molecular weight of 20,000.
Named after John Dalton (1766–1844), who developed the atomic theory of matter.

Kilodalton (kdal)—
A unit of mass equal to 1000 daltons. Most proteins have a mass of between 10 and 100 kdal.

remarkable that more than a thousand different proteins in the simple bacterium *E. coli* can be resolved in a single experiment by two-dimensional electrophoresis (Figure 2-23).

Figure 2-23
Two-dimensional electrophoresis of the proteins from *E. coli*. More than a thousand different proteins from this bacterium have been resolved. These proteins were separated according to their isoelectric pH in the horizontal direction and their molecular weight in the vertical direction. [Courtesy of Dr. Patrick H. O'Farrell.]

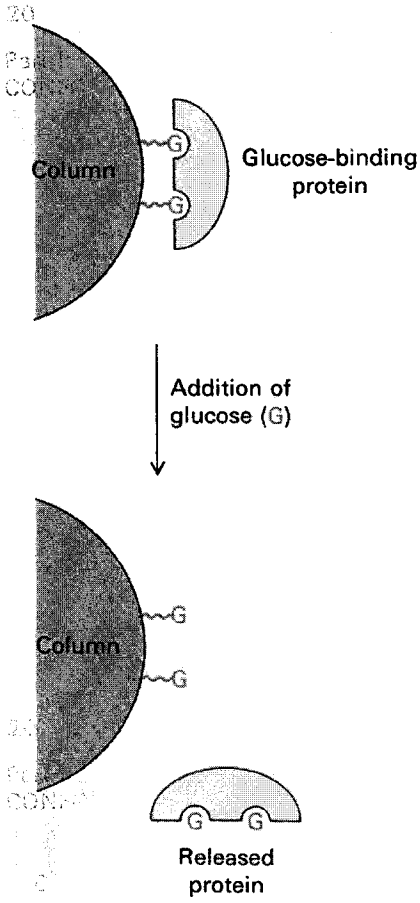


Figure 2-24
Affinity chromatography of concanavalin A (shown in yellow) on a column containing covalently attached glucose residues (G).

Affinity chromatography is another powerful and generally applicable means of purifying proteins. This technique takes advantage of the high affinity of many proteins for specific chemical groups. For example, concanavalin A, a plant protein, can be purified by passing a crude extract through a column that contains covalently attached glucose residues. Concanavalin A binds to such a column because it has affinity for glucose, whereas most other proteins are not adsorbed. The bound concanavalin A can then be released from this 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 2-24). In general, affinity chromatography can be effectively used to isolate a protein that recognizes group X by (1) convalently attaching X or a derivative of it to a column, (2) adding a mixture of proteins to this column, which is then washed with buffer to remove unbound proteins, and (3) eluting the desired protein by adding a high concentration of a soluble form of X.

PROTEINS HAVE UNIQUE AMINO ACID SEQUENCES THAT ARE SPECIFIED BY GENES

In 1953, Frederick Sanger determined the amino acid sequence of insulin, a protein hormone (Figure 2-25). *This work is a landmark in biochemistry because it showed for the first time that a protein has a precisely defined amino acid sequence.* This accomplishment also stimulated other scientists to carry out sequence studies of a wide variety of proteins. Indeed, the complete amino acid sequences of several

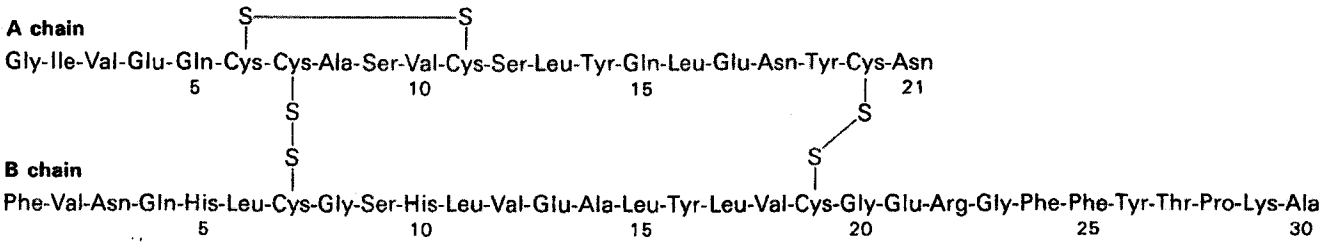


Figure 2-25
Amino acid sequence of bovine insulin.

hundred proteins are now known. The striking fact is that each protein has a unique, precisely defined amino acid sequence. A series of incisive studies in the late 1950s and early 1960s revealed that the amino acid sequences of proteins are genetically determined. The sequence of nucleotides in DNA, the molecule of heredity, specifies a complementary sequence of nucleotides in RNA, which in turn specifies the amino acid sequence of a protein (p. 597). Furthermore, proteins are synthesized from their constituent amino acids by a common mechanism.

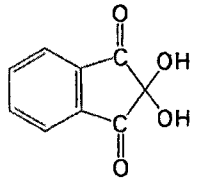
The importance of determining amino acid sequences of proteins is fourfold. First, the determination of the sequence of a protein is a significant step toward the elucidation of the molecular basis of its biological activity. A sequence is particularly informative if it is considered together with other chemical and physical data. Second, the sequences and detailed three-dimensional structures of numerous proteins need to be known so that the rules governing the folding of polypeptide chains into highly specific three-dimensional forms can be deduced. The amino acid sequence is the link between the genetic message in DNA and the three-dimensional structure that is the basis of a protein's biological function. Third, alterations in amino acid sequence can produce abnormal function and disease. Fatal disease, such as sickle-cell anemia, can result from a change in a single amino acid in a single protein. Sequence determination is thus part of molecular pathology, an emerging area of medicine. Fourth, the amino acid sequence of a protein reveals much about its evolutionary history. Amino acid sequences of unrelated proteins are very different. Proteins resemble one another in their amino acid sequences only if they have a common ancestor. Consequently, molecular events in evolution can be traced from amino acid sequences.

Protein—
A word coined by Jöns J. Berzelius in 1838 to emphasize the importance of this class of molecules. Derived from the Greek word *proteios*, which means "of the first rank."

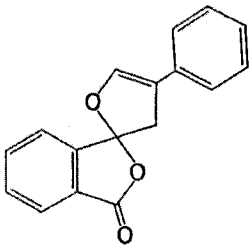
EXPERIMENTAL METHODS FOR THE DETERMINATION OF AMINO ACID SEQUENCE

Let us first consider how the sequence of a short peptide could be determined. Suppose that the peptide has six amino acid residues in the following sequence:

Ala-Gly-Asp-Phe-Arg-Gly



Ninhydrin



Fluorescamine

The abbreviations used are the standard ones given in Table 2-1 (p. 16). First, the *amino acid composition* of the peptide is determined. The peptide is hydrolyzed into its constituent amino acids by heating it in 6 N HCl at 110°C for 24 hours. The amino acids in the hydrolysate are separated by ion-exchange chromatography on a column of sulfonated polystyrene. The separated amino acids are detected by the color produced on heating with *ninhydrin*: α-amino acids give an intense blue color, whereas imino acids, such as proline, give a yellow color. The technique is very sensitive; it can detect even a microgram of an amino acid, which is about the amount present in a thumbprint. The quantity of amino acid is proportional to the optical absorbance of the solution after heating with ninhydrin. As little as a few nanograms of an amino acid can be detected using *fluorescamine*, which reacts with its α-amino group to form a highly fluorescent product. The identity of the amino acid is revealed by its elution volume, which is the volume of buffer used to remove the amino acid from the column (Figure 2-26). A comparison of the chromatographic patterns of the hydrolysate with that of a standard mixture of amino acids shows that the amino acid composition of the peptide is

$$(\text{Ala, Arg, Asp, Gly}_2\text{, Phe})$$

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

The amino-terminal residue of a protein or peptide can be identified by labeling it with a compound with which it forms a stable

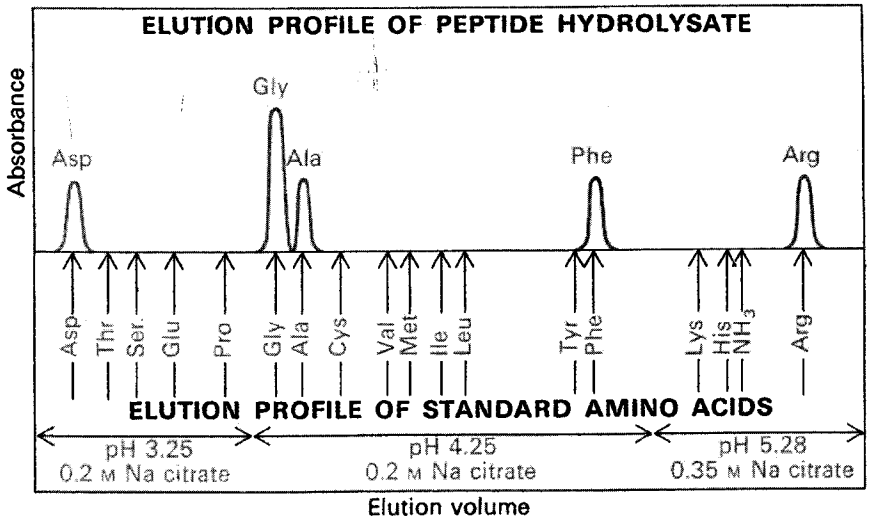
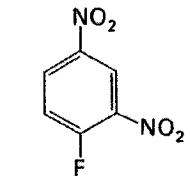


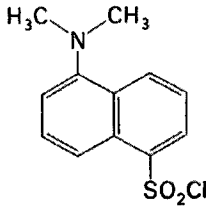
Figure 2-26
Different amino acids in a peptide hydrolysate can be separated by ion-exchange chromatography on a sulfonated polystyrene resin (such as Dowex-50). Buffers of increasing pH are used to elute the amino acids from the column. Aspartate, which has an acidic side chain, is first to emerge, whereas arginine, which has a basic side chain, is the last.

covalent link (Figure 2-27). *Fluorodinitrobenzene* (FDNB), first used for this purpose by Sanger, reacts with the uncharged α-NH_2 group to form a yellow dinitrophenyl (DNP) derivative of the peptide. The bond between the DNP and the terminal amino group is stable under conditions that hydrolyze peptide bonds. Hydrolysis of the DNP-peptide in 6 N HCl yields a DNP-amino acid, which is identified as DNP-alanine by its chromatographic properties.



Fluorodinitrobenzene

Dansyl chloride is now often used to identify amino-terminal residues. It reacts with amino groups to form highly fluorescent and stable sulfonamide derivatives. A few nanograms of an amino-terminal residue can be identified after acid hydrolysis of peptide bonds.



Dansyl chloride

Although the DNP and dansyl methods for the determination of the amino-terminal residue are powerful, they cannot be used repetitively on the same peptide because the peptide is totally de-

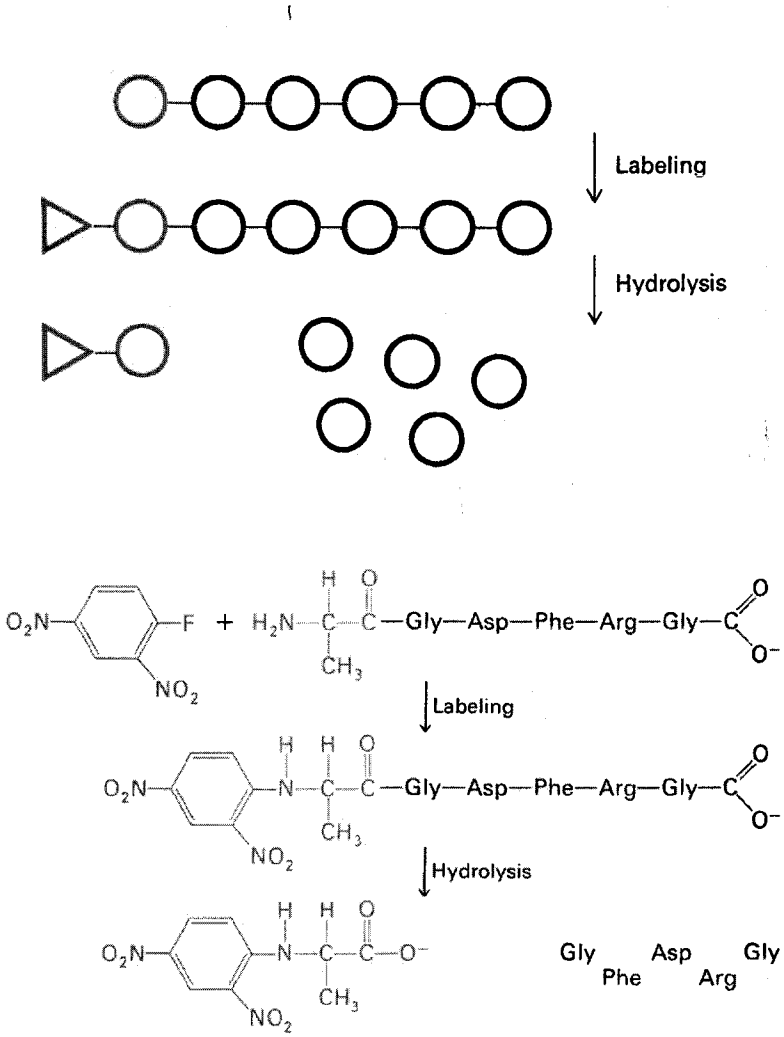


Figure 2-27
Determination of the amino-terminal residue of a peptide. Fluorodinitrobenzene (Sanger's reagent) is used to label the peptide, which is then hydrolyzed. The DNP-amino acid (DNP-alanine in this example) is identified by its chromatographic characteristics.