4

Molecular Weight Determination

INTRODUCTION

The determination of the molecular weight of a molecule would appear to be quite a straightforward procedure. For small molecules this is true: Elemental analysis and approaches using colligative properties yield easily interpretable answers. The situation with a protein is somewhat different: If the primary sequence is known, it is a quite simple matter to determine the molecular weight from the amino acid composition if a single chain is present. Many proteins, however, contain subunits that may or may not be chemically similar. The molecular weight is then either a multiple of the molecular weight of the monomeric polypeptide chain or a combination of different polypeptide chains. Thus, even if the sequence is known (or sequences if more than one type exist) for the polypeptide chain, one does not necessarily know the molecular weight of the native molecule. Distinctions must be drawn between the native molecular weight and the denatured molecular weight. The native molecular weight is that as the protein exists under normal in vivo conditions where it exhibits normal and full activity. The denatured molecular weight can be defined as that of the minimum covalently bonded structure that the molecule is broken down to under denaturing conditions such as the presence of SDS or guanidinium hydrochloride. This molecular weight must not be taken as the polypeptide chain molecular weight since two or more chains may be covalently linked by disulfide bridges and may contain covalently bound nonpolypeptide entities such as carbohydrates, which contribute to the molecular weight. As we shall see, carbohydrate substituents can lead to anomalous molecular weight estimates using some of the approaches described. Similarly, most of the methods for determining molecular weights make

some assumption concerning either the shape or the behavior of the protein relative to some reference: deviations in behavior from this reference lead to misleading molecular weight estimates. These estimates are, in many instances, averages, and depending on how the averaging is done, different answers for mixtures of species of different molecular weights are obtained. The various types of molecular weight averages that can be experimentally determined are described in detail here. Different techniques give different molecular weight averages, and a comparison of those obtained by approaches giving different types of molecular weights average also yield an idea of purity.

The concept of protein purity is somewhat nebulous: As sensitivity of protein detection has been increased, especially in techniques such as polyacrylamide gel electrophoresis (PAGE), it has also been far easier to detect impurities. Many of the experimental approaches for determining molecular weights are also used to estimate purity. Careful consideration of these methods is thus helpful in the context of examining purity as well as determining molecular weights.

MOLECULAR WEIGHT AVERAGES

If a sample contains a distribution of species with different molecular weights (defined in the classical sense), the average molecular weight depends on how the averaging process is done.

If there are $n(M)\,dM$ moles of species with molecular weights between M and $M + dM$, the total number of species present in the sample, n^T, is

$$\int_0^\infty dMn(M) \tag{4-1}$$

where the function $n(M)$ is the molecular weight distribution function. Moments of this distribution function can be defined. The Kth moment is

$$m_k = \int_0^\infty dMn(M)M^K \tag{4-2}$$

The total number of species present, n^T, is the zeroth moment.

Average molecular weights are defined in terms of ratios of higher moments (Kth) to the ($K - 1$)th moment. The *number-average molecular weight*, $M_n = m_1/m_0$:

$$M_n = \frac{\int_0^\infty dMn(M)M}{\int_0^\infty dMn(M)} \tag{4-3}$$

If we consider a discrete distribution of species containing n_i moles of components with molecular weight M_i, we can write the expression for M_n as

$$M_n = \frac{\sum_i n_i M_i}{\sum_i n_i} \tag{4-4}$$

Alternatively, to obtain an expression in terms of molar concentrations, divide by the volume V:

$$M_n = \frac{\sum_i n_i M_i / V}{\sum_i n_i / V} \tag{4-5}$$

Defining the weight concentration, $C_i = n_i M_i / V$, we get

$$M_n = \frac{\sum_i C_i}{\sum_i C_i / M_i} \tag{4-6}$$

The *weight-average molecular weight* is defined by the ratio m_2/m_1,

$$M_w = \frac{m_2}{m_1} = \frac{\sum_i n_i M_i^2}{\sum_i n_i M_i} \tag{4-7}$$

or in terms of weight concentrations,

$$M_i = \frac{\sum_i C_i M_i}{\sum_i C_i} \tag{4-8}$$

The third average that is commonly used is the so-called *Z-average molecular weight*, defined as m_3/m_2:

$$M_z = \frac{m_3}{m_2} = \frac{\sum_i n_i M_i^3}{\sum_i n_i M_i^2} \tag{4-9}$$

which in terms of concentration yields

$$M_z = \frac{\sum_i C_i M_i^2}{\sum_i C_i M_i} \tag{4-10}$$

As will be discussed, the various methods of determining molecular weights usually give either M_n or M_w. However, molecular weights estimated by viscosity measurements give a *viscosity-average molecular weight*, M_v, defined by

$$M_v = \left(\sum_i W_i M_i^a \right)^{1/a} \tag{4-11}$$

where W_i is the weight fraction of the ith species and a is an empirical constant that varies between 0.5 and 2.0, depending on the type of molecular weight average.

From the equation for M_v we get

$$M_v = \left(\frac{\sum_i n_i M_i^{(1+a)}}{\sum_i n_i M_i} \right)^{1/a} \tag{4-12}$$

and hence when $a = 1.0$ this expression is the same as that for M_w, and $M_v = M_w$. For $a < 1.0$, M_v is between M_n and M_w.

Effects of Purity

As indicated earlier, different experimental methods for determining molecular weights can give different types of molecular weight averages. This gives a basis for assessing the purity of a sample in a homogeneous sample $M_n = M_w = M_z$. However, with a mixture of different molecular weights the values for each of these averages *can* be quite disparate. Two examples illustrate this point.

Consider an equal mixture (by weight) of two molecules, one with a molecular weight of 10,000 and the other 100,000. In this instance,

$$M_n = \frac{1 + 1}{1/10{,}000 + 1/100{,}000} = 18{,}181$$

$$M_w = \frac{10{,}000 + 100{,}000}{1 + 1} = 55{,}000$$

and

$$M_z = \frac{10{,}000^2 + 100{,}000^2}{10{,}000 + 100{,}000} = 91{,}818$$

These values should, of course, easily be separated experimentally, allowing one to conclude that a mixture is indeed present.

However, if the mixture consists of equal weights of proteins with molecular weight of 80,000 and 100,000, the values for M_n, M_w, and M_z are 88,888, 90,000, and 91,111, respectively, which would be most difficult to resolve experimentally and could lead to the erroneous conclusion that one was dealing with a homogeneous solution of protein with a molecular weight of $90{,}000 \pm 1500$.

If, instead of equal weights of two species we have equal *numbers* of two types of polypeptide chains, then for our first example $M_n = 90{,}000$, $M_w = 91{,}111$, and $M_z = 92{,}195$, which, as before, would not easily be experimentally separated.

As we examine various methods for estimating molecular weights, these considerations should be remembered. In some of the techniques a clear indication of heterogeneity is obtained, but in others only the appropriate molecular weight average is obtained, and comparison of molecular weight averages can then be a valuable tool in estimating purity.

MOLECULAR WEIGHT DETERMINATION BY SITE TITRATIONS

One of the simplest and earliest methods employed to determine the molecular weight of a protein is based on the assumption that the protein contains a single specific site of some sort. This might be a binding site for a certain ligand, an N- or C-terminal amino acid residue, a cofactor such as a metal ion, or in an extreme case the amino acid present in the lowest amount in the amino acid composition. The molecular weight of the molecule is then calculated on the basis of 1 mol of this specific site per mole of protein. The determination of the number of specific sites present per unit weight of protein and the calculation of the average molecular weight by dividing the weight by the number of molecules present clearly indicates that these methods give a *number-average* molecular weight.

Determination of molecular weight by these approaches has some pitfalls. Consider a situation where the molecular weight of a protein consisting of four chemically identical polypeptide chains is determined by end-group analysis or by titration of a single specific ligand binding site per polypeptide chain. The calculated molecular weight is one-fourth of the true molecular weight. It represents a minimum molecular weight defined in terms of the specific site. The types of measurements involved in this sort of determination are usually quite accurate, and with a rough estimate of the molecular weight by another approach, can be used to give an accurate estimate of the molecular weight. If the protein, instead of consisting of four chemically identical chains consists of two types of chains with different sizes, a quite erroneous molecular weight may be calculated depending on the exact situation. If an $(A - B)_2$ tetramer has a total of two binding sites for a ligand (one for each AB pair) and the molecular weight is determined by site titrations, the molecular weight is one-half of the true molecular weight no matter how disparate the sizes of the A and B chains are. If, however, the molecular weight is determined by the number of moles of dansyl chloride (an amino-terminal labeling reagent) incorporated per gram of protein, quite erroneous estimates are obtained.

METHODS BASED ON COLLIGATIVE PROPERTIES SUCH AS OSMOTIC PRESSURE

Although a number of colligative properties are used in the determination of low molecular weights, osmotic pressure measurements are the only type that have found much use with proteins. To determine the type of molecular weight average obtained by osmotic pressure measurements, let C_i be the concentration of the protein and OP the osmotic pressure. Then

$$\mathrm{OP} = k \sum_i C_i \tag{4-13}$$

where k is the proportionality constant between osmotic pressure and concentration. Since the only quantity known is the total weight concentration of the protein in

solution, C, which is equal to $\sum C_i$ (which in turn equals $\sum M_i C_i$), we get, from Eq. (4-13),

$$\frac{\text{OP}}{C} = \frac{k \sum_i C_i}{\sum_i C_i M_i} \tag{4-14}$$

Since C_i is proportional to N_i one can write

$$\frac{\text{OP}}{C} = \frac{k \sum_i N_i}{\sum_i N_i M_i} \tag{4-15}$$

Earlier we defined $M_n = \sum N_i M_i / \sum N_i$, and hence

$$\frac{\text{OP}}{C} = \frac{k}{M_n} \tag{4-16}$$

and the molecular weight obtained from osmotic pressure (or any other colligative property) measurement is a *number-average* molecular weight.

Osmotic pressure, like all colligative properties, is a measure of the chemical potential, μ, of the solvent, and can thus be used to measure the molecular weight of the solute (at infinite dilution). In the experimental determination of the osmotic pressure of a solute molecule, the solute is separated from pure solvent by a semipermeable membrane and the resultant "pressure" determined as outlined schematically in Fig. 4-1A. The osmotic pressure π results from the chemical potential of the solute and is defined as $\pi = P - P_o$.

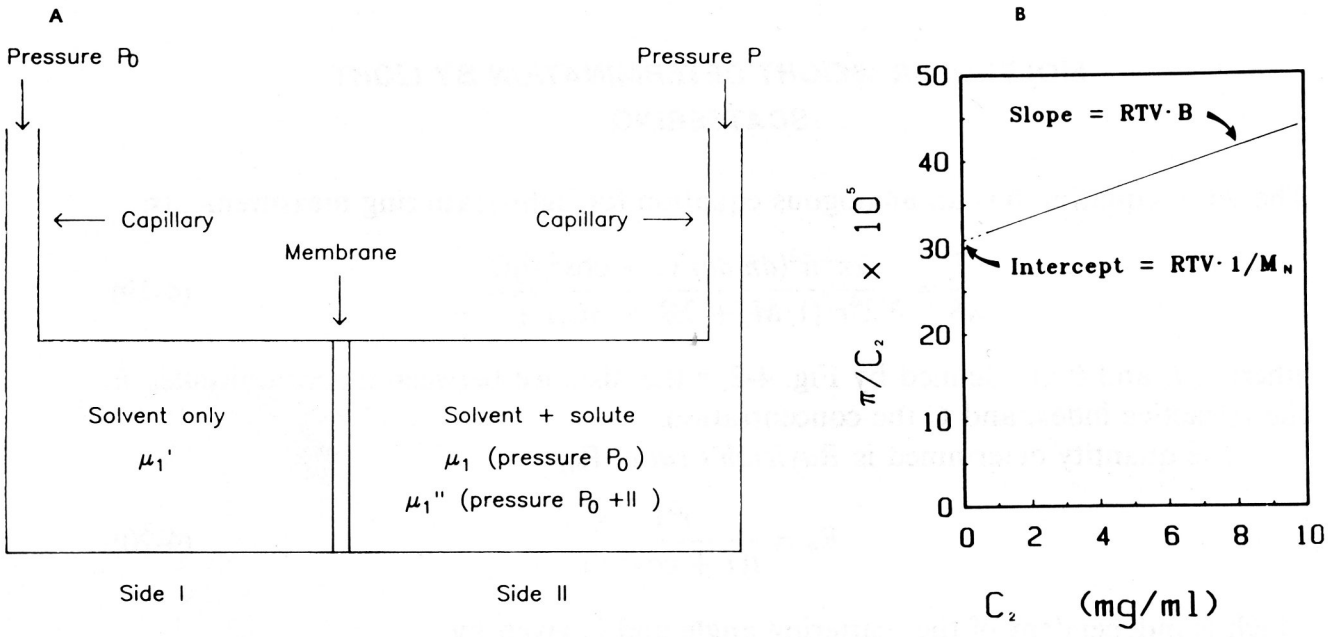

Figure 4-1 A: Schematic of apparatus for osmotic pressure determination. Osmotic pressure $\pi = P - P_0$. B: Plot of π/C_2 versus C_2 to show calculation of molecular weight and the virial coefficient, B.

At equilibrium in an osmotic pressure determination, molecules that can freely pass through the membrane do *not* affect the molecular weight estimates of those molecules retained by the membrane. From van't Hoff's limiting laws for osmotic pressure, extrapolation of the osmotic pressure to zero concentration allows the molecular weight, M_i, of the solute, to be obtained from Eq. (4-17)

$$\frac{\pi}{C_2} = \frac{RT}{M_i} \tag{4-17}$$

In reality, of course, the dependence of the osmotic pressure, π, on solute concentration is given by a virial equation of the form of Eq. (4-18). In this general equation the osmotic pressure $\pi = \mu_1, - F_1^0$.

$$\mu_1 - F_1^0 = -RTV_1^0 C_2\left(\frac{1}{M_i} + B_{c_2} + C_{c_2}^2 + \cdots\right) \tag{4-18}$$

where F_1^0 is the molar free energy, C_2 the concentration of the solute, V the partial molar volume of the solvent, and B and C the virial coefficients. This equation shows that the effect of the solute on the chemical potential decreases as the molecular weight increases. From the limiting slope of the plot of π/C_2 versus solute concentration, an estimate of the virial coefficient, B, is obtained, together with the molecular weight, as shown in Fig. 4-1B.

Although, as indicated previously, the osmotic pressure at a given solute concentration decreases as the solute molecular weight increases, molecular weights up to 200,000–300,000 can be measured quite accurately.

MOLECULAR WEIGHT DETERMINATION BY LIGHT SCATTERING

The virial equation has an analogous equation for light-scattering measurements:

$$\frac{i}{I} = \frac{2\pi^2 \bar{n}^2 (d\bar{n}/dc)^2 (1 + \cos^2\theta) C}{N\lambda^4 r^2 (1/M_i + 2B_c + 3C_{c^2} + \cdots)} \tag{4-19}$$

where i, I, and θ are defined by Fig. 4-2, r the distance between macromolecules, $\bar{n}$ the refractive index, and C the concentration.

The quantity determined is *Rayleigh's ratio*, R_θ,

$$R_\theta = \frac{r^2 i}{I(1 + \cos^2\theta)} \tag{4-20}$$

which is independent of the scattering angle and is given by

$$R_\theta = \frac{K_c}{1/M_i + 2B_c + 3C_{c^2} + \cdots} \tag{4-21}$$

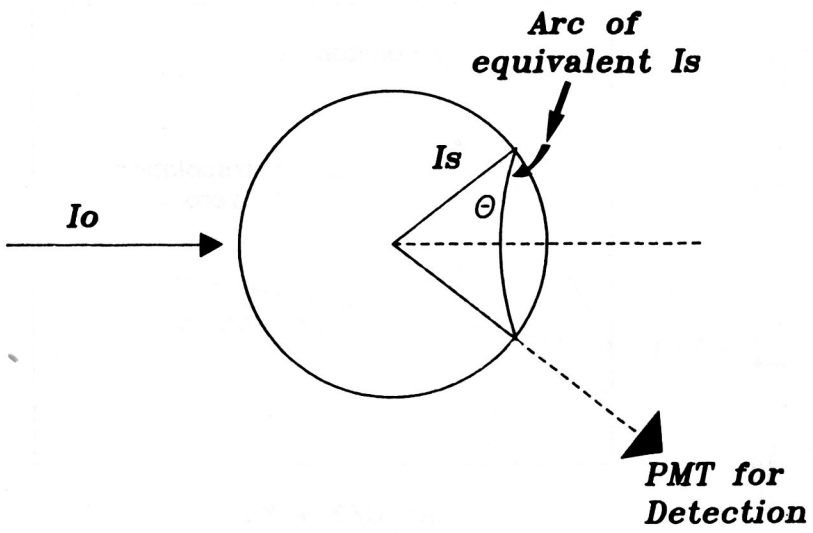

Figure 4-2 Scheme of light-scattering experiment. Scattered light is quantitated using a photomultiplier tube (PMT).

where

$$K = \frac{2\pi^2 \bar{n}^2 (d\bar{n}/dc)^2}{N\lambda^4}$$

The equation for R_θ can be rewritten as

$$\frac{K_c}{R_\theta} = \frac{1}{M_i} + 2B_c + 3C_{c^2} + \cdots \tag{4-22}$$

and a plot of K_c/R_θ versus C gives an intercept of $1/M_i$.

When the scattering molecules are of comparable size to the wavelength of the scattered light, the equation becomes

$$\frac{K_c}{R_\theta} = \frac{1}{M_i}\left(1 + \frac{16\pi^2 R_G^2 \sin^2 \theta}{3\lambda^2}\right) \tag{4-23}$$

where R_G is the radius of gyration of the scattering molecule. This equation is the basis of the Zimm plot (Fig. 4-3).

In the Zimm plot the molecular weight is obtained from the intersection point of the extrapolation at constant θ to $C = 0$ and at constant C to $\theta = 0$. The radius of gyration of the scattering molecule is also obtained from the initial slope of the curve at constant m extrapolated to $C = 0$. In Fig. 4-3 the x axis is plotted as $(\sin^2 \theta/2) + K'C$, where K' is an arbitrary constant added partly for convenience of presentation and partly to account for intermolecular effects.

It is quite apparent that light scattering is proportional to the weight concentration of the solute and its molecular weight, and one can write

$$\text{LS} = k \sum_i C_i M_i \tag{4-24}$$

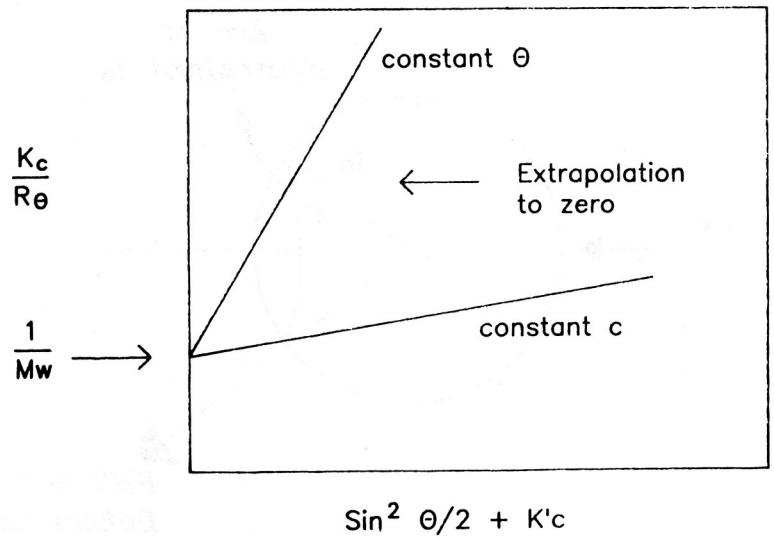

Figure 4-3 Outline of the Zimm plot for the extrapolation of light-scattering data.

Therefore,

$$\frac{LS}{C} = \frac{k\sum_i C_i M_i}{\sum_i C_i} = \frac{k\sum_i g_i M_i}{\sum_i g_i} \tag{4-25}$$

and hence light-scattering measurements give a *weight-average* molecular weight.

SEDIMENTATION METHODS

There are two types of sedimentation methods commonly employed to study the molecular weight of a macromolecule: sedimentation velocity methods and sedimentation equilibrium methods. In each it is necessary to experimentally follow the solute protein during its sedimentation in an ultracentrifuge. In sedimentation velocity the rate of sedimentation is followed, while in sedimentation equilibrium the concentration gradient of the sedimenting macromolecule produced by a variety of opposing forces during centrifugation is established. As is inherently to be expected, sedimentation equilibrium uses lower centrifuge speeds than sedimentation velocity experiments. In most cases the sedimentation of the protein is experimentally followed by absorbance measurements. Although the rigorous derivation of the equations describing the behavior of a macromolecule during sedimentation is beyond the scope of this chapter, it is informative to examine some of them.

Sedimentation Velocity Experiments

In a two-component system containing the solvent (1) and the solute (2), a flux, J_2, of the macromolecule at a particular point X can be described by

$$J_2 = L_{22}\left(W^2 X - \frac{d\bar{\mu}_2}{dX}\right) \tag{4-26}$$

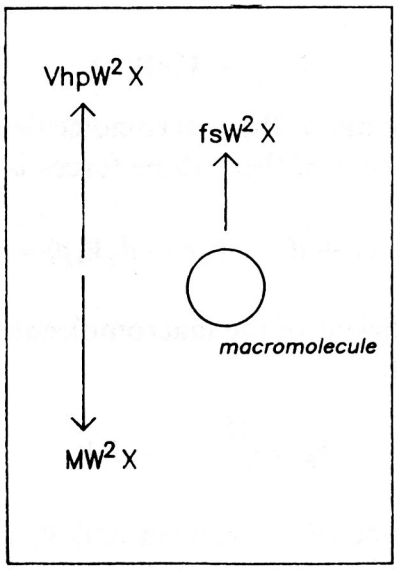

Figure 4-4 Forces on a macromolecule in an ultracentrifuge.

where L_{22} is a phenomenological constant, W the angular velocity, μ_2 the chemical potential of the solute, and X the radius to the point being considered (see Fig. 4-4). The forces due to radial acceleration and diffusion (as the result of a chemical potential gradient) are described respectively by the first and second terms within the parentheses. If there is no diffusion the applied force in the ultracentrifuge is W^2X and produces a constant velocity, V, which is a function of the size and shape of the sedimenting solute molecule. The proportionality constant is defined as

$$S = \frac{V}{W^2X} = \frac{1}{W^2X}\frac{dX}{dt} \tag{4-27}$$

and is known as the sedimentation coefficient. When diffusion occurs, the flux, J_2, becomes

$$J_2 = W^2XSC_2 - D\frac{dc_2}{dX} \tag{4-28}$$

where D is the diffusion constant, and we can now examine the relationship between S and the molecular properties of the solute. As illustrated in Fig. 4-4, there are three forces on the hydrated protein molecule during centrifugation.

The total acceleration force, F_a, on a single molecule is

$$F_a = \frac{M}{N_0}(1 + d_1)W^2X \tag{4-29}$$

where d_1 is the hydration of the molecule, M its molecular weight, and N_0 is Avogadro's number. The resultant acceleration is opposed by a frictional drag force, F_f, and a buoyant force, F_b, which is proportional to the mass of solvent, m_s, displaced by the protein molecule

$$F_f = -fSW^2X \tag{4-30a}$$

and

$$F_b = -V_h p W^2 X \tag{4-30b}$$

where V_h is the hydrated volume of the macromolecule and p is the solvent density.

In the steady state, the sum of these three forces is zero, and

$$\frac{M}{N_0}(1 + d_1 - V_2 p - d_1 V_1 p) = S'f \tag{4-31}$$

where f is the frictional coefficient of the macromolecule and V_h in the expression for F_b has been substituted for by

$$V_h = \frac{M}{N_0} V_2 + d_1 V_1 \tag{4-32}$$

where V_2 is the specific volume of the protein and V_1 is the partial specific volume of the solvent. Since in dilute solution $V_1 \cong 1/p$, we get, after rearranging,

$$S = \frac{M(1 - V_2 p)}{N_0 f} \tag{4-33}$$

Combining this equation with

$$D = \frac{kT}{f} \tag{4-33a}$$

for the diffusion, D, we get

$$M = \frac{SRT}{D(1 - V_2 p)} \tag{4-34}$$

which is the *Svedberg equation.*

So far we have assumed that the sedimentation coefficient (S) and the diffusion coefficient (D) are independent of concentration. In reality this is not true, and for accurate molecular weight determination it is necessary to extrapolate values of S and D to infinite dilution.

The Svedberg equation does not, however, allow the molecular weight of a protein to be calculated in the absence of knowledge of the diffusion and the partial specific volume. If the protein is spherical, it has been shown that S is proportional to $m_i^{2/3}$, which allows the molecular weight of an unknown protein to be estimated by determination of S if the sedimentation coefficient of a protein with known molecular weight is also determined. This approach depends on both the known and unknown proteins having spherical shapes and similar hydrations and partial specific volumes.

Equilibrium and Approach to Equilibrium Sedimentation

If centrifugation is performed using lower forces, a somewhat different situation holds. Rather than all of the material being transported to a packed band at the bottom of the cell, an equilibrium is set up.

At equilibrium, $J_2 = 0$, and using Eqs. (4-33) and (4-33a) for S and D, one can write

$$\frac{C_2 W^2 X M(1 - V_2 p)}{N_0 f} = \frac{kT}{f} \frac{1 + d(\ln \gamma_2)}{d(\ln C_2)} \frac{dC_2}{dX} \tag{4-35}$$

where C_2 is the concentration of the macromolecule and γ_2 is the velocity gradient. This rearranges to yield

$$M\left[\frac{1 + d(\ln \gamma_2)}{d(\ln C_2)}\right]^{-1} = \frac{2RT}{(1 - V_2 p)W^2} \frac{d(\ln C_2)}{dX^2} \tag{4-36}$$

At low solute concentrations (conditions where the activity coefficient term disappears) this equation can be integrated using as a boundary condition the solute concentration at a reference point X_0, to give

$$C_2(X) = C_2(X_0) \exp\left[\frac{M(1 - V_2 p)W^2}{2RT(X^2 - X_0^2)}\right] \tag{4-37}$$

suggesting that a plot of ln C_2 versus X^2 gives as its slope the molecular weight. In practice, the absolute solute concentration gradient is hard to determine, and a form of this equation that employs concentration ratios is usually used:

$$\frac{C_b - C_m}{C_0} = \frac{M(1 - V_2 p)W^2(X_b^2 - X_m^2)}{2RT} \tag{4-38}$$

where C_0 is the uniform initial concentration and C_b and C_m are the equilibrium concentrations at the bottom and at the meniscus, respectively. Again, these ratios cannot be calculated unless the absolute concentration is known at some point for calibration. This is often overcome by using a rotor speed sufficiently high that at equilibrium the meniscus has zero solute concentration, allowing a reference point for detection of the concentrations. This approach is the *Yphanti's meniscus depletion method.*

In an alternative approach, Archibald noted that under an approach to equilibrium conditions, the flux, J_2, at the meniscus and at the bottom of the cell must be zero, hence

$$\frac{W^2 S}{D} = \frac{dC_2}{dX} \frac{1}{XC_2} \qquad \text{when } X = X_b \text{ or } X_m \tag{4-39a}$$

$$M_m = \frac{RT}{(1 - V_2 p)W^2} \left(\frac{dc_2}{dx}\right)_{X_m} \frac{1}{X_m C_m} \tag{4-39b}$$

$$M_b = \frac{RT}{(1 - V_2 p)W^2} \left(\frac{dc_2}{dx}\right)_{X_b} \frac{1}{X_b C_b} \tag{4-39c}$$

where C_m and C_b are the solute concentrations at the meniscus and at the bottom of the cell, respectively. The *Archibald method* gives two estimates of the molecular

weight from a single experiment: these estimates for a homogeneous solution are equal, and this approach is useful for establishing homogeneity.

The weight concentration of solute is used in the derivations described previously and as a result, similar to the discussion on light scattering, it is evident that sedimentation approaches give a weight-average molecular weight.

Problem of Partial Specific Volume

From the previous sections it can be seen that many of these methods depend on a knowledge of the partial specific volume, V_2, of the protein. The equations developed indicate that a factor $(1 - V_2)$ is usually involved, meaning that an error in V_2 results in a *larger* error in the molecular weight estimate. Since V_2 is usually of the order of 0.7, the estimate of molecular weight is subject to about three times the error of the estimate of V_2.

The partial specific volume of a protein can be obtained in one of three ways:

1. *From Density Measurements*: The partial specific volume can be experimentally estimated by measuring the density of a series of solutions with different weight fractions of protein solute,

$$V_2 \cong \frac{1}{p_0}\left(1 - \frac{p - p_0}{wp}\right) \tag{4-40}$$

where w is the weight fraction of the solute, p_0 the solvent density, and p the measured density in the presence of solute.

2. *From Amino Acid Composition*: Estimates of V_2 from amino acid composition are based on the equation

$$V_2 = \sum W_i V_i \tag{4-41}$$

where W_i is the weight fraction of each type of amino acid present in the protein and V_i is the partial specific volume of the individual amino acids. Although this calculation takes no account of volume changes between free amino acids in aqueous solvent compared to that which might pertain to the environment of the residue in the protein, or to volume changes that might result from specific interactions of amino acid side chains in a protein, or to changes that might result from electrostatic interactions, the method appears to work quite well.

3. *From Alternate Solvent Measurements*: This approach is based on the use of H_2O and D_2O as solvents in sedimentation equilibrium experiments. From earlier,

$$M_i(1 - V_2 p_{H_2O}) = \frac{2RT}{w^2}\left(\frac{d \ln C}{d^2}\right) \tag{4-42}$$

when the experiment is run with H_2O as solvent. If the protein is dissolved in D_2O, two events take place: (a) the molecular weight of the protein is increased by deute-

rium exchange and the new molecular weight is related to the molecular weight in H_2O by the ratio

$$k = \frac{M_{i_{D_2O}}}{M_{i_{H_2O}}} \tag{4-43}$$

and (b) the partial specific volume of the protein is decreased by the same relative amount. As a result,

$$k_M = \left(1 - \frac{V_2}{k} p_{D_2O}\right) = \frac{2RT}{w^2}\left(\frac{d \ln C}{d^2}\right) \tag{4-44}$$

These equations can be solved simultaneously to yield

$$V_2 = \frac{k - [(d \ln C/d^2)_{D_2O}/(d \ln C/d^2)_{H_2O}]}{p_{D_2O} - p_{H_2O}[(d \ln C/d^2)_{D_2O}/(d \ln C/d^2)_{H_2O}]} \tag{4-45}$$

To estimate V_2 all that is required is sedimentation equilibrium data with H_2O and with D_2O as solvent, and a value for k: k can be quite reliably estimated from a knowledge of the exchangeable protons in the protein, which are the one amide hydrogen per residue and to a minor extent some side-chain protons. In pure D_2O, $k = 1.0155$ for proteins in general. If the solvent contains D_2O at lower percentages, the value of k is reduced proportionally.

As discussed previously, the accuracy of molecular weight estimates obtained from techniques that require a knowledge of the partial specific volume of the protein depends on the accuracy of the partial specific volume. Most proteins that contain only amino acids (i.e., excluding glycoproteins, etc.) have partial specific volumes between 0.69 and 0.75 $cm^3\ g^{-1}$. If we assume an average value of 0.72 $cm^3\ g^{-1}$, the maximum error for V_2 can be 4.2%, which gives a potential error in the molecular weight estimate of approximately $\pm 12\%$.

GEL FILTRATION METHODS

In Chap. 2 we discussed the basic principles of gel filtration and some of the fundamental equations used to describe the elution behavior of proteins in a gel filtration experiment. The total volume, V_t, of a gel filtration column is expressed by

$$V_t = V_0 + V_{\text{gel matrix}} + V_s \tag{4-46}$$

When a solute is introduced to the column it partitions between the internal and external solvent regions and the distribution can be described by a partition coefficient (PC): The solute mass, SM, found in the internal regions is

$$\text{SM} = \text{PC} \cdot V_s \cdot C \tag{4-47}$$

where C is the solute concentration in the external regions. In terms of the elution volume of the solute, V_e,

$$V_e = V_0 + \text{PC} \cdot V_s \tag{4-48}$$

when a solute is *totally* excluded from the gel phase, PC = 0 and $V_e = V_0$. When the solute can diffuse freely with no size restrictions in the gel phase, PC = 1 and $V_e = V_0 + V_s$.

The elution position, V_e, can also be written in terms of K_{av} by rearranging Eq. (2-4):

$$K_{av} = \frac{V_e - V_0}{V_t - V_0} \tag{4-49}$$

which yields

$$V_e = V_0 + K_{av}(V_t - V_0) \tag{4-50}$$

For a particular gel filtration column,

$$\text{PC} = -A \log M + B \tag{4-51}$$

where A and B are constants according to the nature of the gel filtration material and column size and are established for a particular column by measuring PC for a series of proteins with known molecular weights. As discussed in Chap. 2, K_{av} is usually used in place of PC in such experiments. Figure 4-5 is a plot of K_{av} versus log M for a series of standard proteins. The molecular weight of an unknown protein

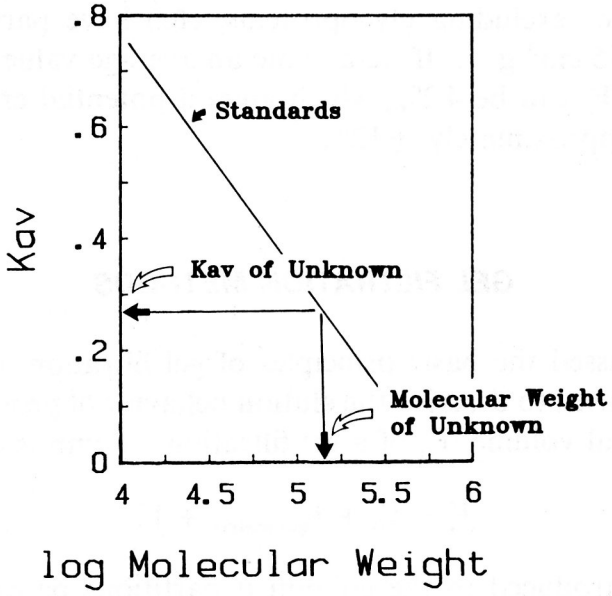

Figure 4-5 Relationship between K_{av} and log molecular weight for molecular weight standards in a gel filtration experiment. Standard proteins used were: ribonuclease (13.7 kD), chymotrypsinogen (25 kD), ovalbumin (43 kD), albumin (67 kD), aldolase (158 kD), catalase (232 kD), ferritin (440 kD), and thyroglobulin (669 kD).

is estimated by determining its V_e, and hence K_{av}, and using a calibration plot from the same column.

Although these equations fit experimental data for a reasonable number of proteins, a better fit to experimental data is obtained if the molecular weight in these equations is replaced by the effective hydrated radius, R_h.

$$PC = -A' \log R_h + B' \tag{4-52}$$

From this it is easy to understand why very large or small proteins deviate from the earlier relationship—they have hydrated radii that are not strictly related to their molecular weights. Similarly, glycoproteins or very asymmetric molecules also deviate from the earlier relationship because of "anomalous" effective hydrated radii.

The molecular weight of proteins can be expressed in terms of the molecular radius, a, as

$$M = Ka^p \tag{4-53}$$

where p varies with the shape of the protein. These relationships form the basis of using gel filtration data to determine shape parameters for proteins, as discussed in more detail in Chap. 13.

POLYACRYLAMIDE GEL ELECTROPHORETIC METHODS

Polyacrylamide gel electrophoresis (PAGE), because of its experimental ease, has become one of the commonest ways of determining the molecular weight of a protein. Because in general one can directly visualize protein after staining of a gel, it is also an excellent method for establishing a level of purity for a sample. Two types of PAGE are most often used: native and SDS-PAGE, which are based on different parameters and assumptions.

Native PAGE

Three parameters influence the movement of a protein in a native PAGE experiment: the charge on the protein (Q), the electric field (E), and the frictional coefficient (f). The charge on the protein depends on its pH and the pI, the electric field is experimentally set for a particular experiment, and the frictional coefficient depends on the pore size of the gel and the size and shape of the protein. The limiting velocity for movement of a particular protein in an experiment is $Q(E/f)$. The total gel concentration, T (which is the sum of the acrylamide and cross-linker concentrations), affects the porosity (and hence frictional effect) of the gel. Using a *relative* mobility, R_f, the mobility of the protein relative to some small molecule that encounters no sieving effect, we can use the *Ferguson equation*,

$$\log R_f = -K_r T + \log R_0 \tag{4-54}$$

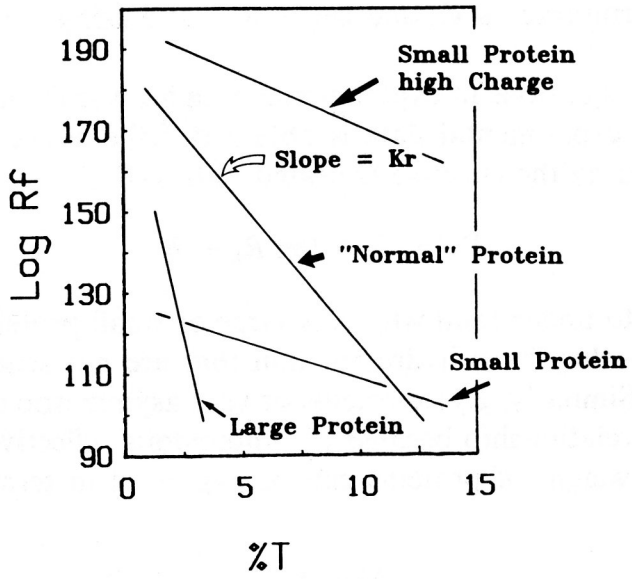

Figure 4-6 Ferguson plot. Typical lines are shown for the various types of protein referred to in the text.

where R_0 is the relative mobility of the protein in the absence of sieving (frictional) effects and K_r is the retardation coefficient. These two parameters are determined experimentally by determining R_f at a series of different gel concentrations and plotting a *Ferguson plot* (Fig. 4-6).

Different overall ranges of slopes of the Ferguson plot are expected for proteins of various types. In Fig. 4-6, line 1 typifies a large protein subject to considerable molecular sieving. Line 2 would be expected for the opposite extreme, a protein with low molecular size. Line 3 would be obtained for a protein of similar size to line 2 but with a much higher charge. Line 4 represents a typical medium-sized protein of normal charge.

The molecular weight of an unknown protein is obtained from the empirical linear relationship between K_r and molecular weight, M_i (see Fig. 4-7):

$$K_r = A + BM_i \tag{4-55}$$

As with gel filtration, A and B are system-dependent parameters determined by using proteins of known molecular weight. As with any method that depends on the properties of standards, the validity of a molecular weight determined by this method is based on the unknown protein having molecular characteristics (i.e., partial specific volume, shape) similar to the standard proteins.

SDS-PAGE

The same fundamental parameters (Q, E, and F) influence polyacrylamide gel electrophoresis in the presence of the anionic detergent sodium dodecylsulfate (SDS) as in its absence. The difference between the two approaches is that several variables

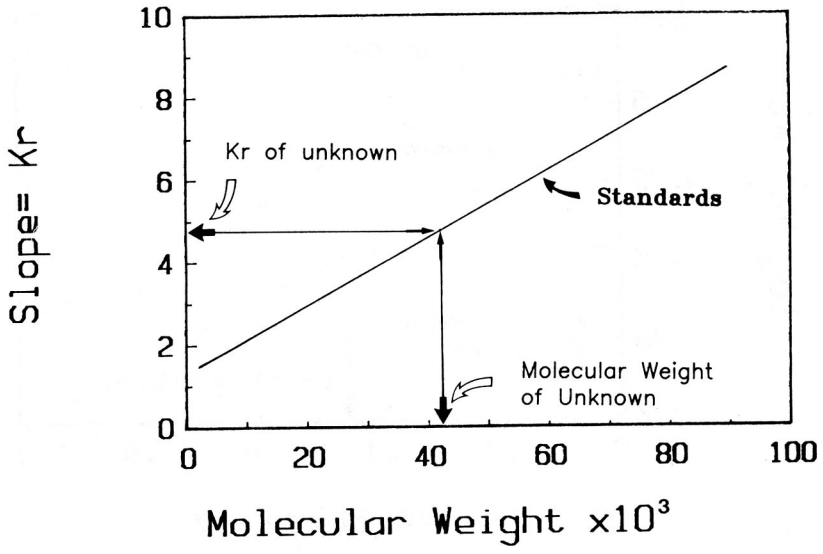

Figure 4-7 Determination of molecular weight from plots of the Ferguson plot versus molecular weight of the standard proteins. The molecular weight of an unknown protein is obtained by extrapolation of its Ferguson plot slope to intersect with the standard plot.

that can affect these parameters are more defined when SDS is present. It has been observed experimentally that most proteins bind a uniform amount of SDS per gram of protein (1.4 g) and that in the presence of a reducing agent (to break disulfide bonds), proteins tend to assume a uniform, "extended-rod" shape when SDS is present. As a result, proteins in the presence of SDS have the same charge per unit mass (and approximately the same charge per unit length). The result is that during electrophoresis only sieving effects are important in determining the mobility of a particular protein (as illustrated in Fig. 4-8). As a result, a plot of R_f versus log M_i can be used directly to give an estimate of the molecular weight of an unknown protein. The determinations based on SDS-PAGE depend on the validity of the assumptions made concerning the amount of SDS bound and the shape of the resultant denatured protein, and factors that affect these (such as glyco or lipo conjugates or a preponderance of basic or acidic side chains in the protein) lead to anomalous molecular weight estimates.

Native versus Denatured Molecular Weights. With several of the techniques that have been discussed here, a choice can be made between determining the molecular weight under native conditions or under denaturing conditions. A comparison of molecular weights determined under native or denaturing conditions can give significant information regarding any subunit structure the protein may have. In particular, sedimentation and gel filtration methods lend themselves to a direct comparison of native and denatured molecular weights since the appropriate solvent systems can readily be used. SDS-PAGE gives a denatured molecular weight of course, but since for the most reproducible results a reducing agent such as β-mercaptoethanol

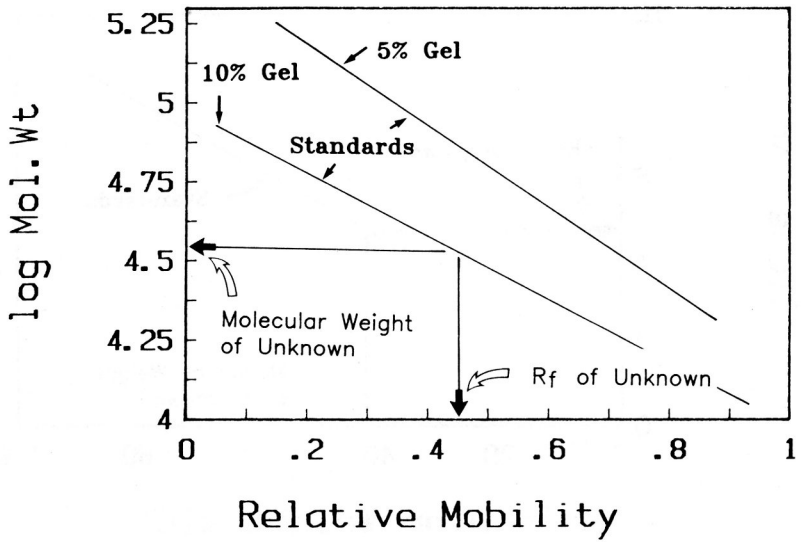

Figure 4-8 Calibration plots of molecular weight standards in SDS-PAGE: the effects of acrylamide concentration. In line 1 a 10% gel is used with molecular weight standards ranging from 14 kD (lysozyme) to 68 kD (bovine albumin). In line 2 a 5% gel is used, with molecular weight standards ranging from 29 kD (carbonic anhydrase) to 210 kD (myosin).

is used, this method does not give information concerning subunit molecular weights where the subunit contains two polypeptide chains covalently linked by one or more disulfides. Although it is possible to run SDS-PAGE in the absence of a reducing agent, it should be kept in mind that polypeptide chains that contain intra-peptide disulfide bonds run anomalously faster than if the intra-peptide disulfide is reduced. This is because such a molecule does not form an extended rod of the same size in the presence of SDS as the comparable reduced molecule and has a resultant smaller radius and value for f. The result is that such a protein has a high R_f value in a particular system. It is this that causes a protein with intra-peptide disulfide bonds to run above the diagonal in a two-dimensional diagonal map used to detect proteins with inter-polypeptide chains (see Fig. 4-9).

Determination of the Molecular Weight of Active Species. A frequently encountered problem in determining molecular weights is the question of the minimum molecular weight of a protein's active species. This is particularly difficult with a subunit containing protein, where the oligomer may itself undergo an association reaction that could possibly affect its activity. The problem is best illustrated by an example: Glutamate dehydrogenase is a hexamer that undergoes a concentration and regulatory ligand-dependent polymerization. The enzyme can utilize either glutamate or norvaline, and it was suggested that the hexamer preferentially uses glutamate, whereas norvaline is utilized only by higher polymers. The problem was resolved by determining the molecular weight of the active enzyme while it was catalyzing the oxidative deamination of either glutamate or norvaline, and showing that the molecular weights of the active species are in fact the same.

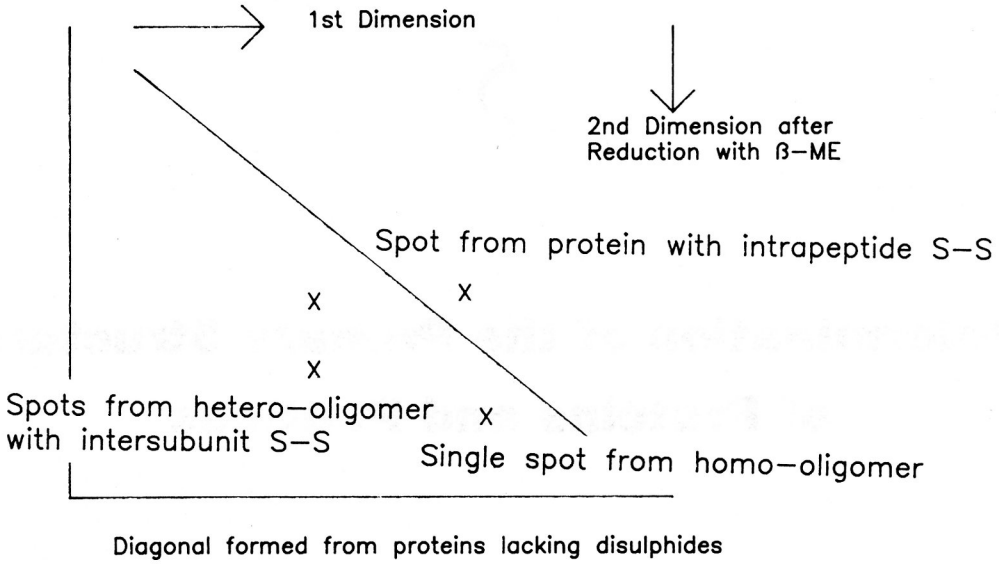

Figure 4-9 Schematic representation of a two-dimensional SDS-PAGE of disulfide-containing proteins.

Several of the methods for determining molecular weights lend themselves to the question: What is the molecular weight of the catalytically active species of an enzyme? Sedimentation methods have been used in what is known as *reactive enzyme centrifugation.* The protein is in effect sedimented through a solution containing all the necessary substrates for reaction to take place and the rate of sedimentation followed by following a reaction product rather than the protein. Thus the rate of sedimentation of the active enzyme is followed. Gel filtration methods have also been employed to allow determination of V_e for the reactive species of the enzyme and hence its molecular weight. The gel filtration matrix is equilibrated with buffer containing all the necessary substrates prior to the protein sample being loaded. V_e is determined by measuring the point at which reaction product is first detected in the eluent.

If one considers the basis of each of these techniques, it should be readily apparent that in circumstances where more than one molecular form of the protein is active, a molecular weight of the *smallest* active species is obtained from sedimentation methods, while the molecular weight of the *largest* active species is obtained using the gel filtration approach.