Protein
Proteins: A General Background

Proteins occupy a central position in the structure and function of all living organisms. Some proteins serve as structural components while others function in communication, defense and cell regulation. The enzyme proteins act as biological catalysts which control the pace and nature of essentially all biochemical events. Indeed, although DNA serves as the genetic blueprint of a cell, none of the life processes would be possible without proteins.

The fundamental unit of proteins is the amino acid. Each amino acid has an amino group (NH$_3^+$) and a carboxylic acid group (COO$^-$) attached to a central carbon atom called the alpha carbon. Also attached to the alpha carbon are a hydrogen atom and an R-group or side chain. The common amino acids have the general structure shown below; the form presented is termed "zwitterion" form in which the species is at its isoelectric point (when the net charge is 0). This form is physiologically significant, since the average pH of the human body is 7.4, which is well in between the pK$_1$ and pK$_2$ dissociation points of the species. Note that every amino acid is at least dibasic (or diacidic, which ever way you look at it) due to two ionizable groups on the "core" structure (the NH$_3^+$ and COO$^-$ groups), which means that it can have at least 3 ionization states. Consider the tribasic amino acid cysteine (cysteine is an amino acid, cystine is a complex of 2 cysteines bound together by a disulfide bond). Cysteine has 3 pK, or 4 ionization states, occurring at pH 1.7 to 2.6, 8.3, and 8.8 to 10.8. Its tribasic nature is due to an additional ionizable group which is a part of the specific side chain (R).

\[
\begin{align*}
\text{NH}_3^+ & \\
\text{I} & \\
\text{R - C}^* & \text{- COO}^- \\
\text{I} & \\
\text{H} & 
\end{align*}
\]

The C stands for a carbon atom; C* is the alpha carbon; H is hydrogen; N is nitrogen, O is oxygen, -NH$_3^+$ is an amino group and -COO$^-$ is the carboxylic acid group. R is a general term for anyone of several different side chains that determine the nature of different amino acids; there are a total of 20 different side chains and hence, 20 different amino acids.

During ionization, different ionizable groups are removed preferentially due to electrochemical thermodynamic considerations. The rationale and process is as follows: the electron withdrawing effect of the positively charged nitrogen causes oxygen to lose electrons more readily (and hence, lose H$^+$), so COOH terminus loses H$^+$ first, followed by the R group and finally the electrophilic nitrogen.

There are 20 amino acids commonly found in proteins and these differ from each other in the nature of the R groups attached to the alpha carbon. A convenient classification of amino acids depends on the number of acidic and basic groups that are present. Thus, the neutral amino acids contain one amino and one carboxyl group. The acidic amino acids have an excess of (acidic) carboxyl groups over (basic) amino groups. The basic amino acids possess an excess of basic amino groups. The major amino acids groups are as follows. The neutral amino acids are glycine, alanine, valine, leucine, isoleucine, serine, threonine, cysteine, methionine, phenylalanine, tyrosine, tryptophan and proline. The acidic amino acids are aspartic acid, asparagine, glutamic acid and glutamine. The basic amino acids are arginine, lysine and histidine. You will notice the commonality of each grouping upon examination of the physical structure of the side chains (R-groups). Each amino acid has an approximate molecular weight of 100 Daltons (this is an average value from all 20 amino acids).

Proteins are composed of amino acids linked into chains by peptide bonds as shown in Figure 40. Two amino acids joined by a single peptide bond form a dipeptide; three amino acids form a tri-peptide, and a large number of amino acids joined together constitute a polypeptide. A protein is a polypeptide chain that contains more than 50 to
The three-dimensional structure of a protein is due to the type and sequence of its constituent amino acids. Since the amino acid sequence of each protein is unique, it follows that different proteins assume different shapes. Thus, there is a remarkable diversity of three-dimensional protein forms. The conformation of a protein is usually of critical importance in the protein’s function. For example, a protein can be unfolded into a polypeptide chain that has lost its original shape. In general, proteins such as enzymes are rendered nonfunctional upon unfolding because the functional activity is dependent on the proteins native shape. This process is called denaturation. Most proteins can be denatured by heating, by certain detergents, and by extremes of pH. The ionic detergent, sodium dodecyl sulfate (SDS), is often used to denature proteins. The denaturing treatment can frequently be reversed, for example by removing the detergent or by neutralizing the pH. During this renaturing process, the polypeptide chain spontaneously refolds into its original conformation and the protein regains its biological activity. A similar folding process occurs in the cell for when a polypeptide is constructed on the ribosomes, it folds into a biologically active conformation. Thus, the three-dimensional folding of a protein and its biological properties are directed by the sequence of amino acid residues along the polypeptide chain.

Biochemists have identified three structural levels that define the three-dimensional shape of a protein. These levels of organization are secondary structure, tertiary structure, and quaternary structure. Figure 42 shows examples of these levels of organization. The major force involved in the formation and maintenance of these structures are various types of weak, noncovalent bonds (hydrogen bond, ionic bond, hydrophobic interaction and Van der Waals bond) that are formed between the amino acid residues and between the amino acid residues and water. Although a noncovalent bond typically has less than 1/20 the strength of a covalent bond, a large number of noncovalent bonds participate in the folding of a single protein into its native conformation (and hence, the reason that noncovalent bonds can act to maintain these structures at such an individually weak strength of bond is that many noncovalent bonds are involved).

The spatial arrangement of the protein backbone that is generated from the folding of the polypeptide chain is called the secondary structure of the protein. The secondary structures of proteins are stabilized by hydrogen bonds in which a hydrogen serves as a bridge between oxygen and nitrogen atoms (\(-\text{C}=\text{O}\)–H–N). A common secondary structure is the \(\alpha\)-helix which consists of a single polypeptide chain coiled into a rigid cylinder. In the \(\alpha\)-helix, each peptide bond along the polypeptide is itself hydrogen bonded to other peptide bonds. Many enzymes contain small regions of \(\alpha\)-helices, while long sections of the \(\alpha\)-helix are often found in proteins involved in cell structure. Another type of secondary structure of proteins is the \(\beta\)-sheet, which is a central organizing feature of enzymes, antibodies and most other proteins that perform nonstructural functions. Here, a single polypeptide chain folds back and forth upon
itself to produce a rather rigid sheet. Hydrogen bonds between neighboring polypeptide chains are a major stabilizing force for the β-sheet conformation.

The tertiary structure of a protein describes the detailed features of the three-dimensional conformation of the polypeptide chain. It is brought about by the interactions between the amino acid side chains which cause the folding and bending of α-helix and β-sheet segments of the protein. One very important interaction at this level of organization involves the hydrophobic and hydrophilic side chains of the amino acid residues. Hydrophobic amino acids, such as phenylalanine and leucine, show limited solubility in water. Thus, these hydrophobic residues in a protein tend to cluster on the inside of the protein in order to avoid contact with the aqueous environment. Hydrophilic amino acids such as glutamic acid and lysine are readily soluble in water, and thus these amino acids arrange themselves on the surface of the protein molecule, where they can interact with water and with other hydrophilic side chains. The consequence of these interactions is that a polypeptide chain typically folds spontaneously into a stable, usually globular structure, with the hydrophobic side chains packed into the central core of the protein and the hydrophilic side chains forming the irregular, external surface.

Some proteins contain more than one polypeptide chain. For example, each molecule of human hemoglobin consists of four polypeptide chains which are held together by a variety of noncovalent bonds. The arrangement of the polypeptides in such proteins is called the quaternary structure.

The three dimensional arrangements of proteins are often further stabilized by covalent interactions, such as disulfide bonds. When two cysteine residues are brought into close proximity in the presence of oxygen, the sulfhydryl groups present on their side chains will spontaneously interact to form a disulfide linkage (Figure 43). This covalent interaction can occur between 2 cysteines on the same polypeptide chain when these residues are brought together during folding or between cysteines on different polypeptides that associate with each other in a multipeptide protein. Thus, intra-chain disulfide bonds stabilize tertiary structures while inter-chain disulfide bonds stabilize quaternary structures.
Enzymes, Isoenzymes and Lactate Dehydrogenase

Enzymes are proteins which catalyze stereospecific reactions involving the conversion of a substrate (S) to a product (P) but do not get altered themselves in the process. This catalysis is due to a lowering of the activation energy of the reaction in such a way that a catalyzed reaction can be accelerated in rate by at least $10^6$ relative to an uncatalyzed reaction (in some literature I have seen $10^{12}$, but the forementioned is a conservative estimate)! This acceleration is due to enthalpic and entropic considerations which effect the energy of activation, Remember, the equilibrium of the catalyzed reaction is not altered, it is only achieved at a greater rate.

$$K_{eq}$$

$$S \xrightarrow{\text{enzyme}} P$$

enzyme
coenz, cofact.
or prosthetic gr.

In order for an enzyme to be active, it may in addition to the protein unit require a cofactor, prosthetic group, or coenzyme. Coenzymes and prosthetic groups are small nonprotein organic molecules which are required by some enzymes for their catalytic expression. These compounds arise from common vitamins. Cofactors as a class are metal ions which are required for catalytic activity of some of the enzymes.

$$\text{coenz, cofactor}$$
or prosthetic gr.

Apoenzyme --------------> Holoenzyme

(inactive) (active)

Coenzymes are bound to the enzymes with noncovalent bonds whereas prosthetic groups are attached by covalent linkages. Both the coenzymes and the prosthetic groups take part and are chemically changed during the course of enzyme catalyzed reaction. On the other hand, cofactors are bound noncovalently to the enzymes to produce an active conformation of the enzyme but are not generally chemically altered during the course of an enzyme catalyzed reaction. Examples of each group are as follows: coenzymes, NADH; cofactors, $\text{Mg}^{2+}$; prosthetics, $\text{P0}_4^{3-}$. Remember, coenzymes are specific for enzymatic reactions (hence, enzyme specific) and due to their size, they can not freely cross membranes; hence, coenzymes can act as regulators.

Enzymes are a delicate group of proteins and hence, the activity of enzymes can be affected by a wide range of parameters. At extreme pH, temperature and in the presence of detergents, enzymes become denatured, or lose their conformationally active shape. Also, by loss of a cofactor, covalent modification, the presence of reducing agents (dithiothreitol or $\beta$-mercaptoethanol) or heavy metals, the catalytic site of an enzyme can be inactivated. Finally, upon cellular breakage, there is a release of proteases, which can cause proteolytic degradation. This effect is minimized by chelators like PHSF or EDTA. In the laboratory, all enzymes must be handled carefully because they are expensive and sensitive reagents that are easily inactivated or destroyed by careless use. Enzymes should always be stored at -20°C in freezers that maintain a constant temperature. For use, the tube containing the enzyme should be placed quickly on ice and the enzyme

---

1Enzymes are generally sensitive to very high or very low pH values and are denatured under these conditions. Enzymes are most active within a narrow pH range. This is dependent on the species of interest; for example, some active sites on enzymes are basic, while some are acidic and hence, the pH of the medium can have different effects on these different enzymes. Remember, we are interested in the effect of pH on the active site region and regulatory sites, regions on the protein that have not been classified to serve a function are still significant to an unknown extent.
should be removed with a sterile pipet tip. After the enzyme is removed, the tube should be immediately be returned to -20°C. Do not let enzymes stay for extended periods on ice and do not expose them to temperatures greater than 0°C.

Enzymes have many different binding sites to interact with other species; these sites can act as substrate or regulatory sites. The relative binding site is where all species of interest (substrates, cofactors, etc.) bind relative to each other, or in close enough proximity to each other to cause the reaction to occur at the active site. The active site (or catalytic site) is a specific spatial arrangement of side chains of amino acid residues which creates a region facilitating the conversion of S to P. The regulatory sites are numerous and merit an expanded discussion; regulatory sites can be inhibitory or stimulatory and can be bound by different species. Noncovalent, allosteric mechanisms are most associated with enzymes, since allosteric mechanisms operate on a hyperbolic effect curve and hence induce greater effects on enzyme activity at lower effector concentrations while the noncovalent nature of the interaction allows for dissociable reactivity of the enzyme with the effector; these species are subject of feedback regulation. An example of noncovalent allosteric mechanisms are Hemoglobin and Myoglobin, proteins which bind to molecular oxygen and act to transport the species to different tissues of interest (Hemoglobin) or within the tissue of interest (Myoglobin); these proteins may not be considered enzymes, but are a good example of allosteric mechanisms. The application of allostery is very important with these proteins; if allosteric mechanisms did not exist, these proteins would not function as they do and hence, we would (nor other oxygen dependent species) not exist! Further discussion concerning these proteins can be found in any Biochemistry text (i.e. Stryer, Mathews & van Holde, Devlin, etc.). Within the noncovalent spectrum are competitive inhibitor species, which bind at the substrate binding site (or bind at a site which blocks the active site), and noncompetitive inhibitory/stimulatory species, which bind at a site away from the active site which induces a conformational change in the enzyme which alters its activity. These noncovalent mechanisms are most important for regulatory enzyme systems in which control of activity has to be rapid; a fast cellular response curve. Also note that the irreversible inhibitors, those that covalently bind to enzymes, are considered toxic and are poisonous; this is the focus of chemical warfare and many biologically mediated protection mechanisms (i.e. Amazonian poison frog). For example, the nerve gases neostigmine and physostigmine inhibit acetylcholinesterase enzyme by irreversible binding, resulting in proliferative "rapid" nerve synapse firing (particularly nasty death!).

Clinically, we have taken advantage of all forms of inhibition by designing drugs that are antimetabolic; drugs which mimic the substrate and preferentially react with the enzyme, causing a buildup of nonsense product which is unusable in the metabolic cascade which follows its production. Certain human pathogens have different nutritional requirements than the host, or have different forms (isoenzymatic forms—to be discussed) of the enzyme. Indeed, sometimes, like during cancer, the only difference between host and pathogenic species is the growth rate. In all cases, the application of inhibition of specific enzymes is noted in therapy. These principles are the basis for chemotherapy in humans. Some clinical examples: sulfanilamide mimics para-aminobenzoic acid and competitively inhibits folate synthesis in pathogenic bacteria while human host metabolizes only injected folate and is unaffected; methotrexate mimics folate and competes for DHF reductase enzyme ultimately inhibiting—folate reduction and inhibiting Nucleic Acid replication in rapidly dividing cancer cells; 5-fluouracil mimics uracil and irreversibly inhibits thymidylate synthetase depleting dTMP stores in rapidly dividing cancer cells ¹.

Isoenzymes are forms of an enzyme which catalyze the same chemical reaction but have different physical and chemical properties. Changes in the plasma levels of different isoenzymes may be indicative of the pathological conditions of tissues where different isoenzymes may be localized and can be used as clinical tools in the diagnosis of certain pathologies. A typical example is lactate dehydrogenase, an enzyme which

¹Also effects normal, healthy cells but kills more cancer cells since they are proliferating at a much more rapid pace (causing an increase in cell number or cell size)!
exists in five common forms and catalyzes the final reaction of anaerobic glycolysis: the reduction of pyruvate to lactate and the concomitant oxidation of NADH to NAD⁺.

\[
\text{pyruvate} + \text{NADH} + \text{H}^+ \rightarrow \text{lactate} + \text{NAD}^+ \\
\text{lactate dehydrogenase}
\]

Each of the five LDH isozymes catalyzes the same reaction, but each has unique kinetic properties. The pattern of the five LDH isozymes in a particular tissue plays a significant role in its metabolism by determining the rate at which pyruvate is converted to lactate. Each of the five major LDH isozymes consists of four polypeptide chains. Biochemical studies have revealed that each polypeptide can be of two types, designated the M and H chains. These designations stem from the predominance of the M chain in the LDH isozymes of skeletal muscle and the predominance of the H chain in the LDH isozymes of the heart. Since there are four chains in each isozyme molecule and two possible subunits (M or H), there are five possible combinations. When all four chains are of the H type, the resultant tetramer has the formula H₄, also referred to as LDH-1. The LDH isozyme with one M chain and three H chains has the formula MH₃ and is designated LDH-2. Similarly, the LDH isozymes with the other combinations of the M and H chains have the formulas M₂H₂ (LDH-3), M₃H (LDH-4), and M₄ (LDH-5).

Each LDH isozyme has a molecular weight of about 140,000 daltons, and each polypeptide chain has a molecular weight of about 35,000 daltons. Despite their similar molecular weights, the M and H chains have very different amino acid compositions and confer different kinetic properties on the LDH molecule. The rate of conversion of pyruvate to lactate increases with the number of M chains. Thus, LDH-5 has the highest rate for converting pyruvate to lactate, while LDH-1 has the lowest rate and is even inhibited by excess pyruvate. The other LDH isozymes have rates that are proportional to the number of M chains.

The relative amounts of the five LDH isozymes in a tissue correlate well with the metabolic requirements of the tissue. In skeletal muscle, where oxygen deprivation is a common occurrence, LDH-5 is the most abundant isozyme. Since LDH-5 can rapidly convert pyruvate to lactate, the predominance of this isozyme allows glycolysis and ATP synthesis to continue in the absence of oxygen. In cardiac muscle, which is a more aerobic tissue, LDH-1 and LDH-2 predominate. Since these isozymes convert pyruvate to lactate very slowly, the pyruvate is preferentially oxidized aerobically to CO₂ and H₂O.

Also, the relative amounts of the five LDH isozymes in a tissue correlate well with the proportions of the M and H subunits in each cell. It has been shown that if the single M and H chains are mixed in specific proportions in vitro, they aggregate to form the different LDH isozymes in proportions predicted by the laws of probability. Like all proteins, the synthesis of the M and H chains is directed by genes. In cells where there is a preponderance of one of the two chains (M or H), we have an example of differential gene activity, with the genes that encode the M and H chains being transcribed at different rates. There is also evidence that during cell differentiation, different LDH isozymes are selectively degraded in certain tissues.

Assays of LDH isozymes are important not only in studies of metabolic regulation and cell differentiation but in clinical evaluations, as well. For example, normally LDH-2 is the predominant LDH isozyme in human serum, with LDH-1 being the second most abundant form. Following a heart attack, however, the damaged heart tissue releases excess LDH-1, so that LDH-1 becomes the predominant LDH isozyme in the blood. Such enzyme assays are used not only for evaluating cardiovascular disease but as an aid in the diagnosis of blood and liver disorders.

Other isoenzymes may also provide the clinician with information for analysis of a particular condition; these tissue specific tests are examined in Table III.
Enzyme Analyses and Protein Purification

Protein purification is a process of concentrating the desired species of interest. The desired protein is separated from undesired cellular proteins based on its inherent properties of stability, solubility, charge, size, shape, density, etc. Remember, any given protein can fall within the range of 10% to 10^3% of a total protein solution, so the goal is to isolate as much of the species of interest without contamination from other protein. This methodology is dependent on the constituency of the protein in solution.

"Why purify a protein?" you may be asking. Purification has many applications from research analysis thru clinical therapy. For example, one may wish to characterize an enzyme to determine its enzymatic role, its dependence on cofactors and its distress caused by inhibitors. Furthermore, one may want to determine the sequence of a protein to analyze the protein chemistry and structure for sequence analysis and probe production. Biotechnologists would gladly know the structure of a protein, and hence, reproduce that protein at a more cost effective rate with better activity characteristics. Clinically, physicians could use purified proteins, antibodies being one example, for treatments to disorders in immunotherapeutic approaches. Not only are there wide reaching applications for purification, but far more reasons for doing so.

There are six steps to protein purification. First, one must develop an assay for the protein of interest; this assay must be able to detect the specific protein and detect the activity of that protein (enzyme). This step is the most tedious and difficult to perform. Second, you must determine your source of proteins and third, you must solubilize your protein. Fourth and fifth, the protein is stabilized and isolated through repeated concentration procedures. Finally, you must make a purity determination.

In developing an assay, you are creating a scenario in which you can quantitate the activity of the enzyme species in which you are interested. Take, for example, the assay developed for β-galactosidase, an enzyme produced by the lacY gene which encodes cleavage function for the sugar lactose to its products glucose and galactose. In this assay, the chromogenic substrate, orthonitrophenyl-β-D-galactoside (ONPG) is used in place of lactose (Table 7). Upon cleavage of ONPG, the products galactose and ONP are produced; ONP is chromogenic (it produces a yellow color) which is quantitated at 420 nm by a spectrophotometer (Figure 58). In non-enzymatic proteins, the purity of the species is determined by its characteristic migration in SDS-PAGE gels, the relative quantities of other proteins (determined by densitometry), and the Western blot analysis with antibodies specific to the protein.

Your source of proteins is dependent on the amount of a given protein in your isolate of interest; some strains produce greater quantities of protein than others and mutants are the most proliferative in protein production. Protein production is also dependent on the growth conditions and stage of growth of an organism. For example, most antibiotics are produced by organisms when in their stationary phase. This is reasonable, since most antibiotics have effects on DNA replication and protein synthesis and, in stationary phase, this is when the organism producing the antibiotic is at its slowest metabolic rate. Most recently, the desired gene is cloned and overexpressed in the host which is then used as a source of the protein; this process entails sub cloning the desired gene into a plasmid and transforming the strain with the recombinant plasmid.

Once you have chosen your source, you must solubilize your protein. If the protein is extracellular in nature, the supernatant can be used as your solubilized medium. If your isolate of interest is intracellular, the cells must be broken open to extract the protein¹. Many different procedures can be implemented to break open cells. During grinding, a mortar and pestle is used to rip the cell wall; this process is similar when using a bead mill with glass beads and a vortexer. Another means of shearing is

¹Proteins which are embedded within the plasma membrane can be extracted from membranes by detergents with lipophilic chains which bind to the protein. An example of such a detergent is Triton, which surrounds the protein and digs into the lipid membrane due to its amphipathic properties.
by using a French press. In a French pressure cell, cells are forced through a small orifice at high pressure, which causes them to tear open. Other methods include ultrasonication, the use of ultrasound to disrupt cells by shearing and cavitation, and enzymatic lysis, in which lysis is mediated by lysozymal digestion of the bacterial cell wall. One commonality between all methods of solubilization is buffering of protein: when breaking open the cells, they must be kept in buffer with a pH and ionic strength common to the intracellular environ of the cell. Next, the cellular debris must be removed. This is accomplished by low and high speed centrifugation.

Now that the protein solution has been solubilized, the protein of interest is stabilized by taking advantage of certain conditions where the desired protein is more stable than the rest of the proteins in solution. This step resolves removal of the undesired proteins through stabilizing the protein of interest and denaturing and precipitating out the undesired proteins. Agents used to accomplish this phase include temperature, ionic strength of medium, pH of medium, oxidation, heavy metal concentration, protease inhibition/contamination of undesired protein and nucleic acid degredation by RNase/DNases.

The isolation and concentration procedures are based on the characteristics of solubility, size and charged of the protein. **Differential solubility (hydrophobic interaction)** involves the use of inorganic salts, like ammonium sulfate, or organic solvents, like ethanol or acetone, and takes advantage of the unique charge of each native protein. By adding salts or solvents, the ordered water molecules are pulled away from the protein leaving insoluble hydrophobic patches which will precipitate out of solution upon centrifugation. In other words, salt is added to solution until the isoelectric point of the protein is reached (when the protein has no charge). At this point, being electrically neutral, the protein does not interact with the polar species and hence, precipitates out of solution. The advantages of differential solubility are that it is inexpensive and relatively easy to do with high yields; the diadvantages are that the isolate is not very pure. Chromatography involves the separation of molecules based on their size and or charge in a flowing phase and a stationary phase. **Ion exchange chromatography** involves the separation of molecules based on charge. The column matrix is made up of an insoluble component to which charged groups have been covalently bound (exchange resin). A cation-exchange resin is a negatively charged resin which binds cationicly (+) charged species (proteins); an anion-exchange resin is a positively charged resin which binds anionicly (-) charged species (proteins). In principle, charged molecules (proteins) adsorb to the ion exchange resin and by manipulating the environment (changing the ionic charge and/or pH) with a washing solution, the species of interest is isolated. The manipulation is carried out in a stepwise gradient. **Affinity chromatography** involves separation of molecules based on their affinities for a specific ligand. This specific ligand, which is designed in vitro, is covalently bound to an insoluble matrix composed of polyacrylamide and agarose. In principle, the molecules which non-covalently bind to the ligand will be retained in the column; the unbound material will be washed through. Upon washing with eluting buffer, a solution which alters the pH, ionic strength, or increases the concentration of free ligand2, the species of interest is eluted from the column. **Simple column chromatography** involves separation of molecules based on their size. Proteins are percolated through a solid support with a buffer rinse. In the process, smaller proteins elute first, while larger proteins elute last. In a modification of this procedure, **gel permeation chromatography** utilizes an inert crosslinked polymer (dextran) which contains pores. In principle, the molecules are separated according to their ability to diffuse into the pores of the matrix. Larger molecules will move through the column without being held back by the column matrix while smaller molecules will be slowed down while moving through the column due to the diffusion in and out of the pores and hence, larger proteins elute first. In both procedures, simple column and gel-permeation chromatography

---

2By adding free ligand, the non-covalent interactions exhibited between bound ligand and protein will be competing against the unbound ligand, which is in greater concentration and hence, will out compete the bound ligand species and result in elution of the protein complexed to free ligand.
Chromatography, blue dextran dye is used, since it is eluted first in both columns and acts as a running indicator for initiating fractionating your samples. Immunoprecipitation is another method of purification by which a specific protein can be selectively precipitated from a complex protein mixture, provided that specific antibodies directed against the protein are available. Both conventional and monoclonal antibodies may be employed for immunoprecipitation. Furthermore, native-PAGE electrophoresis (discussed separately), ultracentrifugation and ultrafiltration are techniques utilized to isolate and concentrate a protein. In ultracentrifugation, the rate of sedimentation depends on the molecular weight, density, and frictional coefficient of the protein; knowing these parameters, one can utilize ultracentrifugation to isolate and characterize the molecular weight of the species (sedimentation is in Svedbergs "S"). In ultrafiltration, the species is separated according to molecular dimensions by filtering the sample across membranes with defined pore sizes; gas or centrifugal force is applied to force the solution containing the protein species through the pores.

Once the species has been isolated and concentrated, the purity of the sample must be determined. The determination of purity is only as good as the criteria used for its establishment and hence, many methods are used. One method is electrophoresis (native, SDS-PAGE, isoelectric focusing, 2-Dimensional) and Western blotting (discussed separately). Another process is by crystallization, or homogeneity of sample. Still another is by analysis of specific activity of protein (enzymes), in which an assay is performed to determine the activity of the sample in the presence of a substrate which is converted to an observable product. Finally, GC (Gas Chromatography) or HPLC (High Performance/Pressure Liquid Chromatography) can be performed, in which the sample is injected into a heated, pressurized port which acts to separate proteins based on physical-chemistry characteristics. The HPLC is the method of choice (if you have an easy $100,000 to spend on a decent setup), since it not only determines purity of sample, but also separates samples and concentrates them for the investigator at a very fast pace with minimal effort for the investigator (you just inject your sample with a syringe into the entry port on the machine and it does the rest!). There are four HPLC methods available to the researcher; reversed phase, ion exchange, size exclusion and hydrophobic interaction HPLC. Each of these are equivalent to their counterparts in conventional chromatography with the following advantages: small amounts (≤1 nmol) of a given protein can be handled and purified, less time is required for a separation (≤3 hr), and smaller chromatographic peak volume is achieved. Reversed-phase HPLC is useful primarily for purifying small proteins (molecular weight <20,000 Da) and protein fragments.
Quantitation of proteins

How much Protein?
1. Colorimetric methods:
   a. Bradford method
   b. Lowry method
   c. Spectroscopy at 280nm
   d. Bio-Rad assay
   e. Pierce assay
2. Electrophoretic methods:
   a. Types (1 dimensional):
      1. SDS-PAGE
      2. Native-PAGE
      3. Isoelectric focusing
      4. Discontinuous types (variable)
      5. 2-Dimensional types (variable)
   b. Detection on gel:
      1. Coomassie blue stain
      2. Silver stain
   c. Electroblootling from gel to membrane:
      1. Membrane blotted to:
         a. Nitrocellulose
         b. Nylon
      2. Detection on membrane:
         a. India ink
         b. Iron stain
         c. Gold stain
         d. Antibodies (Western blot)

Isolation of proteins and polypeptide fragments

How is a lirotein or liolypeptide fragment isolated?
1. Electrophoresis:
   a. electroelution from gel
   b. electroblootling from gel to membrane; wash species off membrane
2. Chromatography:
   a. Conventional:
      1. gel permeation
      2. ion exchange
      3. hydrophobic interaction
      4. affinity
      5. simple column
   b. HPLC:
      1. reversed phase
      2. ion exchange
      3. size exclusion
      4. hydrophobic interaction
      5. Analysis of protein purity, molecular weight and subunit structure

Is the protein pure? How many protein subunits? What is the M.W. of the protein/its subunits?
1. Electrophoresis:
   a. comparison with molecular sizes for standard proteins by densitometry after staining
2. Chromatography:
   a. direct estimation of molecular weight

\footnote{For a protein without subunits or with identical subunits, detection of a single protein band after 1-D gel electrophoresis under denaturing conditions or a single spot (revealed by silver nitrate stain) after 2-D gel electrophoresis indicates that the protein is pure. If the protein consists of multiple subunits of different molecular sizes, purity is confirmed by detecting a single, stainable band after gel electrophoresis under non-denaturing conditions.}
**Separation of Proteins by Electrophoresis**

Electrophoresis is the movement of charged molecules under the influence of an electric field. In the most common form of electrophoresis, the sample is applied to a stabilizing medium which serves as a matrix for the buffer in which the sample molecules travel. Both agarose and polyacrylamide gels can be used as stabilizing media for the electrophoretic separation of proteins on the basis of charge or size. The following describes each process in theory.

All polypeptide chains contain at least two ionizable groups: the amino and carboxyl groups at their termini. In addition, the R-groups of most amino acid residues can be charged. These charges are responsible for the migration of proteins in an electric field. At high pH, the carboxyl (-COO⁻) groups are negatively charged while the amino groups (−NH₂) are not charged. At low pH, acidic groups are uncharged (−COOH) and basic groups (−NH₃⁺) are positively charged. Thus, there must be an intermediate pH at which the protein bears no net charge and does not migrate in an electric field. The pH at which an amino acid or protein does not migrate in an electric field is called the isoelectric point (pI); this property of proteins is manipulated in isoelectric focusing, in which the gel medium is a pH gradient (i.e. pH 11 to 2) where proteins lack migration at their isoelectric point. At a pH above the isoelectric point, a protein is negatively charged and when applied to sample wells at the negative electrode end of the gel, will travel towards the positive electrode. The rate of migration of a protein species in an electric field depends upon its charge density (the ratio of charge to mass); the higher the charge density, the faster the protein will travel. For example, serum albumin which has an isoelectric point of 4.7, will carry a strong negative charge in a buffer of pH 8.6 as compared to gamma globulin which has an isoelectric point of 7.2. Therefore, at pH 8.6 albumin will migrate toward the positive electrode at a much faster rate than gamma globulin. The agarose gel is commonly used in the charge separation of proteins since low percentage gels form a sponge-like network which serves as a medium for the buffer but which has pores large enough to allow even the largest proteins to pass unimpeded.

Electrophoretic separation of proteins by molecular size requires several modifications in the basic procedure described above. First, proteins are covered with the anionic detergent sodium dodecyl sulfate (SDS), which masks their native charges with its own negative charges. Proteins which contain disulfide bonds must also be treated with reducing agents to cleave these bonds so that polypeptide chains dissociate from each other and unfold. Proteins that have been treated with SDS and a reducing agent assume a rod like structure and carry the same charge density imposed by anionic SDS. Since such proteins no longer possess their native shape and charge, they are referred to as denatured proteins. The final modification involves a reduction in the pore size of the agarose gel so that the gel matrix now serves as a molecular sieve. Thus, electrophoretic separation of denatured proteins sorts them according to size since it relies on the ability of uniformly charged proteins to fit through the pores of gel matrix; this procedure is termed **SDS-polyacrylamide gel electrophoresis**. In a similar procedure, **native polyacrylamide gel electrophoresis (Native-PAGE)**, proteins move in the electric field according to their native charge and size; since these proteins are not acted upon by SDS or a reducing agent, they retain their native conformation and charge. Remember, native page retains 4° structure, whereas SDS-PAGE destroys 4° structure (separates polypeptides from each other) and 3° structure (polypeptides lose their 3 dimensional conformation and become linearized). Hence, native-PAGE enables the investigator to perform enzyme activity assays since the protein is in its native (physiologically natural) form. Furthermore, both PAGE systems utilize different types of gel medium. Resolving gels, poured first and consuming 3/4 of the gel, has a smaller pore size (10-15%) and a higher pH (9.0); stacking gels, poured last and consuming 1/4 of the gel (by the end containing the sample wells), has a larger pore size (2.5%) and a smaller pH (7.0). This process, termed **discontinuous (native/SDS PAGE)**, allows the investigator to selectively screen larger or smaller proteins of interest by using the stacking gel and resolving gel. Consider again the electrophoresis of albumin and gamma globulin. Albumin consists of a single polypeptide of molecular weight 66,000
daltons while gamma globulin is a multipeptide protein containing two 53,000 dalton polypeptides and two 23,000 dalton polypeptides. Native gamma globulin migrates toward the positive electrode slower than albumin at pH 8.6 since it carries a weaker negative charge. When it is denatured, however, gamma globin has the same charge density as albumin and its polypeptide chains dissociate from each other. Thus, electrophoretic migration will only be affected by the sieving property of the gel matrix and the smaller, 23,000 dalton gamma globulin polypeptide will move the fastest toward the positive electrode, followed by the 53,000 dalton gamma globulin polypeptide and then by albumin.

The vertical electrophoresis unit is composed of an acrylic cell with central platform, platinum electrodes, a central gel-casting tray, a sample well-forming comb and a safety lid with power cords. The gel is made in the casting tray on the central platform of the electrophoresis cell. Each gel contains 6 or 12 separate sample wells depending on the sample well-forming comb. The Horizontal Electrophoresis unit is composed of an acrylic cell with central platform, platinum electrodes, a removable gel-casting tray, a sample well-forming comb and a safety lid with power cords. The gel is made in the casting tray and then placed on the central platform of the electrophoresis cell. Each gel contains 6 or 12 separate sample wells depending on the sample well-forming comb.

The power supply is a general purpose electrophoresis power source which produces a constant voltage output; the range of output is 20 to 170 volts. Voltage selection is controlled by the switch located in the center of the front panel. The ammeter, also located on the front panel, permits the current to be monitored during an electrophoretic run. The unit can reach a maximum of 500mA.

A number of different types of stabilizing supports have been used in the electrophoretic separation of proteins. These include filter paper cellulose acetate and gels composed of either starch, polyacrylamide, agar or agarose. The agarose gel is an ideal solid support for the separation of proteins on the basis of charge and the polyacrylamide gel is generally used for the separation of proteins on the basis of size. However, the preparation of high percentage (5-6%) agarose gels extends the utility of this non-toxic support to SDS electrophoresis applications as well. These high percentage agarose gels have small pores, and thus, can be used to separate SDS treated proteins on the basis of size. Agarose is a natural polysaccharide of galactose and 3,6-anhydrogalactose derived from agar, which in turn is obtained from certain marine red algae. Agarose gels are frequently run in the "submarine mode where the gel is completely immersed in buffer. This feature reduces heat development in the gel which could otherwise lead to protein band distortion.

The separation of proteins on the basis of molecular weight is highly dependent on their denaturation with the anionic detergent sodium dodecyl sulfate (SDS). Furthermore, the pH of the buffer is also important in electrophoretic separation, since it will influence the net charge of the SDS-denatured protein. Thus, the pH and SDS concentration in the electrophoresis chamber buffer (Tris-Acetate-SDS pH 8.3), and gel buffer (Tris-Borate pH 8.6) must be carefully regulated. The ionic strength of the buffer is also important in electrophoresis. High ionic strength buffers permit fast migration and can promote the sharpening of protein zones. However, high ionic strength buffers may also cause high heat production which can lead to band distortion. Moderate ionic strength buffer permits optimal resolution of protein bands in minimal time.

In addition to SDS denaturation, protein samples must be treated with an agent effective at eliminating disulfide bonds prior to electrophoretic separation. Thus the sample buffer incorporates the reducing agent, β-mercaptoethanol as well as sodium dodecyl sulfate. The protein sample buffer used in all the experiments described below also contains 10 to 20% glycerol to ensure that the samples will layer smoothly at the bottom of the sample wells. The tracking dye, bromphenol blue, is also present in the sample buffer to enable the investigator to follow the progress of an electrophoretic run.

As an adjunct to this discussion, one must realize that proteins can be run in a gel in 1 dimension or 2 dimensions. One dimension refers to the migration of proteins in a gel from one side of the gel to the other, two dimensions refers to the migration of proteins in 1 dimension, rotation of the gel in the electric field (90°), and running the gel in this direction to completion. Furthermore, the running (agar) medium is manipulated
such that the first direction is a different constituency from the second (for example, the medium is an isoelectric focusing gel in one direction and SDS-PAGE in another direction). An example of a two dimensional electrophoresis is provided in Figure 47.

Most proteins are not colored, and therefore it is necessary to visualize them in some way in order to determine their position in the agarose gel after electrophoresis. The most commonly used stain for the detection of proteins is Coomassie blue. The staining solution also contains acetic acid which serves to precipitate and immobilize the proteins in the gel so that the protein bands do not become blurred by diffusion. After the proteins in the gel have stained, the unbound dye must be rinsed from the gel by a process known as destaining. A dilute solution of acetic acid and methanol is often used for the destaining of the agarose gels.
Protein Evolution

The theory of organic evolution is based upon the belief that present-day organisms have descended with modifications from forms of life that lived in the past. Phylogeny is the evolutionary history of a species, genus or larger group and taxonomy is the science of classification of organisms according to the degree of kinship and evolutionary relationships. The doctrine of organic evolution is one of the most important generalizations in science. It is supported by evidence drawn from genetics, paleontology and geographical distribution and from comparative anatomy and embryology. Results from the modern biochemistry laboratory have also provided strong evidence for the doctrine of organic evolution and have suggested possible mechanisms by which evolutionary changes occur.

Each protein carries in its amino acid sequence information pertaining to its own evolutionary history and origin, and clues to the evolutionary history of the organism in which it is found. Indeed, the millions of proteins existing today are in effect living fossils. Two of the methods that have been used to study the evolution of proteins at the molecular level are discussed below.

A comparison of the amino acid sequence of the same protein in different organisms has provided a direct way to study molecular evolution. A comparison of the amino acid sequence of cytochrome C from over 80 species has revealed that the amino acid sequence of this protein from different species varies and the degree of variation corresponds to the distance that separates the two species on the evolutionary tree. That is, the greater the taxonomic difference, the more the cytochromes are likely to differ in their order of amino acid residues. For example, the cytochrome C in man differs from the cytochrome C found in yeast in 44 out of the 104 amino acid residues. The numbers of amino acid replacements in cytochrome C of 12 species are compared in table 4. This type of information has led to the construction of family trees of organisms that agree remarkably well with those obtained from the classical anatomical record. In fact, on a number of occasions, comparative protein studies have been used to clarify and expand on phylogenetic relationships that were derived from classical analysis.

Immunology & the Western blot

The immune system consists of a diverse set of cells, tissues and organs and about $10^{12}$ antibody molecules. The major function of the immune system is to protect the organism from viruses, bacteria, protozoans and larger parasites. When these organisms enter the vertebrate body, macromolecules on their surfaces induce the production of specific antibodies that appear in the serum of the infected animal. The antibodies, in turn, combine with these foreign macromolecules thereby rendering the invading organisms inactive and noninfective. Central to this protective mechanism is the ability of the immune system to distinguish between foreign macromolecules and those that are a natural part of the individual's body.

The macromolecules that elicit antibody production are called antigens and are most often proteinaceous in nature. Although antigens are frequently components of foreign organisms, purified foreign proteins will serve as antigens in that they will stimulate the formation of antibodies when injected into a suitable test animal such as a rabbit. Each antigen possess features that are recognized by the antibody and these features constitute the antigenic determinants or epitopes. An antigenic determinant recognized by an antibody molecule may be a unique shape or a sequence of about 5 to 10 amino acid residues on the protein molecule. It follows that each protein possess a large number of potential antigenic determinants and when a foreign protein is injected into an animal, antibodies to different determinants on the protein are produced and appear in the serum. This antibody containing serum is called an antiserum. Thus, an antiserum generated by immunization with one purified protein usually contains a large number of different antibodies which recognize different determinants along the protein.
Antibodies are frequently used to study evolutionary relationships because they recognize unique antigenic determinants along a protein molecule. For example, human serum albumin injected into a rabbit makes the animal produce anti-human albumin antibodies. These antibodies are directed against determinants of human albumin (shapes and/or amino acid sequence) that are not found on the rabbit albumin molecule. When the anti-albumin serum is mixed with human albumin, the antibodies in the serum react strongly with the albumin and an antibody-antigen complex is formed. Similarly, the antisera will react strongly with albumin isolated from gorilla, orangutan and baboon. In contrast, the anti-human antisera will react only weakly with albumin from deer, horse and ox and not at all with chicken or frog albumin!

The first step in Western Blotting is to electrophoretically separate the serum protein samples in agarose gels. Next, the gel is placed against a specialized membrane made of nitrocellulose and incubated overnight. During the incubation, the proteins diffuse out of the gel and are trapped on the nitrocellulose membrane. As a result, a replica (blot) of the electrophoretically separated proteins is produced on the nitrocellulose membrane (Figure 44).

The next step is to incubate the membrane with antibodies which react with the proteins trapped on the membrane. Most experiments use antibodies that were generated in rabbits against human serum proteins. These antibodies will bind to the membrane-trapped human serum proteins and to the serum proteins from the other species that are structurally related to the human proteins. Gelatin is added to the membrane before and during antibody incubation in order to minimize non-specific protein-antibody interaction.

Since the antigen-antibody complexes are not colored, they must be treated in some way in order to visualize them. A commonly used method involves the coupling of a colored molecule to a second antibody. The second antibody was generated by immunizing animals with the primary, antigen-reactive antibody. The advantage of using a second antibody instead of just coloring the first antibody is two-fold. First, since the primary antibody contains several epitope sites, several second antibody molecules can bind to each primary antibody. If the second antibody is color treated and then reacted with the primary antibody-antigen complex, then for every antigen site present, two to ten color molecules will be bound as compared to only one if a colored primary antibody were used. Second, it is more convenient to color one second antibody that can then be used with all primary antibodies produced in the same species than it is to individually color each primary antibody that reacts with a different antigen.

Enzymes which catalyze the reaction of soluble, colorless substances to insoluble, colored products are often coupled to second antibodies to permit visualization. For example, the second antibody described above is coupled to Horseradish Peroxidase (HRP) which catalyzes the reaction diagrammed in Figure 13. This antibody, which was generated in goats by immunizing with rabbit antibodies, binds to the primary, antigen binding rabbit antibody. Following incubation of the membrane with the HRP-coupled second antibody, the final step in the Western Blotting procedure is to incubate the blot in a color development solution containing hydrogen peroxide and 4-chloro-l-naphthol. The immobilized HRP then converts the 4-chloro-l-naphthol to an insoluble purple product which is deposited at the site of the antigen bands and allows antigen visualization.
**Protein Measurements**

There are many useful methods for estimating the concentration of proteins: the Lowry protein assay, the Bradford (or Coomassie Blue Dye Binding) assay, the A$_{280\text{ nm}}$ assay, the Bio-Rad (Rapid) assay and the Pierce Protein assay. These assays are useful for estimating the concentration of proteins; they do not the serve a purpose for detecting the quality, or activity, or proteins. The qualitative assays utilize a metabolizable substrate, which is converted to a detectable product in solution. Please do not confuse between qualitative and quantitative protein assay procedures.

Protein in many biological materials can be measured using a procedure described by Lowry et al (1951, *J Biol Chem*, 193:265). This Lowry assay procedure involves the solubilization of protein from cells in dilute alkali and an estimation of the protein by a modification of the procedure for measuring aromatic amino acids by the Folin-Ciocalteau phenol reagent (Folin and Ciocalteau, 1927, *J Biol Chem*, 73:627). In the presence of copper in alkaline solutions, the color due to the reduction of the phenol reagent is increased 10-15 times. Bovine serum albumin is used as a standard. It should be kept in mind that different proteins give a different response (more or less color per unit weight of protein). This procedure is many times more sensitive than other methods including the biuret reaction, phenol reagent and the use of ninhydrin, and is usually the method of choice for estimating the concentration of impure proteins. As little as 0.2 ug of protein can be measured by this method.

A more rapid assay is the Bradford assay (also known as the Coomassie Blue Dye-binding assay) as described by Bradford. This assay involves the binding of proteins to Coomassie blue which results in an absorption shift of the dye from approximately 465 nm to 595 nm (dependent on dye lot). This method is very sensitive and much more rapid than the Lowry assay. However, it is much more dependent on the use of an appropriate standard.

Probably the most simplistic assay, the Bio-Rad Protein Assay, is based on the observation that the absorbance maximum for an acidic solution of Coomassie Brilliant Blue G-250 shifts from 465 nm to 595 nm when binding to protein occurs. Bradford first demonstrated the usefulness of this principle in protein assay. Spector found that the extinction coefficient of a dye-albumin complex solution was constant over a 10-fold concentration range. Thus Beer’s Law may be applied for accurate quantitation of protein by selecting an appropriate ratio of dye volume to sample concentration. Over a broader range of protein concentrations, the dye-binding method gives an accurate, but not entirely linear response. The Bio-Rad Protein Assay is useful for several reasons. First, it is much easier to use. It requires one reagent and five minutes to perform as compared to three reagents and 30 to 40 minutes typical for the Lowry assay. Second, because the absorbance of the dye-protein complex is relatively stable, the Bio-Rad assay does not require the critical timing necessary for the Lowry. Third, the Bio-Rad assay is free from most of the interferences which limit the application of the Lowry assay.

The Pierce BCA Protein Assay is the most sensitive method for the spectrophotometric determination of protein concentration. This unique reagent system combines the well known reaction of protein with Cu$^{2+}$ in an alkaline medium (yielding Cu$^{+}$ with a highly sensitive and selective detection reagent for Cu$^{+}$, namely bicinchoninic acid). The BCA Protein Assay offers the researcher flexible and easy assay procedures which eliminate the precisely timed reagent additions and vortexing necessary with the Lowry method. Additional advantages include compatibility with ionic and non-ionic detergents, a stable working reagent, less protein-to-protein variation than with other methods, broad linear working ranges with excellent sensitivity and the ability to change the protocol which provides flexibility. In principle, bicinchoninic acid (BCA), in the form of its water-soluble sodium salt, is a sensitive, stable and highly specific reagent for the cuprous ion (Cu$^{+}$). Macromolecular structure and four peptides (cysteine, cystine, tryptophan, tyrosine) have been reported to be responsible for color formation in protein samples when assayed with BCA.

The reaction scheme (Figure 68) shows how the BCA Protein Assay combines the well-known biuret reaction (protein reducing Cu$^{2+}$ in an alkaline medium to produce Cu$^{+}$) with the
The purple reaction product, formed by the interaction of two molecules of BCA with one cuprous ion (Cu$^{+}$), is water soluble and exhibits a strong absorbance at 562 nm. This allows the spectrophotometric quantitation of protein in aqueous solutions.

Finally, for pure proteins in solution (i.e. solutions free of nucleic acids and other materials that absorb ultraviolet light), measurement of the absorption of 280 nm light may be used. The absorption at 280 nm is more convenient and is non-destructive, but this method is less sensitive.
Protein Sequencing

There are many procedures available for determining the constituency of a protein (Moore and Stein) or the short amino acid sequences of the amino terminus (Sanger), but neither of these procedures are as effective as the Edman method of sequencing. Edman devised a method for labeling the amino-terminal residue and cleaving it from the peptide without disrupting the peptide bonds between the other amino acid residues. The Edman degradation sequentially removes one residue at a time from the amino end of a peptide (Figure 45). Phenyl isothiocyanate reacts with the uncharged terminal amino group of the peptide to form a phenylthiocarbamoyl derivative. Then, under mildly acidic conditions, a cyclic derivative of the terminal amino acid is liberated, which leaves an intact peptide shortened by one amino acid. The cyclic compound is a phenylthiohydantoin (PTH) amino acid, which can be identified by chromatographic procedures. Furthermore, the Edman procedure can then be repeated on the shortened peptide.

Analyses of protein structures have been markedly accelerated by the development of sequenators, which are automated instruments for the determination of amino acid sequence. In a liquid-phase sequenator, a thin film of protein in a spinning cylindrical cup is subjected to the Edman degradation. The reagents and extracting solvents are passed over the immobilized film of protein, and the released PTH-amino acid is identified by high-performance liquid chromatography (HPLC). One cycle of the Edman degradation is carried out in less than two hours. By repeated degradations, the amino acid sequence of some fifty residues in a protein can be determined. A recently devised gas-sequenator can analyze picomole quantities of peptides and proteins. This high sensitivity makes it feasible to analyze the sequence of a protein sample eluted from a single band of an SDS-polyacrylamide gel.

Peptides much longer than about fifty residues cannot be reliably sequenced by the Edman method because not quite all peptides in the reaction mixture release the amino acid derivative at each step. If the efficiency of release of each round were 98%, the proportion of "correct" amino acid released after sixty rounds would be only 0.3 (0.98^60) -a hopelessly impure mix. This obstacle can be circumvented by specifically cleaving a protein into peptides not much longer than fifty residues.

Specific cleavage can be achieved by chemical or enzymatic methods (Table 5). The peptides obtained by specific chemical or enzymatic cleavage are separated by chromatography. 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. The necessary additional information is obtained from overlap peptides (Figure 46). An enzyme different from trypsin is used to split the polypeptide chain at other linkages. For example, chymotrypsin cleaves preferentially on the carboxyl side of aromatic and some other bulky nonpolar residues. Because these chymotryptic peptides overlap two or more tryptic peptides, they can be used to establish the order of the peptides. The entire amino acid sequence of the polypeptide chain is then known.

These methods apply to a protein consisting of a single polypeptide chain devoid of disulfide bonds. Additional steps are necessary if a protein has disulfide bonds or more than one chain. For a protein made up of two or more polypeptide chains held together by noncovalent bonds, denaturing agents, such as urea or guanidine hydrochloride, are used to dissociate the chains. The dissociated chains must be separated before sequence determination can begin. Polypeptide chains linked by disulfide bonds are first separated by reduction with β-mercaptoethanol or dithiothreitol. To prevent the cysteine residues from recombining, they are then alkylated with iodoacetate to form stable S-carboxymethyl derivatives. Distinct polypeptides are then separated by chromatography according to size (or other features which distinguish between the polypeptide types).

Many proteins are modified after synthesis; these changes are termed post-translational modification. Amino acid sequences derived from DNA sequences are rich in information, but they do not disclose posttranslational modifications. Chemical analyses of proteins themselves are needed to delineate the nature of these changes! Hence, DNA and protein sequencing complement each other to elucidate protein function in structure.