1

Basic Concepts of Protein Purification and Characterization

INTRODUCTION

Before examining the ways in which proteins may be purified, two points must first be covered that are linked in the concept of a "specific activity." Specific activity is defined and discussed in more detail later, but for the moment it can be regarded as activity per unit amount of protein. Of paramount importance therefore in the determination of a specific activity are the experimental determination of the concentration and the activity of a protein. Each of these is considered before discussing specific activity in more detail and examining sources of protein to be purified. It is the goal of purification to obtain protein having the maximum specific activity, and achievement of this is often monitored by obtaining a constant specific activity during several purification steps.

DETERMINATION OF PROTEIN CONCENTRATIONS

The accurate quantitation of protein concentrations is, of course, increasingly important as purity is attained since the accuracy of the final specific activity of a "pure" preparation of protein depends in large measure on the accuracy of its concentration determination. This determination would appear to be a simple operation; however, many of the methods that have been used suffer from interference by reagents such as the detergents or salts that are commonly used during purifications, or from contaminants such as the nucleic acids or nucleotides that are frequently associated with proteins. Much effort has been directed toward developing approaches to circumvent such problems.

Here we consider the pitfalls and advantages of several commonly used techniques for determining protein concentration. These are the biuret protein determination method, the Lowry method, the Bradford method, the Amidoschwarz method, and the 280-nm absorbance method.

Biuret Method

The biuret method is one of the earliest colorimetric methods for determining protein concentrations and is still a rapid, if rather insensitive approach for use during the early stages of a purification, especially when ammonium sulfate precipitation (Chap. 2) is employed since it is unaffected by the presence of this salt. In this method a colored complex involving the complexation of copper in alkaline solution with peptide bonds and tyrosine side chains is used. As with many of the methods discussed here, a calibration curve with known concentrations is used to allow the concentration of the unknown to be determined. This aspect leads to many of the inaccuracies of such methods since the choice of standard protein affects the calibration curve. In the biuret method the use of two different standard proteins having quite different tyrosine contents yields different calibration curves. The nature of the standard protein should always be indicated.

Lowry Method

The approach first described by Lowry et al. in 1951 is also based on the formation of a copper–protein complex under alkaline conditions that subsequently reduce the folin–phenol reagent at pH 10 with color development that can be quantitated spectrophotometrically at 750 nm. Protein concentration of an unknown sample is determined by comparison with a standard curve. The folin–phenol reagent is a phosphomolybdate-containing mix, and maximum color development usually takes about 30 minutes. A fundamental problem is that the color developed depends on the nature of the protein—not all proteins give the same color intensity. Some of the problems encountered using nonionic and cationic detergents (which under Lowry conditions tend to cause precipitate formation) have been overcome by the inclusion of 0.5% sodium dodecylsulfate in the alkali reagent, which prevents precipitate formation and does not affect color development. This adaptation is particularly useful in the estimation of protein concentrations in detergent solubilized membrane preparations.

Bradford Method

The observation that the dye Coomassie Brilliant Blue G-250 exists in two different color forms (a red and a blue) and that the red form is converted to the blue form upon binding to protein, led to the development of the Bradford dye binding method. The binding of the dye to protein causes a shift in the absorption maximum of the dye from 465 nm to 595 nm, and the resultant increase in absorption at 595 nm is monitored. The dye–protein complex has a high extinction coefficient,

leading to excellent sensitivity, and the colored complex develops rapidly and has relatively good stability. Although the color intensity is quite pH dependent, falling off considerably at high pH, appropriate buffering leads to accurate protein estimation. Although cations and carbohydrates do not significantly interfere with the color, various detergents do cause problems. If the concentration of detergent is small, the interference is minimized by the inclusion of the same amount of detergent in the calibration samples. For the most accurate determinations, the buffers for the sample and the standard curve should be the same.

Because of its ease of use, its high sensitivity (the accurate determination of 0.1 μg of protein is possible), and its overcoming of many of the Lowry method problems, the Bradford method of protein estimation is often the method of choice.

For both the Lowry and Bradford methods we have indicated that a variety of small molecules interfere with color formation. This can lead to quite large errors in protein concentration determination and often, compounds that decrease the sensitivity of the Bradford assay (e.g., CsCl and guanidine HCl) increase the sensitivity of the Lowry assay. Other compounds, such as Tris and dithiothreitol, have a significant effect on the Lowry but not on the Bradford method.

Some of these problems involve the effects of solvent composition on the color of the various noncovalent complexes involved in color formation. Such effects can be overcome by employing covalent labeling with a fluorescent reagent such as *O*-phthalaldehyde, which reacts with amines to give an adduct that fluoresces under alkaline conditions (pH 9 to 11). The derivative is essentially insensitive to all common buffer components except Tris.

Amidoschwarz Method

The Amidoschwarz approach is particularly useful since it is not affected by the majority of reagents that interfere with the Lowry and the Bradford methods. Its basis is simple. Protein is precipitated by trichloroacetic acid, the precipitate collected by filtration onto filter paper and stained with Amidoschwarz 10B dye, and after suitable washing to remove excess dye the protein–dye complex is eluted and quantitated by absorbance measurements at 630 nm. As with the Lowry and Bradford methods, the protein concentration is estimated by comparison with a standard curve, but because of the precipitation step prior to staining, the color development is virtually independent of buffer salts, detergents, and so on, which interfere with the previously discussed methods. Furthermore, because quite large volumes can be used, concentrations as low as 0.5 μg/ml can be conveniently determined. The major practical drawback of this method is that it is somewhat tedious; however, under circumstances that interfere with the Lowry and Bradford assays, this method represents a convenient way to estimate protein concentrations accurately. The only condition so far shown to lead to problems is the estimation of low-molecular-weight proteins where the precipitates are incompletely retained during the filtration step. This represents a problem with proteins such as insulin (mol. wt. 5770) and α-lactalbumin (mol. wt. 14,000).

280-nm Absorbance Method

Most proteins contain one or more aromatic amino acids (tyrosine, tryptophan, and phenylalanine) that absorb light in the region 250 to 300 nm. The absorption spectrum of most proteins features a broad absorption usually centered around 275 to 285 nm, and spectrophotometric measurement at 280 nm is a frequently used method of estimating protein concentration. As will be discussed in more detail, a pure protein has a characteristic and defined extinction coefficient at 280 nm that can, in pure solutions, be used to quantitate concentration very accurately based on spectrophotometric measurements. However, the extinction coefficient at 280 nm varies from approximately 0.5 to 1.5 for a 1-mg/ml solution, making it difficult to obtain accurate estimates of total protein concentration for a mixture of proteins (which, of course, is the normal situation encountered during a purification). However an "average" value of the extinction coefficient of 1.0 for a 1-mg/ml solution is frequently employed for the purposes of estimating protein concentrations.

The prime advantage of absorbance measurements is their ease, but unfortunately, proteins are not the only species in solution that absorb at 280 nm. Particularly troublesome are nucleic acids, nucleotides, and detergents. Various detergents used in protein solubilization absorb at 280 nm, frequently giving quite high background absorbances and making accurate protein determinations difficult. Although the use of appropriate blanks can often alleviate detergent problems, nucleic acids and nucleotides represent a more difficult situation. Nucleotides absorb strongly at 260 nm with resultant absorbance at 280 nm from the tail of their spectrum. The average protein has a 280:260 absorbance ratio of approximately 2.0, and the 280:260 ratio of a protein containing solution is frequently used to give some estimate of possible contamination by nucleotides. In fact, this approach is often used with pure proteins that bind nucleotides tightly, as a way of estimating how much nucleotide is bound per protein molecule.

DETERMINATION OF EXTINCTION COEFFICIENTS OF PURE PROTEINS

Clearly, for the quantitation of a pure protein solution by absorbance measurements at 280 nm, its extinction coefficient must be known. There are two commonly used methods to determine this for a protein: the dry weight method and the amino acid analysis method. In the first, the absorbance of a protein containing solution is first determined after thorough dialysis versus a buffer of known composition that is not volatile, to give an absorbance per milliliter. Next, volumes of protein containing solution and dialysate are evaporated to dryness, heated at 110°C for 24 hours and cooled in a vacuum desiccator over phosphorus pentoxide, weighed, and the heating and weighing cycle repeated until a constant weight is obtained. The weight of protein in the original solution is thus obtained, allowing direct calculation of the extinction coefficient.

Alternatively, from the known molecular weight of the protein and amino acid analysis, a theoretical extinction coefficient can be calculated from the extinction coefficients of tyrosine, tryptophan, and phenylalanine. Usually, extinction coefficients are somewhat lower than those from dry weight determinations, probably because of environmental effects on the absorption properties of the aromatic amino acids.

Because of the experimental vagaries of either of these methods, they are best used in conjunction with some method of active site titration, discussed in more detail in Chap. 4. Where an accurate molecular weight can be obtained, active site titration, in conjunction with absorbance measurements, gives increased reliability of extinction coefficient determinations.

Amino Acid Analysis

The amino acid analysis of a polypeptide chain is achieved through some type of hydrolysis followed by analysis of the resultant amino acids using either ion-exchange chromatography or, for much greater sensitivity, high-performance liquid chromatography (HPLC).

There are two basic approaches to the hydrolysis procedure: (1) chemical hydrolysis, and (2) enzymatic hydrolysis, both of which suffer from drawbacks. With chemical hydrolysis it is the fact that the usually harsh hydrolysis conditions lead to destruction or chemical modification of some of the amino acids. For example, under typical acid hydrolysis conditions (5.7 *M* HCl, 108°C for 22 hours), the amides (asparagine and glutamine) are hydrolyzed to aspartate and glutamate, up to 10% losses of threonine or serine occur, there are variable losses of tyrosine or tryptophan (protective agents are usually added to help prevent these), and cysteine and cystine may be interconverted, oxidized to cystine oxides, or in the case of cysteine, react with tryptophan.

Early procedures for amino acid analysis employed multiple time points of hydrolysis and specific types of hydrolysis to optimize determination of certain amino acids. However, a general procedure has evolved that allows reasonable amino acid analysis to be made from a single hydrolysis of a protein. This approach is comprised of the following stages: (1) alkylation of cysteine, (2) acid hydrolysis in the presence of an amino acid protectant, and (3) neutralization and formation of *S*-sulfocysteine prior to analysis.

Alkylation of Cysteine. The two most commonly used methods of alkylation for amino acid analysis involve iodoacetate (a) or 4-vinyl-pyridine (b).

(a) $P—CH_2—SH + ICH_2—COO^- \longrightarrow P—CH_2—S—CH_2—COO^- + HI$

(b) $P—CH_2—SH + CH_2—CH\theta(N) \longrightarrow P—CH_2—S—CH_2—CH_2\theta(N)$

These yield derivatives that are fairly stable and more polar than cysteine (which is usually eluted with proline): *S*-carboxymethylcysteine is more acid than cysteine and elutes before aspartate, while *S*-β-(4 pyridylethyl) cysteine is more basic than cysteine and elutes after ammonia.

Alkylations are usually performed with a 10-to 20-fold excess of alkylating reagent over thiol at alkaline pH values. The reactions can be done in the presence of up to 8 *M* urea or 5 *M* guanidine hydrochloride, which is subsequently removed by dialysis or gel filtration.

Hydrolysis. Although a variety of hydrolysis conditions have been used, one of the best is 4 *M* methanesulfonic acid at 115°C for 22 hours. Here the addition of protective reagents for unstable amino acids is essential, with the most common additives being tryptamine (to protect tryptophan) and phenol (to protect tyrosine). One of the major problems with hydrolysis conditions selected to minimize breakdown of products is that these "mild" conditions may be insufficent to achieve complete cleavage of more resistant peptide bonds, such as the hindered ones of Ile-Ile, Ile-Val, or Val-Val, where time periods of up to 96 hours may be necessary to achieve complete hydrolysis.

Glycoproteins may be of particular trouble in hydrolysis since some amino acids are destroyed by hydrolysis in the presence of large amounts of carbohydrate. While protective agents such as tryptamine do work, there is a correlation between yield and percent carbohydrate—a plot of log (tryptophan recovery) versus percent carbohydrate gives a straight-line relationship.

As one may gather, the estimation of tryptophan content is most problematical in acid hydrolysis and most subject to error. For the most accurate estimation of tryptophan from protein hydrolysis, alkaline hydrolysis rather than acid hydrolysis gives the best results. It is carried out in vacuo at either 110°C or 136°C in 4.2 *N* NaOH containing 40 mg/ml starch. Tryptophan is then estimated by ion-exchange chromatography at pH 5.4 to separate it from a lysine derivative *N*-(DL-2-amino-2-carboxyethyl)-L-lysine formed during alkaline hydrolysis. As discussed in Chap. 7, the tryptophan contents of proteins can also be obtained from various chemical modifications.

Because of problems with chemical hydrolysis, digestion with mixtures of proteolytic enzymes has been used as an alternative. Although this approach is successful in some cases, low yields of aspartate and glutamine often result, possibly due to the resistance of Asp-Asp bonds to proteolysis, the cyclization of glutamine residues or α-β rearrangement of peptide bonds involving aspartate. A further problem is presented by autolysis of the digestive enzymes and corrections must be made to accommodate such effects. Disulfide-containing proteins that may be particularly resistant to proteolysis must be reduced and alkylated prior to enzymatic digestion.

Analysis. While the analysis of amino acids by ion-exchange chromatography has been well documented, sensitivity in the picomole range is achievable by HPLC. The strategy for such analysis often involves a pre-column derivatization which must react with all amino acids, including the imino acids proline and hydroxyproline, in a reproducible and quantitative manner to give a stable derivative. A particularly useful reagent for this purpose is DABS-Cl (dimethylaminoazobenzenesulfonyl chloride), which in addition to fulfilling these requirements produces derivatives that are detectable at 420 nm in ethanol with very high sensitivity.

In an alternative approach, the amino acids obtained from hydrolysis are first separated and then detected in a post-column derivatization procedure.

DETERMINATION OF ACTIVITY

The determination of the "activity" of a protein can represent a difficult problem. While many proteins are enzymes, where some catalytic reaction can be followed, others have biological functions that do not fall into this category. Apart from the class described as enzymes, proteins can, in terms of the way one attempts to assay for them, be placed into two groups: binding and structural proteins. Each of these three classes is considered in turn.

Enzymes

The activity of an enzyme can often be conveniently measured by following either the production of a product or the removal of a substrate. With certain classes of enzymes (e.g., dehydrogenases) the natural substrates are chromophoric and exhibit spectral changes that can be followed directly. For example, alcohol dehydrogenase catalyzes the reduction of acetaldehyde by the coenzyme NADH:

$$\text{acetaldehyde} + \text{NADH} \rightleftharpoons \text{ethanol} + \text{NAD}$$

NADH has an absorption band centered at 340 nm with an extinction coefficient of $6.22 \times 10^3\ \text{cm}^{-1}\ M^{-1}$, while NAD has no absorbance at this wavelength. When alcohol dehydrogenase is added to a mixture of acetaldehyde and NADH, there is a time-dependent loss of absorbance at 340 nm, as seen in Fig. 1-1.

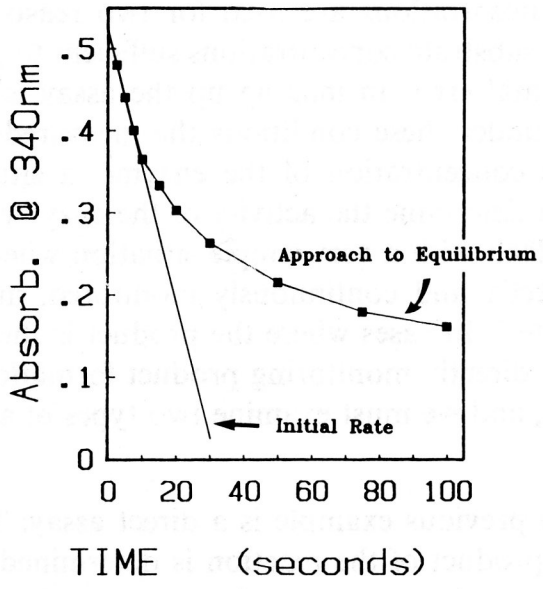

Figure 1-1 Complete time course of the reduction of acetaldehyde catalyzed by alcohol dehydrogenase using NADH as coenzyme.

If the reaction is allowed to proceed to equilibrium, the "rate" progressively slows until equilibrium is reached. Clearly, the reaction "rate" changes during the time course of the reaction as a consequence of both utilization of substrate and approach to equilibrium. To enable reproducible rate determinations, two aspects of the reaction are determined: (1) the initial rate, as shown in Fig. 1-1, and (2) the rate at saturating substrate concentrations. This rate (the "maximum rate") is calculated using concentrations of, in this case, acetaldehyde and NADH that give an experimentally determined maximum rate (Fig. 1-2).

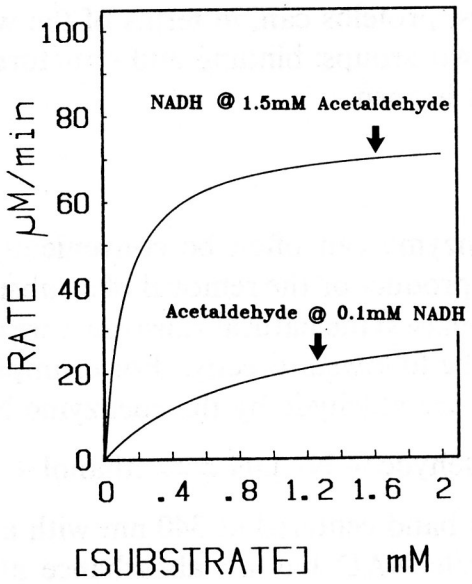

Figure 1-2 Dependence of the rate of the reaction catalyzed by alcohol dehydrogenase on the concentrations of NADH and acetaldehyde.

These substrate concentrations are used for two reasons, one pragmatic and the other theoretical. If substrate concentrations sufficient to give this maximum rate are used, any experimental error in making up the assay is minimized. From the theoretical standpoint, under these conditions the measured rate of the reaction is dependent only on the concentration of the enzyme, a situation necessary if the enzyme assay is used to determine the activity of the enzyme.

Our example has dealt with a very simple situation where the concentration of a substrate can be directly and continuously monitored, making it quite easy to determine an "initial rate." In cases where the product is the absorbing species, the rate can be followed by directly monitoring product formation. Many enzymes are not this simple to assay, and we must examine two types of assay procedures, direct and indirect.

Direct Assays. The previous example is a direct assay: The concentration of a particular substrate or product of the reaction is determined directly—in that case continuously. Consider another enzyme, galactosyl transferase, which catalyzes the reaction

$$\text{UDP-galactose} + N\text{-acetylglucosamine} \rightleftharpoons \text{UDP} + N\text{-acetyllactosamine}$$

None of these substrates or products can be conveniently monitored spectrophotometrically; however, the reaction rate can be followed directly using radioactively labeled UDP-galactose (with the label in the galactose moiety). The reaction is initiated by addition of the enzyme and allowed to proceed for a particular time interval before termination [in this case by the addition of EDTA since galactosyl transferase is a Mn(II)-requiring enzyme]. The reaction mixture is then chromatographed on an ion-exchange column (which binds UDP-galactose), the flow-through collected, and the amount of radioactivity determined. From the specific activity of

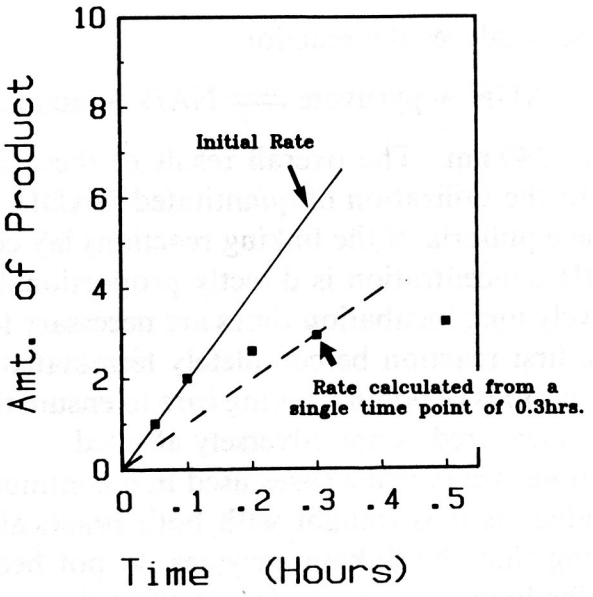

Figure 1-3 Time dependence of product accumulation during a discontinuous direct assay of enzyme activity.

the UDP-galactose the amount of *N*-acetyllactosamine produced in the time interval used can be measured and hence the reaction rate calculated. This rate is, however, not necessarily the initial rate discussed previously (see Fig. 1-3), and to establish that the initial rate is being determined the amount of product formed at several time intervals must be measured to show that at the particular time point chosen for routine assays the amount of product falls on a linear product versus time plot, proving that indeed the initial rate is being determined.

Such direct assays, where reaction mixtures are incubated, terminated, and then the amount of product or substrate measured after some type of separation procedure, are frequently employed, and provided that the separation procedure is effective, yield good results (usually this is controlled for by running blank reaction mixtures with no enzyme and subtracting "blanks" from values determined in the presence of enzyme).

Indirect Assays. Sometimes it is not convenient or possible to run direct assays, and indirect ones must be resorted to. In an indirect assay, product or substrate concentration is determined by linking the reaction to some second enzyme (or in

some cases several enzymes) to give an easily measured variable. Very frequently, reactions are linked to dehydrogenases due to of the ease of measuring their reactions.

Consider the example just used, galactosyl transferase. Its activity can also be determined in a linked assay as follows. After reaction termination, in this case through the addition of EDTA to chelate the required metal ion of the enzyme, pyruvate kinase and lactate dehydrogenase are added, together with phosphoenolpyruvate (PEP) and NADH. Although pyruvate kinase normally uses ATP/ADP, it also functions with UDP and catalyzes the reaction

$$\text{UDP} + \text{PEP} \rightleftharpoons \text{UTP} + \text{pyruvate}$$

Lactate dehydrogenase catalyzes the reaction

$$\text{NADH} + \text{pyruvate} \rightleftharpoons \text{NAD} + \text{lactate}$$

which is monitored at 340 nm. The overall result of these reactions is that UDP production is linked to the utilization of quantitated NADH. Provided that conditions are such that the equilibria of the linking reactions lay completely to the right, the decrease in NADH concentration is directly proportional to the UDP concentration. Where relatively long incubation times are necessary for the linking enzyme, it is essential that the first reaction be completely terminated. This can usually be accomplished by brief periods of boiling, taking care to ensure that the product whose concentration is to be measured is not adversely affected.

Such linked reactions are in some cases used in a continuous manner; however, this is not recommended, as it is fraught with both practical and theoretical difficulties such as ensuring that the linking enzymes do not become rate limiting, or that the substrates of the linking reactions do not affect the rate of the reaction being followed.

At this point it is beneficial to make several comments about enzyme activity measurements and to work through some calculations for the rates of enzyme reactions. The use of saturating substrate concentrations in reaction mixtures to minimize experimental errors has been emphasized. It is also important that reaction rates be measured under conditions where a sufficiently small amount of substrate is utilized so that the rate does not change during the assay as a result of substrate depletion. Similarly, product buildup, which may lead to product inhibition, is to be avoided. In general, a convenient way to test that these factors do not become a problem is to measure activity at a series of protein concentrations: The rate should be directly proportional to the protein concentration, as in Fig. 1-4. Deviations below the line indicate that substrate depletion or product accumulation may be occurring. Deviations from linearity can also result from protein aggregation or subunit dissociation affecting the rate of the catalyzed reaction.

Calculations of Enzyme Activity. Consider again the example of alcohol dehydrogenase in Fig. 1-1. From the linear extrapolation of the initial rate it is found that an absorbance change of 0.53 occurs in 27 minutes. From this a "rate" of 0.0196 A/min is calculated. Given that the millimolar extinction coefficient of NADH is

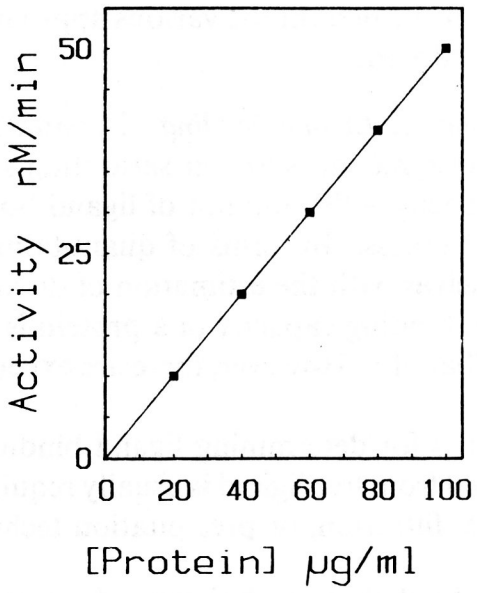

Figure 1-4 Dependence of measured rate on protein concentration for a typical enzyme assay measured using a discontinuous system.

6.22 cm^{-1}, the rate equals

$$0.0196/6.22 = 0.00316 \text{ m}M/\text{min}$$

That is, in terms of *concentration*, the rate in the spectrophotometer cuvette is 0.00316 mM/min. For convenience, assume that this measurement is made in a 1-ml cuvette. The rate in terms of *amount* of NADH utilized per minute is

$$0.00316 \text{ m}M = 0.00316 \text{ mmol/liter} = 3.16 \ \mu\text{mol/liter} = 3.16 \text{ nmol/ml}$$

So the rate, in terms of amount, is 3.16 nmol/min *per milliliter* of reaction mixture.

If a radioactive substrate is employed, as in the example of the direct assay for galactosyl transferase, the concentration of product formed is calculated from the specific activity of the isotope. Assume that the UDP-galactose has a specific activity of 3.7×10^5 cpm/mol and that during the course of an activity measurement, as outlined in Fig. 1-3, an initial rate of 284 cpm/min is determined in a 100-μl reaction mixture.

The *rate* of the reaction in terms of μmol/min is

$$284/3.7^5 = 0.00077 \ \mu\text{mol/liter} = 0.77 \text{ nmol/min}$$

If this rate is expressed in terms of *concentration*, the rate equals

$$0.77 \text{ nmol/min per } 0.1 \text{ ml} = 7.7 \text{ nmol/min/ml} = 7.7 \ M/\text{min}$$

Binding Proteins

The function of these proteins is usually assessed by their ability to bind a specific ligand or an analog of that ligand. Equilibrium binding of a ligand to a protein can be determined in various ways, some of which are dealt with from a theoretical

standpoint in Chap. 17. For the present the various approaches are introduced briefly from the experimental standpoint.

Direct Methods of Studying Ligand Binding. The analogous parameter in ligand binding to the rate of an enzyme measured at saturating substrate concentrations is the maximum binding capacity—the amount of ligand bound by the protein in the presence of a large ligand excess. In terms of quantitating the amount of protein, we need not concern ourselves with the estimation of dissociation constants. Determination of the maximum binding capacity of a protein is not a simple matter, as is discussed thoroughly in Chap. 17. However, there are experimental difficulties as well as theoretical problems.

In the direct techniques for determining ligand binding, some method to separate physically bound ligand from free ligand is usually required. This may be achieved by equilibrium dialysis, gel filtration, or precipitation techniques.

Equilibrium Dialysis. In dialysis techniques, of which there are a variety, the protein-containing solution is separated from buffer by a semipermeable membrane that allows equilibration of the ligand. Ligand initially can be added to either side of the membrane, and after reaching equilibrium aliquots are withdrawn from both the free buffer and the protein-containing solution, and the concentration of ligand is determined (by radioactivity, spectral or chemical methods, or enzymic assay). The amount of bound ligand is obtained by subtracting the free ligand concentration from the total ligand concentration (bound + free) in the protein-containing solution. Equilibrium is ensured by running appropriate controls where ligand is added to only one of the compartments in the absence of any protein. The dialysis may be passive or forced by application of pressure to the protein-containing compartment in an apparatus such as the Amicon filter.

Gel Filtration. Separation of free ligand from protein-bound ligand can often be achieved rapidly and effectively by gel filtration using a matrix, where the free ligand penetrates the matrix but the protein and protein-bound ligand are excluded. The application of gel filtration techniques to ligand binding is usually achieved in one of two ways. Either the protein is passed through a gel filtration column equilibrated with a set concentration of ligand and the concentration of protein and ligand monitored in the effluent, or a small sample of protein plus ligand is applied to the top of a syringe or centrifuge tube column of matrix equilibrated in buffer containing no ligand. The column is then rapidly centrifuged, forcing the protein (and bound ligand) to be rapidly eluted, and the eluted protein and ligand concentration is determined. In the latter case the appropriate "blank" values must be subtracted.

Precipitation. In many instances the approaches already discussed cannot readily be applied. This is most often the case where the ligand is a protein or other large molecule. In such instances it may be possible to precipitate the complex between receptor and ligand but not the ligand free in solution. This approach may also be applied to small ligands where the complex can be precipitated without loss of affinity

for the ligand and the precipitate washed to remove free ligand. This is frequently the case with proteins such as hormone receptors, which have very high affinity for their specific ligands.

Two points to bear in mind concerning these techniques need to be mentioned here as well as one experimental problem in determining the maximum binding capacity of a protein.

Nonspecific Binding. In many instances, especially with impure protein preparations, both specific (i.e., to the binding site of the desired protein) and nonspecific binding may occur. Specific binding, especially in the case of hormone receptors, is usually defined as being readily displaceable by a competing ligand, whereas nonspecific binding is usually nondisplaceable.

Chemically or Isotopically Altered Ligands. Often, the natural ligand for a protein cannot be used conveniently and an analog, either radioactively or spectrally labeled, is employed. Although this is usually not a problem, on occasion the alteration of the ligand may induce inhomogeneity: a mixture of two forms may arise where one is active in terms of binding and one inactive, leading to erroneous calculations of the total amount of ligand present in a particular compartment. It is therefore essential to determine that all of the ligand used can be bound by the protein. This is done by titrating a small amount of ligand with increasing amounts of protein to determine what proportion of the ligand can be bound. With small ligands this is not normally a problem unless the modification has produced racemization of an asymmetric carbon, but it is often encountered with protein or peptide ligands that have been radioactively labeled, where, due to differing extents of modification, some of the protein may have been inactivated. Where assays of biological function can be correlated with binding assays, additional confidence in the ligand is obtained.

Finally, we must consider an overall experimental problem in the determination of the maximum binding capacity of a protein. As with kinetic assays to determine activity, large excesses of ligand must be used to ensure saturation. This results in trying to measure the difference between two large numbers to assess the amount of bound ligand—obviously not the ideal situation, but one about which very little can be done except exercising experimental dexterity and using large numbers of replicates, the latter of which is usually inconvenient in the course of protein preparation.

Indirect Methods. As is discussed in considerable detail in Chap. 17, there are a wide variety of indirect methods for studying ligand binding, ranging from titrations employing spectral parameters of either the protein or ligand, to studies involving the ability of ligand to alter in some way an identifiable protein property. Although these approaches are frequently of use in the context of ligand binding studies, only those involving spectral parameters are of much use in the routine assay of protein binding activity. Briefly, these methods, whether they use absorbance, fluorescence, or magnetic properties, depend on the protein–ligand complex exhibiting some spectral difference from the free protein and ligand solutions that can be used to follow ligand binding. Although clearly such an approach cannot be used to quantitate

the amount of protein when it is not pure, the method can be used to follow the protein activity and to calculate a "specific activity." If one makes assumptions about the number of binding sites for ligand, it can also be used to quantitate the protein.

Structural Proteins

The determination of the "activity" of structural proteins represents the most difficult aspect of protein purification: frequently, activity measurements are only possible by immunoassay using antibody to the purified material.

The problems may best be illustrated by reference to the purification of laminin, a glycoprotein obtained from basement membranes. In essence, laminin was discovered as an impurity in collagen preparations from tumor cells: It was purified by standard approaches to give a "pure" protein by protein chemistry criteria—antibodies were made, and, via immunofluorescence studies, an in vivo localization of the laminin was achieved. Subsequently, immunoassay techniques were used to follow the purification.

In other instances, assays of biological function may be employed to follow purification, as has been used for example with interleukin-II, a T-lymphocyte growth factor required for the stimulation and growth of responding T cells, which is produced by a particular subset of T cells.

Finally, it is sometimes possible to "purify" a protein based on some particular chemical property—a unique cofactor or unusual amino acid composition, for example, the hydroxyproline content of collagen. Clearly, there are no firm rules for the assay of structural proteins, and definition of activity has to be rather pragmatic.

SPECIFIC ACTIVITY

Finally, we come to a definition of specific activity. Depending on the state of knowledge concerning the molecular weight of the protein, the specific activity can be defined by one of the equations

$$\text{specific activity} = \frac{\text{activity}}{\text{mg protein}} \tag{1-1}$$

$$\text{specific activity} = \frac{\text{activity}}{\text{mol protein}} \tag{1-2}$$

In either instance, as the purity of the protein preparation increases, the apparent specific activity increases, reaching a defined value for the pure protein. As indicated earlier and considered in more detail in Chap. 2, it is often considered that a protein is pure when a constant specific activity is reached in a purification. If the specific activity of the pure protein is known, the apparent specific activity can be used to calculate the percent purity at a given point in a purification scheme.

Calculation of Specific Activities

From the example of alcohol dehydrogenase discussed earlier, a *specific activity* can be calculated if the concentration of enzyme or the amount of enzyme in the reaction mixture is known. In the example we calculated a rate of 0.00316 mM/min. If this rate was achieved with a protein concentration of, for instance 4 μg/ml, one can calculate the specific activity,

$$0.00316/4 = 0.00079 \text{ m}M/\mu\text{g/ml/min}$$

If the molecular weight of alcohol dehydrogenase is known, the molarity of the enzyme can be calculated and the specific activity expressed in terms of much simpler units. Assume that the molecular weight is 70,000; then 4 μg/ml gives a molecular concentration of

$$70{,}000 \text{ g/liter} = 1\ M, \qquad 70{,}000 \text{ mg/ml} = 1\ M, \qquad 1 \text{ mg/ml} = \tfrac{1}{7} \times 10^{-4}\ M$$

$$4\ \mu\text{g/ml} = \tfrac{4}{7} \times 10^{-7}\ M = 5.7 \times 10^{-8}\ M$$

Since the rate is expressed in millimolar units, this becomes 5.7×10^{-5} mM for a 4-μg/ml solution. The *specific activity* is therefore

$$\frac{0.00316}{(5.7 \times 10^{-5})\ \text{min}^{-1}} = 55.4\ \text{min}^{-1}$$

In this type of calculation care must be taken to define the "molecule" of the enzyme. In these calculations a molecular weight of 70,000 was used: This is the *dimer* molecular weight (the dimer contains *two* active sites). The specific activity calculated on the basis of the overall molecular weight can be divided by the number of active sites per molecule to give the active site specific activity: in this example, $55.4/2 = 27.7\ \text{min}^{-1}$.

Finally, even when activities are measured with saturating substrate concentrations as described earlier, temperature and pH must be defined, and if buffer salts affect the enzyme activity, these must also be defined. Thus the active-site specific activity would be given as $27.7\ \text{min}^{-1}$ at 25°C, pH 7.0 in phosphate buffer. If maximum-rate substrate concentrations are not used, these should also be given.

SOURCES OF PROTEIN

Before discussing fractionation procedures for proteins (Chap. 2), there are several generalities that can be made concerning sources of proteins for purification. Often the goal is to obtain, in pure form, a particular enzyme (or other activity) for further study, with little regard to the starting material. In such a case it is extremely helpful initially to screen a variety of species and tissues within a species to obtain a starting material with the highest initial specific activity. Once a starting tissue has been selected it is necessary to establish whether the protein is soluble (i.e., in the

cytosol), membrane bound, or entrapped in a subcellular organelle. When the protein is cytostolic the cells must be ruptured to release the cytosol; this can be achieved by homogenization, sonic shock, or freeze-thawing.

Many proteins can be described as membrane bound, and it is often convenient as well as beneficial to isolate the appropriate membrane fraction prior to isolating the protein, as this can give a much higher apparent specific activity of the starting material relative to a whole-cell extract.

Initial stages of membrane preparation involve membrane disruption, usually by mechanical means such as homogenization. The homogenate is then centrifuged at low speed to remove nuclei and whole cells. Different subcellular organelles such as mitochondria and microsomes can then be obtained using differential centrifugation. Finally, discontinuous density gradient centrifugation (frequently using sucrose gradients) can be employed to yield various membrane fractions. A typical scheme is given in Table 1-1. Different membrane fractions are identified by marker enzymes, which are summarized in Table 1-2.

Membrane-bound enzymes usually need to be solubilized prior to further fractionation. Two common approaches are used for so-called peripheral and integral membrane proteins. Peripheral proteins can usually be solubilized by high ionic

TABLE 1-1 Preparation of membrane markers

Step	Procedure
1	Homogenize cells (Dounce homogenizer)
2	Centrifuge at 1000*g* to remove nuclei and whole cells
3	Differential centrifugation
	(a) 10,000*g*, 15 min → Mitochondrial preparation
	(b) Supernate from (a): 113,000*g* for 30 min → microsomal preparation
	Discontinuous sucrose gradient centrifugation, frequently with sucrose gradients ranging from 20 to 50%

TABLE 1-2 Membrane markers

Membrane	Marker enzyme
Mitochondrial	
Inner	Succinate–cytochrome *c* reductase
	Rotenone sensitive; NADH–cytochrome *c* reductase
Outer	Monoamine oxidase
	Rotenone-insensitive NADH–cytochrome *c* reductase
Endoplasmic reticulum	RNA and protein synthesis enzymes
	NADPH–cytochrome *c* reductase
Plasma	5-Nucleotidase
	Lectin binding
	Oxytocin (or hormone) binding
Golgi	Glycosyl transferases

strength, while integral proteins require detergent solubilization. In many instances differential solubilization, either of peripheral versus integral proteins or of different integral proteins, by different detergents, can be of considerable help in enhancing the specific activity of the starting extract.

Secreted Proteins

A number of proteins of considerable biological interest are secreted from cultured cells and can be harvested from culture filtrate. With bacterial cells this usually represents an excellent source of such proteins. With mammalian cells, however, a major problem is that the cultures frequently require quite high percentages of serum to grow, and thus secreted proteins may represent a small proportion of the extracellular proteins.

A similar problem is encountered with the secreted material from solid tissue cells such as hepatocytes, which often require a collagen-based matrix to grow. A particular problem with extracellular material from solid tissues such as the liver is encountered in the attempted isolation of glycocalyx material, a loose aggregate of glycoprotein and proteoglycan material that surrounds the plasma membrane and may be regarded as an extracellular organelle. Since solid tissue cells are often harvested by techniques such as collagenase perfusion to degrade the extracellular support of the cells, the existence and isolation of glycocalyx material were not recognized or achieved until gentler means of tissue disruption were introduced.

Protease Problem

In a number of instances the isolation of a particular protein is hindered by its susceptibility to various nonspecific (and in some cases specific) proteases which abound in the cell. Such protease activities may reside in lysosomes or be associated with membrane fractions and are often released or activated (e.g., by detergents) during the course of the initial stages of a protein purification. Their effects are frequently combated by the use of various protease inhibitors that can be added during the isolation procedures. Some of the more troublesome proteases and ways to inhibit them are outlined in Table 1-3.

TABLE 1-3 Proteases and their inhibitors

Protease	Specificity	Inhibitor
Chymotrypsin	Phe, Tyr, Trp	TPCK, DFP, PMSF
Trypsin	Lys, Arg	TLCK, DFP, SBTI
Pepsin	Phe, Trp, Tyr	*p*-Bromophenacylbromide, Diazo-*p*-bromoacetophenone
Papain	None	Iodoacetic acid, heavy metals
Thermolysin	Hydrophobics	EDTA, 1, 10-phenanthroline
Carboxypeptidase B		1, 10-phenanthroline, 2,2-dipyridyl
Cathepsin A		Antipain
Urokinase		ε-Aminocaproic acid
Kallikreins		α-2-macroglobulin, DFP

TABLE 1-4 Small-molecule substrates for proteases

Protease	Substrate	Monitored by:
Papain	L-Pyroglutamyl-L-Phe-L-Leu-*p*-nitroanilide	Liberation of *p*-nitroaniline monitored at 410 nm
Elastase	CBZ-Ala-*p*NP	Liberation of *p*-nitrophenol monitored at 400 nm
Chymotrypsin	Ac-Trp-*p*NP, *p*-nitrophenyl-ethyl carbonate	Liberation of *p*-nitrophenol monitored at 400 nm
Trypsin	CBZ-Lys-*p*NP	Liberation of *p*-nitrophenol monitored at 400 nm

It is important to consider the possibility of proteases in any purification scheme and it is useful to be able to assay for their activity in the initial extract as well as during subsequent stages in a purification. Protease assays are also helpful in determining the efficacy of added protease inhibitors. Two general types of protease assays are employed: one involving chromagen or radioactively labeled protein and the other small molecule protease substrates that release a chromophore after hydrolysis by a protease. In the former the labeled protein substrate is added and " soluble" label detected after precipitation of protein by trichloroacetic acid. In the latter type of assay, the release of chromophore as a result of the hydrolytic activity of the protease is monitored. A number of small-molecule protease substrates have been developed, which are summarized in Table 1-4.

Once a suitable starting material is obtained, an initial specific activity must be determined together with the total amount of the desired protein present in the starting material. When the specific activity of the pure protein is known, this amount can be calculated in terms of milligrams of protein. However, this is often not the case, and the total amount of active protein can be expressed as *units* of protein, where the definition of the *unit* must be made in terms of the activity being measured. During the subsequent purification stages (Chaps. 2 and 3) both the specific activity and overall yield (the percentage of the original amount of active protein) should be monitored. Such comparisons from step to step in the purification require that assays be performed under standard conditions of substrate concentration, pH, temperature, and buffer components, since each of these factors can affect the measured activity.