SELF-ASSOCIATING SYSTEMS IN THE ANALYTICAL ULTRACENTRIFUGE

BECKMAN COULTER
Life Sciences

SELF-ASSOCIATING SYSTEMS IN THE ANALYTICAL ULTRACENTRIFUGE

DONALD K. MCRORIE
PAUL J. VOELKER
BECKMAN COULTER

CONTENTS

FIGURES

TABLE

GLOSSARY

A	Absorbance
B	Second virial coefficient
c	Solute concentration (g/L)
c_r	Solute concentration at radial distance r
c_{r_0}	Concentration of the monomer at the reference radius r_0
E	Baseline offset
G	Gibbs free energy
H	Enthalpy
K_a	Association equilibrium constant
M	Monomer molecular weight
M_n, M_w, M_z	Number-, weight- and z-average molecular weights
$M_{w,app}$	Apparent weight-average molecular weight
R	Gas constant (8.314 J mol^{-1} K^{-1})
r	Radial distance from center of rotation
r_0	Reference radial distance
S	Entropy
T	Temperature in Kelvin
$\bar{v}$	Partial specific volume
ρ	Solvent density
ϕ	Isopotential partial specific volume
Σ	Summation symbol
X^2	Goodness of fit statistic
ω	Angular velocity (radians/second)

INTRODUCTION

Studying associating systems in the analytical ultracentrifuge allows one to fully characterize the thermodynamics of these association reactions at equilibrium. An advantage to the experimenter is that the macromolecule can be studied directly in solution under conditions of choice. By varying these conditions, several parameters can be determined: molecular weights, stoichiometry, association constants (tightness of binding), nonideality coefficients, and thermodynamic parameters such as the changes in Gibbs free energy, enthalpy, and entropy associated with binding (ΔG, ΔH, and ΔS). These parameters can be studied with both self-associating systems and associations of unlike species (hetero-associations). This primer will describe both simple and complex methods needed to examine self-associating systems in an approach requiring little or no advance knowledge of the characteristics of the interactions involved. Hetero-associations require a different approach and will be covered in a separate review.

The initial characterization of an association reaction involves several experiments to obtain information about the homogeneity of the system, monomer molecular weight, reversibility of the reaction, stoichiometry, and nonideality. These experiments require various run conditions and the use of data analysis procedures and diagnostic plots that provide estimates for these characteristics. The most straightforward initial approach is to measure molecular weights under denaturing vs. native conditions. A simple average molecular weight determination will reveal if an association is taking place and provide an initial estimate of stoichiometry. Further diagnostic plots may give better estimates of stoichiometry, reversibility, and nonideality.

When initial experiments yield sufficient information, more detailed analysis can be undertaken. Through the use of nonlinear regression techniques, a more accurate analysis of the system is accomplished by direct fitting of the primary data to a model describing the association. By comparing goodness of fit of the experimental data to the calculated data, a model best describing the association can usually be discerned. Care must be taken in statistical analysis to ensure that the fit of the data to the selected model is significantly better than to alternative models. If not, more experiments may be needed to distinguish between models. Fitting in this manner can give accurate determination of average molecular weights, (M_n, M_w, and M_z), association constants, and nonideality as measured by virial coefficients. Also, this procedure allows one to confirm stoichiometries estimated from other experiments and to incorporate baseline errors in the data.

In all calculations, several parameters are needed: ω^2, determined from the rotor speed; R, the gas constant; T, the temperature in Kelvin; $\bar{v}$, the partial specific volume determined from the sample composition or by measurement, and ρ, the density determined from the solvent composition. The last two are the most variable in a system and can therefore lead to the greatest error in calculations.

From a single experiment, only the buoyant molecular weight is measurable directly in the analytical ultracentrifuge. This value, $M(1 - \bar{v}\rho)$, is the molecular weight of the sample corrected by a buoyancy factor due to displaced solvent. In the case of multiple species, $M(1 - \bar{v}\rho)$ will be a statistical average of the molecular weights of all species present in solution. Different molecular weight averages can be determined by various treatments of sedimentation equilibrium data. More detailed analyses of the associations as described above require that the data from several experiments be examined simultaneously.

Three parameters necessary for the analysis of self-associating systems are not determined by run conditions. These are $\bar{v}$, ρ, and monomer molecular weight. Calculation of these parameters is needed to begin an analysis. Calculations of $\bar{v}$ and ρ are described in Appendices B and C, respectively. If sample composition is not known, monomer molecular weight can be determined experimentally from a run in denaturing conditions. Also, alternative methods such as mass spectrometry give accurate subunit molecular weight determinations.

MOLECULAR WEIGHT DETERMINATION BY SEDIMENTATION EQUILIBRIUM

When sedimentation and diffusion come to a state of equilibrium, no apparent movement of solute occurs. The equilibrium concentration distribution is dependent on the buoyant molecular weight, $M(1 - \bar{\nu}\rho)$; angular velocity, ω^2; and temperature. Since the concentration distribution is dependent on the buoyant molecular weight, it is obvious that accurate values of $\bar{\nu}$ and ρ are necessary for the determination of molecular weight from equilibrium conditions.

From the Lamm (1929) equation describing movement of molecules in a centrifugal field, the following equation can be derived for a single, thermodynamically ideal solute:

Equation 1

$$\frac{\ln(c_r)}{r^2} = \frac{M(1-\bar{\nu}\rho)\omega^2}{2RT}$$

A plot of $\ln(c)$ vs. r^2 will yield a straight line with a slope proportional to M. For nonideal or associating systems, or when multiple species are present, a straight line is not obtained, and more rigorous analysis is needed. Nevertheless, as shown in Figure 1, this analysis provides a first approximation of M and can indicate thermodynamic nonideality or polydispersity in the sample.

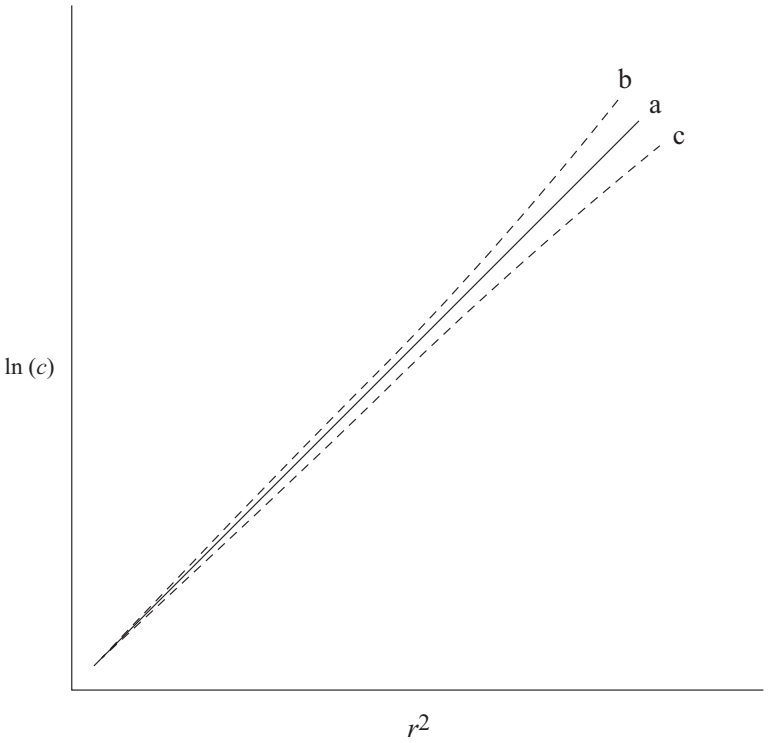

Figure 1. *Graph of ln(c)* vs. r^2 showing curves from ideality

From the graph it is apparent that nonideality and polydispersity have opposite effects so that the presence of both in a sample may be offsetting and thus not discernible. It should also be noted that small amounts of aggregation, such as 10% or less of a sample present as a dimer, will not be detectable by this method.

The slope of the ln(*c*) *vs.* r^2 plot is one of a group of statistical averages of the molecular weight and is known as the weight-average molecular weight, M_w. Other treatments of sedimentation equilibrium data allow determination of number-average (M_n), *z*-average (M_z), and higher order molecular weight averages (Correia and Yphantis, 1992) (see Figure 2).

A

$$M_n = \frac{\sum_i n_i M_i}{\sum_i n_i} = \frac{\sum_i c_i}{\sum_i c_i / M_i} \quad \text{(number-average)}$$

$$M_w = \frac{\sum_i n_i M_i^2}{\sum_i n_i M_i} = \frac{\sum_i c_i M_i}{\sum_i c_i} \quad \text{(weight-average)}$$

$$M_z = \frac{\sum_i n_i M_i^3}{\sum_i n_i M_i^2} = \frac{\sum_i c_i M_i^2}{\sum_i c_i M_i} \quad (z\text{-average})$$

B

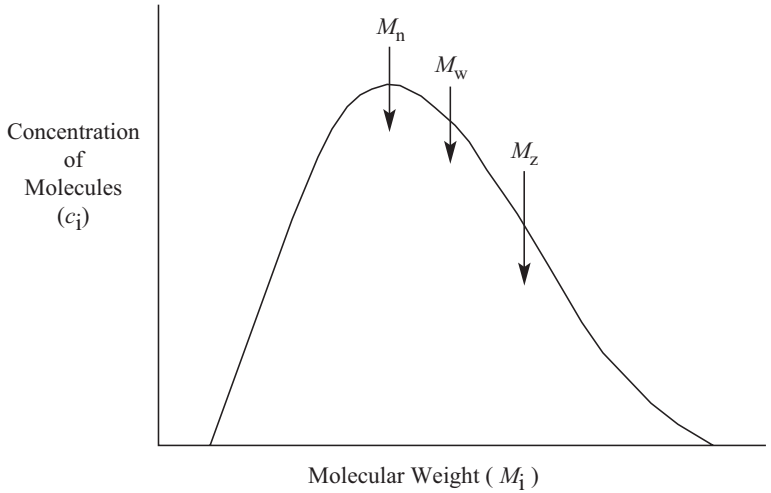

Figure 2. Molecular weight averages. a) Calculation of molecular weight averages; b) graph showing distribution in a polydisperse system.

If the solution is polydisperse, then sedimentation equilibrium experiments yield an average molecular weight, not that of any single component. M_n , M_w, and M_z give increasing significance, respectively, to those components in a mixture with the highest molecular weights. Thus, for a polydisperse system, $M_n < M_w < M_z$. If a solution is monodisperse, then $M_n = M_w = M_z$. A comparison of the molecular weight averages can therefore provide a good measure of homogeneity.

If the plot of ln(*c*) *vs.* r^2 is nonlinear, tangents to the curve yield a weight-average molecular weight for the mixture of species present at each radial position. In this manner the user can obtain molecular weight

as a function of increasing solute concentration moving down the cell. Overlaying plots of data from samples of different starting concentrations will provide information about the presence of more than one species in the cell, the ability to distinguish polydisperse and self-associating systems, and, in the latter case, an estimate of the monomer molecular weight and stoichiometry.

INITIAL EXPERIMENTAL APPROACH TO ANALYZING SELF-ASSOCIATING SYSTEMS

An initial level of analysis involves a characterization of the sample using diagnostic graphs. This level of analysis is qualitative and simply determines if the sample is homogeneous, ideal, and whether or not a self-association is occurring. In many cases, estimates of monomer molecular weight and stoichiometry are also possible at this level, but quantitative determination of thermodynamic parameters will require more rigorous nonlinear least-squares fitting procedures that will be described in more detail in a following section. Qualitative analyses do, however, provide excellent starting guesses for models in more complex analyses.

To determine these parameters, data from runs under both denaturing and nondenaturing conditions need to be compared. The two most commonly used denaturants are 8 M urea and 6 M guanidine-HCl. Both have effects on v which should be taken into account during data analysis. In addition, data from samples run at multiple starting concentrations and multiple rotor speeds will be required. This approach is described by Laue (1992).

Typically, a range of starting concentrations with absorbances from 0.1 to 1.0 is employed. The simplest way to obtain this range of concentrations is to perform a serial dilution. Generally, the absorbance variation is scaled for a single wavelength, but if the extinction coefficients are known for several wavelengths, this added information can be used to convert all absorbances to the same concentration scale using the Beer-Lambert Law,* which permits an even wider range of concentrations to be used.

Rotor speeds are chosen to straddle the estimated optimum rotor speed for the sedimentation equilibrium run. Figure 3 shows optimum speeds for an equilibrium run if either the molecular weight or sedimentation coefficient can be estimated for the sample.

---

* Beer-Lambert Law: $A = \log(I_0/I) = \varepsilon cl$, where I_0 = intensity of the incident radiation; I is the intensity of light transmitted through a pathlength l in cm, containing a solution of concentration c moles per liter; ε is the molar extinction coefficient with units liter mole^{-1} cm^{-1}; A is the absorbance.

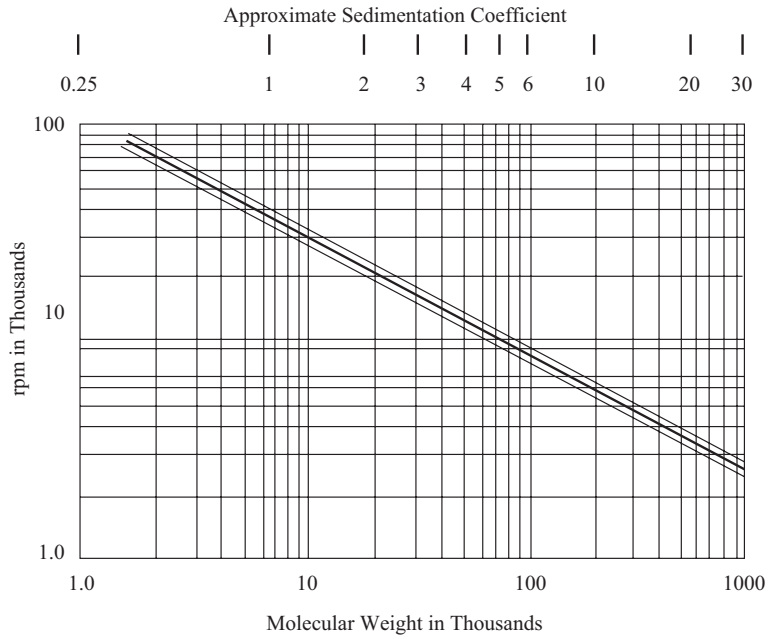

Figure 3. Optimum speeds for equilibrium runs if either molecular weight or sedimentation coefficient can be estimated. (Reprinted from Chervenka, 1970.)

A set of at least three speeds is chosen to yield significantly varied data for the diagnostics (Laue, 1992). For three speeds, the ratio of the squares of the two slower speeds should be 1.4 or greater, and the ratio of the squares of the fastest and slowest speeds should be 3 or greater. For example, if an optimum rotor speed is estimated as 20,000 rpm, a good choice of three speeds would be 16,000, 20,000, and 30,000 rpm [$(20{,}000/16{,}000)^2 = 1.56$ and $(30{,}000/16{,}000)^2 = 3.52$]. Data should be acquired at the lowest speed first, then at progressively higher speeds to minimize the time to reach equilibrium. If data need to be acquired at a slower speed, it is advantageous to stop the rotor and gently shake the cells rather than simply lowering the speed, because redistribution from diffusion is quite slow.

MOLECULAR WEIGHT *VS.* CONCENTRATION DIAGNOSTIC GRAPHS

The data collected as described in the preceding section allow the apparent molecular weight to be plotted as a function of sample loading concentration and rotor speed. Since the sample was run under denaturing conditions, M_w can also be calculated relative to monomer molecular weight. These graphs provide an initial characterization of the system with respect to: 1) homogeneity, 2) nonideality, 3) self-association (with limited information about stoichiometry), and 4) polydispersity.

Laue (1992) uses a plot of the apparent weight-average molecular weight vs. the midpoint absorbance (Figure 4). If the sample obeys the Beer-Lambert Law, absorbance and concentration will be proportional.

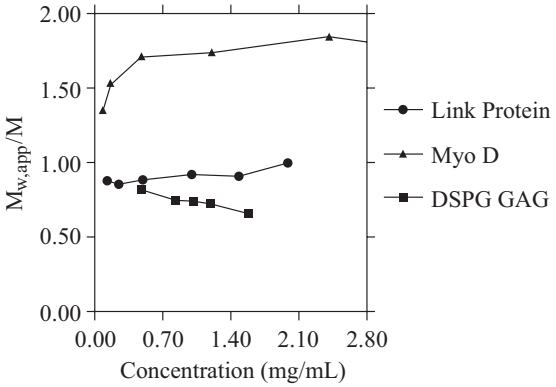

Figure 4. Diagnostics graph providing a qualitative characterization of the solution behavior of macromolecules. (Reprinted from Laue, 1992.)

This graph is useful in the detection of three possible conditions in the sample run with multiple concentrations. If the molecular weight remains constant with changing absorbance (concentration), a single ideal species is indicated. If the molecular weight decreases with increasing absorbance, this indicates thermodynamic nonideality. Finally, if the molecular weight increases with increasing concentration, a self-association may be occurring. In this last case, if a wide enough concentration range is examined, the molecular weight at low concentration will approach that of the

smallest species in the reaction, and at higher concentrations it will approach that of the largest species. If the monomer molecular weight has been determined under denaturing conditions, this provides estimates of the stoichiometry of these limiting species. As is the case for plots of ln(*A*) *vs.* r^2, association and nonideality have opposing effects. So, if both conditions are present, determination of an upper limit for oligomerization is difficult. Similar problems may arise due to limited solubilities at higher concentrations.

More information can be obtained from plots of molecular weight vs. concentration from a single sample by examining the apparent weightaverage molecular weight as a function of increasing radius. One method is to use nonlinear regression techniques to calculate parameters that allow determination of molecular weight for each radial position in the cell, then to graph the molecular weight vs. the corresponding absorbance value (Formisano *et al.*, 1978). Another method is to take subsets of the data, determine molecular weight from linear regression analysis with a plot of ln(*A*) *vs.* r^2, and finally graph the molecular weight against the midpoint absorbance of the subset. Other methods are used in some cases to avoid the systematic errors that occur in some calculations (Dierckx, 1975). M_w *vs.* concentration calculations provide information similar to that obtained from a single point per sample, but, in addition, a diagnostic graph that distinguishes between self-association and polydispersity may be obtained (Figure 5).

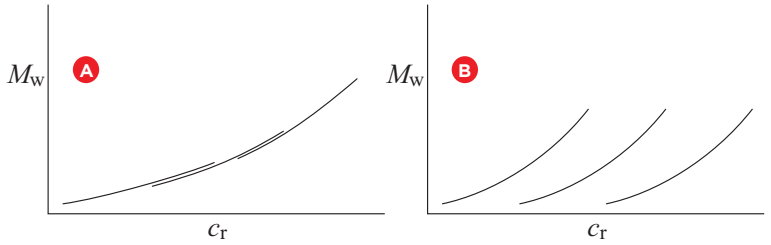

Figure 5. Multiple data sets graphed in terms of molecular weight vs. concentration. a) Self-associating system; b) polydisperse system.

For a self-associating system, the apparent weight-average molecular weight will increase with increasing concentration, and the plots of M_w *vs.* absorbance will coincide. As with the previous diagnostic graphs, the limiting molecular weights will approach that of the monomer at the lower concentrations and that of the largest species at the higher concentrations. Nonideality will cause a downward trend in the slope and will prevent an accurate estimation of the stoichiometry from this plot.

In the case of polydispersity, however, the M_w *vs.* absorbance plots will not coincide but will be displaced to the right with increasing sample starting concentration. Nonoverlapping weight-average molecular weight distributions will be observed for different loading concentrations. This observation is due to the fact that the same molecular weight distribution is present regardless of sample concentration.

Polydispersity and reversibility of a self-association reaction can be confirmed by graphs of apparent molecular weight as a function of rotor speed (Figure 6).

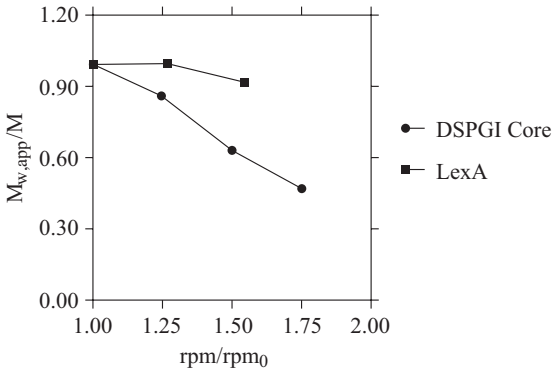

Figure 6. Diagnostics graph providing a qualitative characterization of the solution behavior of macromolecules. (Reprinted from Laue, 1992.)

For both homogeneous, noninteracting samples and homogeneous, self-associating samples, molecular weight is independent of rotor speed as long as all associating species are detectable in the concentration gradient. Polydisperse samples, however, display a systematic decrease in molecular weight with increasing rotor speed.

Caution is required due to the opposing effects of association and nonideality on concentration dependence of molecular weight. The presence of both phenomena can produce apparently ideal behavior. The effect of nonideality may not be apparent except in higher concentrations. In this case, the molecular weight will appear to decrease with increasing concentration. This situation makes estimation of association stoichiometry difficult. These diagnostic plots should only be considered qualitative and not used for numeric determination of any molecular weight, association, or nonideality parameters.

NONLINEAR LEAST-SQUARES ANALYSIS

Least-squares methods are one way to obtain a statistical fit of the experimental data to a proposed model and to obtain best estimates for unknown parameters (Johnson, 1992; Johnson and Faunt, 1992; Johnson and Frazier, 1985; Johnson *et al.,* 1981). The major advantage of these methods is that data can be analyzed directly without transformation. Also, more complex models can be tested when obvious differences between experimental and fitted data exist. Computer analysis has greatly facilitated these methods. Without the calculating power of the computer, only the conventional graphical analyses with their inherent assumptions and approximations would be possible. Three basic features are needed for simple nonlinear least-squares analysis: 1) an algorithm for calculating least-squares, 2) a mathematical model to describe the system, and 3) statistical analysis to measure goodness of fit of a proposed fitted model to the experimental data. A practical approach to data analysis using these methods and the Beckman Optima XL-A Data Analysis Package is described in Appendix A.

LEAST-SQUARES METHODS

In this analysis, a series of curves is calculated to locate a "best" fitting model of the data. Each iteration leads to a better approximation of the curve parameters until the approximations converge to stable values for the parameters being varied. For least-squares analysis, the differences between the fitted function and the experimental data are squared and summed, and the parameters varied so as to minimize this sum. Ideally, the reduction continues until a global minimum is reached. Minimization of least-squares does not always provide the correct set of model parameters. Therefore, additional statistical and graphical analysis is usually needed in addition to curve fitting techniques.

The graphs in Figure 7 show the sum of squares minima in relation to two parameters.

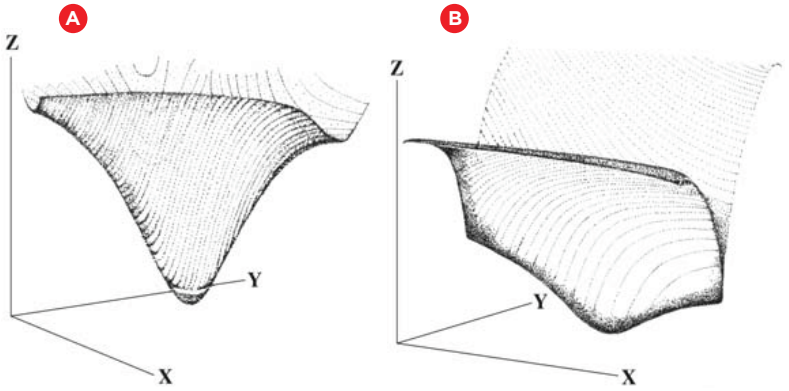

Figure 7. Error surface graphs showing the sum of squares graphed on *z*-axis. Two variables are shown on the *x*- and *y*-axes. a) Well-defined minimum; b) well-defined for y but not *x*.

The values for the two parameters are on the *x*- and y-axis, respectively, and for the sum of squares on the *z*-axis. In Figure 7a the error space is well defined for both parameters, and the algorithm would be able to calculate them to a high precision. Figure 7b, however, shows an example where the parameter plotted on the y-axis is well defined, but the parameter plotted on the x-axis is not. In this case, the algorithm would be able to calculate the y parameter to a high precision, but would have more difficulty calculating the x parameter to the same precision. In some cases, the error space can be flat or contain several minima that may yield different answers depending on starting guesses.

Many numerical algorithms are available for determining parameters by least-squares and are too numerous to be listed here. The Marquardt method (1963) is the most commonly used. It combines the advantages of two other methods, Gauss-Newton and Steepest Descent (Bevington, 1992), to obtain a more robust convergence. The Nelder-Mead algorithm (1965), also known as the Simplex method, is a geometric, rather than a numeric, procedure. The XL-A Data Analysis Package has incorporated a modification of the Gauss-Newton method developed by Johnson *et al.* (1981) in the multifit program analyzing multiple data sets simultaneously. Also accessible is the Marquardt algorithm for single data file analysis.

MATHEMATICAL MODELS

Single ideal species

When a method of least-squares analysis is used, a mathematical equation, or model, describing the distribution of the macrosolute in the cell is needed for the fitting procedure. The initial model chosen can be one that seems best to describe the system as discerned from the diagnostic plots described earlier. From goodness of fit analyses, alternative models are then analyzed to find the best description of the experimental data.

As described previously, the Lamm partial differential equation describes all movement of molecules in a centrifugal field. At equilibrium, no apparent movement of solute occurs due to the equalization of sedimentation and diffusion. From this observation, an exponential solution to the Lamm equation can be derived (Equation 2). This equation, with cr and radius as the dependent and independent variables, respectively, is directly related to the data as it is obtained from the analytical ultracentrifuge.

Equation 2

$$c_r = c_{r_0} e^{\left[\frac{\omega^2}{2RT} M(1-\bar{v}\rho)(r^2 - r_0^2)\right]}$$

where c_r = concentration at radius r
c_{r_0} = concentration of the monomer at the reference radius r0
ω = angular velocity
R = gas constant (8.314×10^7 erg/mol·K)
T = temperature in Kelvin
M = monomer molecular weight
$\bar{v}$ = partial specific volume of the solute
ρ = density of solvent.

Ultracentrifuge data can be fitted to this equation using optical absorbance in place of concentration, provided the sample obeys the Beer-Lambert Law across the full range of absorbance. Final results are usually converted back to concentration.

Equation 2 describes distribution of a single ideal species at equilibrium. Fitting the data to this model using nonlinear regression analysis yields an apparent weight-average molecular weight for all solutes in the cell when the baseline offset is constrained to zero. Including the baseline offset will result in determination of a z-average molecular weight. Recalling from Figure 2 that these average molecular weights are expressed in terms of concentration, the molecular weights are not well defined averages if the extinction coefficients are not known for all components because the data used are absorbance values.

Baseline offset results from a difference in the absorbance between the reference and the solvent in which the sample is dissolved. Usually, extensive dialysis procedures are used to minimize the difference. If, however, a difference remains, the baseline term must be included in the mathematical model. With sufficient data, the baseline can be varied as an additional parameter, but the calculated values of other parameters such as molecular weight are affected greatly by the baseline value. The baseline value can be determined experimentally by a high-speed run where the meniscus is depleted of all sample and the absorbance read directly from the data near the meniscus. The experimental approach is limiting when the meniscus cannot be depleted.

Self-association

The model equation for a self-associating system is similar to that of a single ideal species except that the total absorbance at a given radius is the sum of absorbances of all species at that radius. Each term of the summation will be a distribution function similar to that for a single ideal species. Take, for example, the simple equilibrium:

$$\text{monomer} \leftrightarrow n\text{-mer}$$

At equilibrium, the total absorbance as measured in the analytical ultracentrifuge can be shown as the sum of two species (Figure 8).

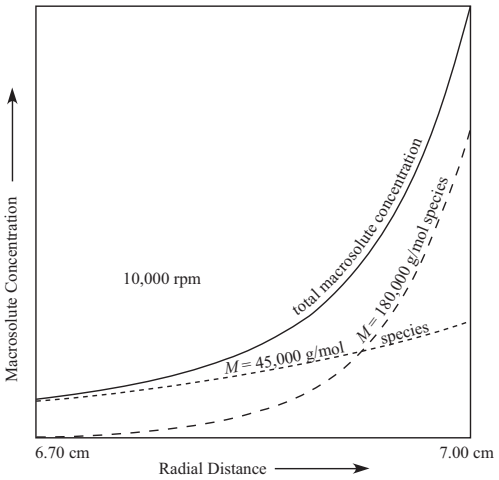

Figure 8. Concentration distributions of equilibrium in the analytical ultracentrifuge for a monomer-tetramer model showing monomer (----), tetramer (- - -), and total concentrations (———).

The equation that describes the total macrosolute distribution for this monomer–*n*-mer equilibrium is as follows:

Equation 3

$$c_{\text{total}} = c_{\text{monomer},r_0} e^{\left[\frac{\omega^2}{2RT} M(1-\bar{v}\rho)(r^2 - {r_0}^2)\right]} + c_{\text{n-mer},r_0} e^{\left[\frac{\omega^2}{2RT} nM(1-\bar{v}\rho)(r^2 - {r_0}^2)\right]}$$

Each exponential in the summation describes the equilibrium distribution of an individual species in the equilibrium, the first being the monomer and the second the *n*-mer. The stoichiometry is set in the model by an integer value for *n*, such that the n-mer molecular weight is *n* times the monomer molecular weight, *M*.

For the monomer–*n*-mer equilibrium reaction, the association constant is:

Equation 4

$$K_a = c_{\text{n-mer}}/(c_{\text{monomer}})^n$$

Substituting into equation 3, gives a new model to solve directly for K_a without the $c_{\text{n-mer}}$ term:

Equation 5

$$c_{\text{total}} = c_{\text{monomer},r_0} e^{\left[\frac{\omega^2}{2RT} M(1-\bar{v}\rho)(r^2 - {r_0}^2)\right]} + K_a (c_{\text{monomer},r_0})^n e^{\left[\frac{\omega^2}{2RT} nM(1-\bar{v}\rho)(r^2 - {r_0}^2)\right]}$$

Rearranging the equation constrains the $c_{\text{monomer},r0}$ and K_a terms to positive numbers:

Equation 6

$$c_{\text{total},r} = e^{\left[\ln(c_{\text{monomer},r_0}) + \frac{\omega^2}{2RT} M(1-\bar{v}\rho)(r^2 - {r_0}^2)\right]} + e^{\left\{n\left[\ln(c_{\text{monomer},r_0})\right] + \ln(K_a) + \frac{\omega^2}{2RT} nM(1-\bar{v}\rho)(r^2 - {r_0}^2)\right\}}$$

Similar models for more than two species in equilibrium can be derived by using the same method of describing the total concentration as the summation of all species present. Usually, not more than three species can be distinguished by analytical ultracentrifugation data However, a model is included in the data analysis software that describes the equilibrium distribution of up to four ideal species (equation 7):

Equation 7

$$c_r = c_{\text{monomer},r_0} e^{\left[\frac{(1-\bar{\nu}\rho)\omega^2}{2RT} M(r^2 - {r_0}^2)\right]} + (c_{\text{monomer},r_0})^{n_2} K_{a,2} e^{\left[\frac{(1-\bar{\nu}\rho)\omega^2}{2RT} n_2 M(r^2 - {r_0}^2)\right]} + (c_{\text{monomer},r_0})^{n_3} K_{a,3} e^{\left[\frac{(1-\bar{\nu}\rho)\omega^2}{2RT} n_3 M(r^2 - {r_0}^2)\right]} + (c_{\text{monomer},r_0})^{n_4} K_{a,4} e^{\left[\frac{(1-\bar{\nu}\rho)\omega^2}{2RT} n_4 M(r^2 - {r_0}^2)\right]} + E$$

where c_r = concentration at radius r
$c_{monomer,r0}$ = concentration of the monomer at the reference radius r_0
M = monomer molecular weight
n_2 = stoichiometry for species 2
$K_{a,2}$ = association constant for the monomer–n-mer equilibrium of species 2
n_3 = stoichiometry for species 3
$K_{a,3}$ = association constant for the monomer–n-mer equilibrium of species 3
n_4 = stoichiometry for species 4
$K_{a,4}$ = association constant for the monomer–n-mer equilibrium of species 4
E = baseline offset

Rearranging as before to constrain $c_{monomer,r0}$ and K_a values:

Equation 8

$$c_r = c_{\text{monomer},r_0} e^{\left[\frac{(1-\bar{\nu}\rho)\omega^2}{2RT} M(r^2 - {r_0}^2)\right]} + (c_{\text{monomer},r_0})^{n_2} K_{a,2} e^{\left[\frac{(1-\bar{\nu}\rho)\omega^2}{2RT} n_2 M(r^2 - {r_0}^2)\right]} + (c_{\text{monomer},r_0})^{n_3} K_{a,3} e^{\left[\frac{(1-\bar{\nu}\rho)\omega^2}{2RT} n_3 M(r^2 - {r_0}^2)\right]} + (c_{\text{monomer},r_0})^{n_4} K_{a,4} e^{\left[\frac{(1-\bar{\nu}\rho)\omega^2}{2RT} n_4 M(r^2 - {r_0}^2)\right]} + E$$

The stoichiometries n_2-n_4 are defined by the user, and the respective association constants are for the monomer–*n*-mer equilibrium of each aggregate. Association constants for *n*-mer–*n*-mer equilibria can be calculated from respective monomer–*n*-mer constants, but these mechanisms of reaction cannot be confirmed from equilibrium concentration data. For example, with a monomer-dimer-trimer equilibrium, the model (equation 8) will calculate the association constants for the monomer-dimer and monomer-trimer equilibria. If there is evidence that a dimer-trimer equilibrium is present, it can be calculated:

Equation 9A

$$K_{1\text{-}2} = c_{\text{dimer}}/(c_{\text{monomer}})^2,\ K_{1\text{-}3} = c_{\text{trimer}}/(c_{\text{monomer}})^3$$

Equation 9B so

$$K_{2\text{-}3} = c_{\text{trimer}}/c_{\text{monomer}} \bullet c_{\text{dimer}} = K_{1\text{-}3}/K_{1\text{-}2}$$

Usually, association constants are expressed in terms of concentration. Since data from the analytical ultracentrifuge are in absorbance units, some assumptions are usually made in the calculation of these constants. The extinction coefficient for an *n*-mer in the monomer–*n*-mer equilibrium is assumed to be n times that of the monomer. So, if an association constant is calculated in terms of the absorbance data, conversion to one based on concentration must include this assumption (Becerra *et al.*, 1991; Ross *et al.*, 1991). Equation 10 shows this conversion if the extinction coefficient of the monomer is known:

Equation 10

$$K_{1\text{-}2,\text{conc}} = c_2/{c_1}^2 = K_{1\text{-}2,\text{abs}}\varepsilon l/2;$$
$$\text{Beer-Lambert Law: } A = \varepsilon c l \text{ or } c = A/\varepsilon l$$

Similarly, for a monomer-trimer system the conversion is:

Equation 11

$$K_{1\text{-}3,\text{conc}} = c_3/{c_1}^3 = K_{1\text{-}3,\text{abs}}\varepsilon^2 l^2/3$$

Nonideality

Nonideal behavior of an associating system resulting from charge or crowding can be incorporated into the model (equation 12). The nonideality is described by the second virial coefficient, *B* (Haschmeyer and Bowers, 1970; Holladay and Sophianopolis, 1972, 1974).

Equation 12

$$c_{r,total} = e^{\left[\ln(c_{monomer,r_0}) + \frac{(1-\bar{\nu}\rho)\omega^2}{2RT} M(r^2 - r_0^2) - BM(c_{total,r} - c_{total,r_0})\right]}$$
$$+ e^{\left[\ln(c_{monomer,r_0}) + \ln(K_{a,2}) + \frac{(1-\bar{\nu}\rho)\omega^2}{2RT} n_2 M(r^2 - r_0^2) - BM(c_{total,r} - c_{total,r_0})\right]}$$
$$+ e^{\left[\ln(c_{monomer,r_0}) + \ln(K_{a,3}) + \frac{(1-\bar{\nu}\rho)\omega^2}{2RT} n_3 M(r^2 - r_0^2) - BM(c_{total,r} - c_{total,r_0})\right]}$$
$$+ e^{\left[\ln(c_{monomer,r_0}) + \ln(K_{a,4}) + \frac{(1-\bar{\nu}\rho)\omega^2}{2RT} n_4 M(r^2 - r_0^2) - BM(c_{total,r} - c_{total,r_0})\right]} + E$$

This model can also be used for fitting all previously described equilibria with appropriate constraints. For example, setting the virial coefficient value to zero effectively removes all nonideality terms, and the model is the same as that described for an ideal self-associating system. Also, constraining any of the association constant values to an extremely small number, e.g., 1×10^{-20}, effectively removes the exponent term describing the distribution of the corresponding n-mer. This constraint can make the same model usable for one to four species in the cell.

Fitting data to a model

Once a model has been chosen along with the algorithm for analysis, the user is ready to begin the fitting procedure. The basis for fitting is to set initial guesses for all parameters in the model and either to constrain these to the set value or to allow them to float in the least-squares calculation. The constrain/vary choice will be dictated by experimental conditions and the thermodynamic parameters to be determined. Certain parameters are determined by experimental conditions and must be provided prior to the fitting procedure. These include angular velocity (ω2) determined from the rotor speed, temperature, and characteristics of the solute and solvent ($\bar{\nu}$ and ρ). A reference radius will also be chosen for the fitting procedure. In all cases, the absorbance or concentration at this reference radius will be allowed to float in the calculations. All other parameters are chosen by the user.

The amount of data will be the factor limiting the complexity of the models that can be distinguished statistically. For example, with a single data set, floating variables should be limited to two. Even then, confidence in answers obtained is not always good enough. So, as a general rule, multiple data sets with varying conditions (usually speed and starting sample concentration) should be used. In this case, an algorithm is needed that can fit to the multiple data sets simultaneously.

Analysis of multiple files separately can also be used to check the reversibility of an associating system. If the same association constant is obtained from runs at multiple speeds and starting concentrations, the association is reversible.

Confidence intervals

Confidence intervals are a measure of the precision of individual parameters based on a single set of data. However, the interval determined for a parameter can also serve as a measure of the accuracy of the estimated parameter (Johnson and Faunt, 1992). A number of methods of varying complexity exist to evaluate confidence intervals and, thus, the validity of the approximation (Johnson and Faunt, 1992; Straume and Johnson, 1992b). Confidence intervals in most cases are not symmetrical, so the magnitudes of the two intervals from a determined parameter are not always equivalent. This asymmetry makes these intervals more realistic than symmetrical standard deviation determinations or linear approximations of confidence intervals that are symmetrical around the determined parameter.

Contour maps

Contours are a method of profiling a three-dimensional surface in a twodimensional format. In this way, the user can visualize the sum of squares error space in relation to two parameters (Bates and Watts, 1988; Johnson and Faunt, 1992). Using the error surface maps illustrated in the leastsquares section, one can calculate confidence intervals corresponding to the magnitude of the sum of squares on the *z*-axis. Planes drawn parallel to the *x*-*y* axes at increasing magnitudes along the *z*-axis will intersect the error surface. When viewed down the *z*-axis, the intercepts will appear as concentric contours. The contours will show graphically the magnitude of the confidence intervals on the *x*- and *y*-axes in relation to the minimum of the sum of squares.

Several methods exist for estimating the contours. The actual contours for nonlinear regression are asymmetrical, but to save computer time, in many cases, linear or symmetric approximations are used to calculate elliptical estimations (Figure 9).

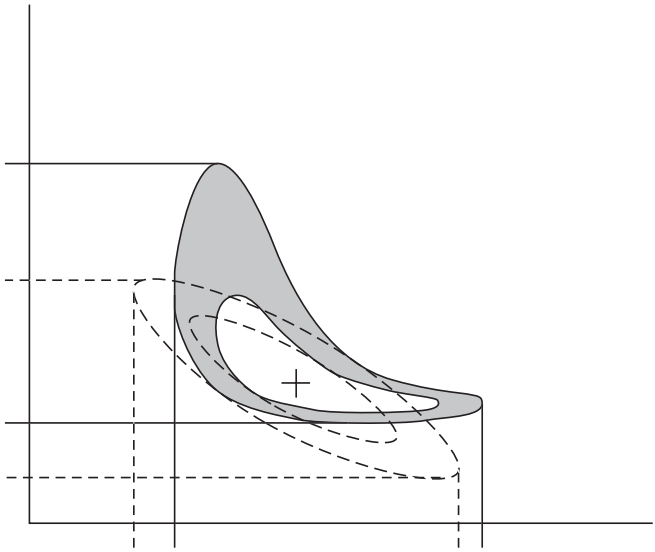

Figure 9. Actual contour maps (———) in relation to linear approximations (- - -) to demonstrate differences in confidence intervals; + represents the least-squares estimate. (Redrawn from Bates and Watts, 1988, with permission of John Wiley & Sons, Inc.)

The spacing and shape of the contours can indicate how well the two parameters being examined are defined by the error space. Closer and less elongated (rounder) contours indicate better defined parameters.

GOODNESS OF FIT

The first fitting attempt for any experiment is usually to the single ideal species model. The M_w and M_z determined will give the first indication of possible inappropriateness of the model for the experimental data. If these molecular weight averages are not equivalent, multiple species are indicated. If a monomer molecular weight is known from previous experiments, M_w will provide the first indication of the stoichiometry of an associating system. But, as mentioned previously, nonideality will have opposite effects on the analysis, and this fact should be considered in interpretation of results.

RESIDUALS

The most sensitive graphical representation for goodness of fit, and the best indicator of possible alternative models, is the residual plot. Residuals are the difference between each experimental data point and the corresponding point on the curve calculated from the model equation. Figure 10 shows an example of the desired residuals with a fit from a plausible model.

A random distribution of points about the zero value is a desired diagnostic for a good fit. Also shown are typical patterns of systematic errors characteristic of associating and nonideal systems. Thus, the patterns of residual plots can suggest additional models for fitting.

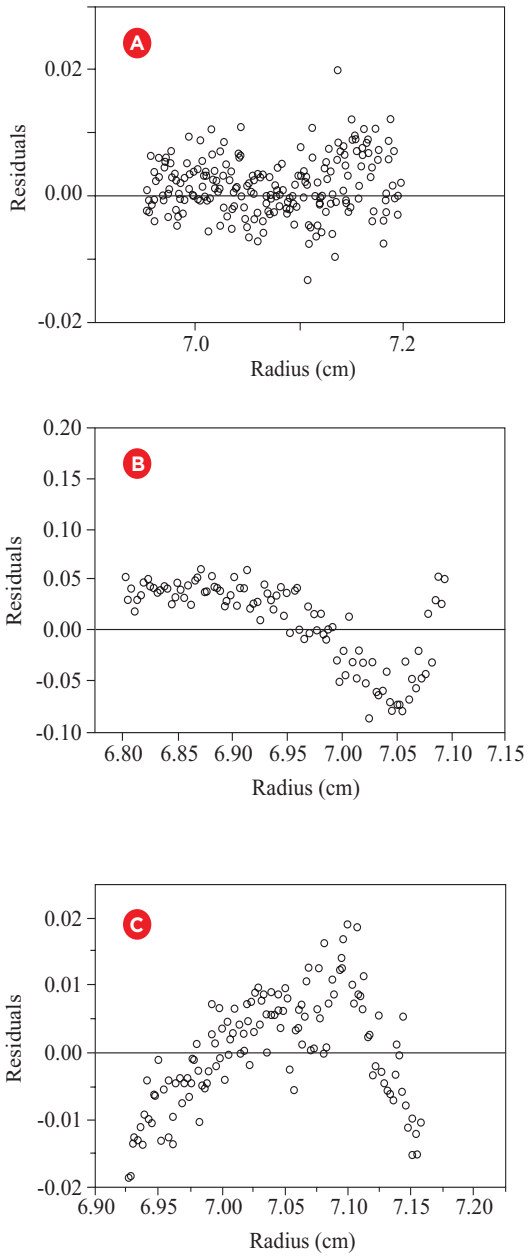

Figure 10. Residuals from desired fit. a) Good fit; b) aggregation; c) nonideality.

CHI SQUARE

Several mathematical analyses are used for determining goodness of fit (Straume and Johnson, 1992a). Of these, the χ^2 test is probably the most common. Once it is clear that there are no systematic trends in the residual plot, the χ^2 test provides a quantitative measure for the goodness of fit of a particular model. The χ^2 statistic is defined as:

Equation 13

$$\chi^2 = \sum_i \frac{(\text{observed residual} - \text{expected residual})_i^2}{\text{expected residual}_i}$$

over the defined confidence interval or the sum over all data points of the residuals squared, normalized to the error estimate for each point. The standard error is used in weighting the individual data points. By dividing this weighted χ^2 value by the number of degrees of freedom, the reduced χ^2 value, or variance, is obtained. The number of degrees of freedom is defined as $n - \acute{n} - 1$, where n is the number of data points and $\acute{n}$ is the number of parameters being determined in the analysis. The value of the reduced χ^2 should approach one if the mathematical model accurately describes the data.

PARAMETER CORRELATION

Correlation of unknown parameters is another important diagnostic. Statistically, the dependence of one parameter on another can be calculated in a correlation coefficient. Calculation of these coefficients with multiple fitting parameters involves use of covariance matrices to obtain the final correlation matrix with correlations between each pair of parameters (Bard, 1974; Bates and Watts, 1988). This calculation can be complicated and requires use of a computer. Absolute correlation between variables results in a correlation coefficient of ± 1.00. No correlation results in a value of 0.00. In nonlinear regression techniques, correlation coefficients can be determined between all parameters. Ideally, these coefficients should be low enough to show little or no correlation, and the user must decide according to the model being fitted how much correlation is acceptable. As a general rule, no coefficient having an absolute value greater than 0.95 would be acceptable. The accuracy of values for highly correlated parameters is greatly reduced.

The parameters most likely to show the highest correlation in a selfassociating system are the association constants. Constraining the value of one or more of the correlated parameters, while ensuring goodness of fit to the proposed model, can help reduce the coefficient values.

SUMMARY

Data analysis is the most important aspect of characterizing a self-associating system using the analytical ultracentrifuge. The first level of analysis is a qualitative one using diagnostic graphs. At this level questions are addressed about homogeneity, nonideality, and whether or not an association reaction is occurring. Also, the reversibility of an association and an estimate of monomer molecular weight can be determined. The next level of analysis involves nonlinear regression analysis for a quantitative determination of monomer molecular weight, association constants, stoichiometries, and nonideality coefficients. At this level, the most accurate information can be obtained using fitting procedures with multiple data sets varying both speed and starting concentration. It is necessary to test a number of possible model equations describing the associating system to find the model that best describes the equilibrium. In many cases, varying constrained parameters will accomplish this task. Finally, goodness-of-fit graphics and statistics help to distinguish the model that best describes the system. If a significant statistical difference between models cannot be established, the simplest model should be used to describe the system until more data are obtained.

APPENDIX A: A RATIONAL APPROACH TO MODELING SELF-ASSOCIATING SYSTEMS IN THE ANALYTICAL ULTRACENTRIFUGE

This section is intended to provide a rational approach to modeling sedimentation equilibrium data for determining stoichiometry, association constants, and in certain situations, the degree of nonideality of reversible self-associating systems.

1. WHAT QUESTIONS NEED TO BE ANSWERED?

Before trying to analyze any data, and before running any equilibrium experiments, one should have a clear idea of the questions to be asked about a particular system. This helps to design the experiment with respect to the question. If, for example, very little is known about the system, the experiments should be exploratory in nature and answer more qualitative questions. The first step might be to estimate sample purity, or the extent of associative behavior, *i.e.*, is the interaction weak, moderate, or strong. As more is known about the system, follow-up experiments can be tailored to answer more specific quantitative questions.

Having a sense of which questions are pertinent can also determine how much time needs to be spent on a system. It may not always be necessary to understand to the last decimal point everything about a system.

2. DON'T VARY EVERYTHING AT ONCE

The self-association fitting equation, provided as part of the XL-A Data Analysis Package, can be used to model up to four interacting species. The values of interest usually include one or more of the following: the molecular weight of the monomer, the stoichiometry of the system, and the association constants. These properties are expressed as parameters in a model equation. These parameters are determined by solving the appropriate equation, identified by the best-fit curve through the data, using a nonlinear curve fitting algorithm. A series of iterative guesses are made for each parameter to minimize the least-squares response between calculated and experimental data sets. Since these

parameters are often unknown, the first impulse in solving for each of them is to simultaneously vary all the parameters in the model and let the fitting routine sort out the numbers.

Although tempting, this approach can lead to problems on two fronts: 1) the program may have difficulty converging since the error space may appear flat, or 2) if a fit is attained, the accuracy of the results may be suspect. With respect to accuracy, the algorithm may have converged on a local rather than a global least-squares minimum, or the large number of varied parameters may have resulted in fitted values that are highly correlated with each other and, therefore, statistically questionable.

3. DETERMINE *M* AND *n*

As a rule, the number of parameters allowed to vary during a fit should be kept to a minimum. This often requires advance knowledge of the values of some parameters. One of the most important parameters to determine is the monomer molecular weight, *M*. This value can either be calculated from the formula molecular weight or determined by complementary techniques, or it can be determined through the use of the analytical ultracentrifuge. The analytical ultracentrifuge can be used to determine *M* by running the sample under denaturing conditions and fitting directly to the molecular weight parameter. [A lower-than-expected estimate of *M* can mean that a baseline offset, *E*, needs to be included in the fit (see next section).] Alternatively, the equilibrium gradient can be transformed to a M_w *vs.* concentration plot and the curve extrapolated to the ordinate in order to obtain an estimate of *M* (see Figure 11). Reading the curve in the other direction (to the highest gradient concentration) and dividing the molecular weight at this point by *M* affords a measure of the highest associative order, *n*, of the system, assuming a high enough loading concentration has been used. (Samples are normally run at multiple concentrations and speeds.) It should be cautioned that this technique provides only a rough estimate of *n*, and that the value obtained is not above suspicion. An estimate of *n* that is lower than expected can result from the molecular weight at the highest gradient concentration being depressed by nonideality effects or high-molecular weight aggregates that have pelleted. Or, an estimate that is higher than expected can result from *M* being depressed by a baseline offset, *E* (see next section). An important consequence of estimating *n* is that it provides the information necessary to narrow the field down to a couple of potential associative models. For example, an estimate of $n \cong 3.6$ may suggest a monomer-dimer-tetramer or higher order of association that isn't fully assembled. This information also allows one to dismiss irrelevant models; for example, fits to monomer-dimers or monomer-trimers would be inappropriate at this point.

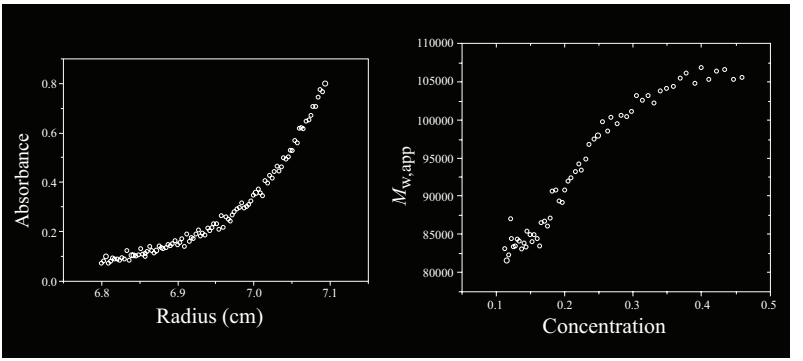

Figure 11. A simulated monomer-dimer association. a) Absorbance scan; b) plot of M_w vs. concentration.

The stoichiometry of a system can also be determined using a technique known as a species analysis. This technique involves first expressing each species in the model in terms of absorbances. The algorithm then converges directly on each of the absorbance terms for a series of preselected models, *e.g.*, monomer-dimer, monomer-tetramer, etc., or one extended model, *e.g.*, monomer-dimer-tetramer-octamer. In this manner, a variety of potential models can be quickly evaluated and dismissed in terms of physical reality. For example, a suspected association state that converges to a large negative absorbance can probably be ruled out. (A large negative absorbance usually means a value several orders of magnitude larger than the baseline offset; see next section.) Following this initial prescreening, a more refined model containing association constants can be used for estimating more substantive values.

4. BASELINE CONSIDERATIONS

The baseline offset, *E*, is included in a model when a correction term is needed to account for the presence of any absorbing particulates left undistributed in the gradient. Left uncorrected, E can lead to a low estimate for $M_{w,app}$ when fitting to a single ideal species model, or a high estimate for n when reading an M_w *vs.* concentration plot; in the latter case, *E* has a stronger negative effect on *M* at the lower end of the gradient. Since even small values of *E*, such as 10^{-2}, can play an adverse role if left uncorrected in a fit, it is important to test for its presence. *E* can be measured by overspeeding a run and reading the absorbance of the trailing gradient (the meniscus-depletion method).

5. SINGLE DATA FILES, THEN MULTIPLE DATA FILES

When dealing with a variety of files collected at multiple concentrations and speeds, it is often useful to begin by analyzing single data files. This affords the opportunity to inspect each data file individually, and either to accept, edit, or reject individual files before they can adversely impact a multiple fit with other data files. After examining individual files, it is helpful to begin the analysis of multiple data files by first grouping the data sets by speeds and channels (if multichannel centerpieces were used), before attempting to incorporate all of the data files in a fit. Files should be rejected only if there are obvious problems. By grouping files in terms of speeds and channels, future concerns over certain fit parameters can either be avoided or rationalized by revealing them at their source, *e.g.*, a baseline offset inconsistent with trends observed in other files, or a window scratch at a certain location.

Grouping files of different concentration has the advantage of incorporating individual files that collectively span the associative range of the system, which can facilitate the identification of the correct model. Grouping files in terms of concentration and/or speed can also be used as a diagnostic to evaluate the homogeneity and reversibility of a system (see section 11).

Certain parameters contained in a model are treated differently depending on whether single or multiple data files are employed. When fitting single data files, it is important to constrain *E* to a known value as measured by the meniscus depletion method. Problems can arise if *E* is allowed to vary without knowing its value. The reason for this is that *E* is highly correlated with other parameters, and the accuracy of each of these values can be compromised for the sake of a fit. If the value for *E* is unknown, it is better to leave it at 0. When fitting multiple data files, the opposite is the case; the baseline term should be allowed to vary. Since multiple runs are usually made at multiple speeds and concentrations, it is reasonable to assume that the baseline term will be different for different concentrations. However, the same sample should have the same value of E regardless of speed. By allowing the baseline offset to vary, its value is adjusted with respect to the conditions of each file, and the accuracy of the fit is enhanced. When allowed to vary, *E* is usually given an initial guess of 0; it converges reasonably close to each of its measured values.

There also exists some commonality between the two fitting routines. The absorbance at the reference radius, A_0, is treated the same way for fits to either single or multiple data files; it should be allowed to vary.

Analysis of ideal models for $M_{w,app}$ can be done using either single or multiple data files. Analysis of more complicated models that involve additional varying parameters, such as E and K_a, should be evaluated using multiple data files. Multiple data files have the advantage of introducing more data points to a fit. Also, a global least-squares minimum often can be easier to reach.

6. START WITH THE SIMPLE MODEL FIRST

It is always a good idea to start with the simplest case first: the single ideal species model. In addition to providing an apparent weight-average molecular weight, $M_{w,app}$, the residuals of the fit can substantiate or refute any preconceived notions about the behavior of the system (see Figure 10).

7. MODEL IN STEPS

One approach to modeling an associating system using multiple data files is to converge on the parameters in a series of iterative steps. This approach can help overcome problems the algorithm may have in converging on too many parameters. Ideally the monomer molecular weight and associative order of the system are known and constrained to their appropriate values. If these values are unknown, determining the correct model can be very difficult.

As mentioned earlier, files can be grouped in terms of speeds and/or concentration. This grouping becomes useful for diagnosing the associating system as well (see section 11).

In the first step, the absorbance at the reference radius and the appropriate association constant are allowed to vary. To ensure that the algorithm is moving down the error space, parameter guesses are given realistic values. Values for parameters obtained from the first fit are used as guesses for the next fit when an additional parameter is allowed to vary, *e.g.*, the baseline offset. For situations where the baseline terms are known, the fitted parameters can be validated. If there is close agreement, this can lend additional confidence to the other fitted parameters. Alternatively, the first convergence step may include the baseline term, with the second step incorporating the association constant. There is no single approved method for analyzing all associating systems. Each system has its own peculiarities.

8. CHECK FOR PHYSICAL REALITY

Although a fit can be evaluated using a variety of sophisticated statistics, one of the easiest and most often overlooked methods is simply to check the fitted parameters in terms of physical reality. Is the molecular weight or association constant consistent with expectations; is A_0 consistent with the observed gradient; and is the baseline term close to zero or the experimentally measured value?

9. PROBLEMS IN FITTING

Sometimes getting a fit to converge can be particularly difficult. Assuming the number of parameters allowed to vary has been kept to a minimum, the problem may lie in the initial guesses. If the initial guesses are too unrealistic, the algorithm can have trouble getting off the flat part of the error space. If the association constant is suspected to be at fault, there are ways of improving a guess. By knowing the weight-average molecular weight (estimated from a single ideal species fit) and dividing by the monomer molecular weight, the extent of association may be more closely approximated. If, for example, this ratio comes out low with respect to the known stoichiometry, this can indicate a weak association, and a lower estimate of the association constant may prove a better initial guess. (An initial guess of 10^{-1} or 10^{-2} for K_a is usually used as a default.)

In general, problems in fitting may occur with gradients that are either too shallow or too steep, resulting from extremes in molecular weight and/or run speed. This is because the algorithm fits to an exponential equation and assumes that the gradient follows an exponential profile over the entire solution column height, which may not always be the case.

There may be problems in fitting to a model if the monomer is present in minute amounts. Since the models are written in terms of monomer concentration, a measurable amount of monomer must be present to avoid an ambiguous fit. The fit may be improved by expressing the model in terms of the predominant species present in the system. In cases where the monomer has assembled irreversibly to a dimer, for instance, the stoichiometry of the system may be misidentified as monomer-dimer. It may actually be a dimer-tetramer system in which the dimer is the lowest molecular weight species present. A run made under dissociating conditions using guanidine hydrochloride, for example, may prevent this type of error.

Another condition that can be problematic occurs when the time allowed for a fit is insufficient. The number of loops (iterations) the algorithm goes through before quitting is 100. Sometimes by refitting the same guesses for a second set of iterations or increasing the number of iteration loops, a fit will converge.

Even with the best of intentions and an ideal set of conditions, the association may be too complicated to analyze. These models are ideally suited for discrete associations, involving up to three species (although we allow the capability of modeling up to four species). Trying to model to an indefinite system of more than five species, containing intermediate irreversible associations, can be too much for this approach.

If this turns out to be the case, at least the equilibria have been revealed to be highly complex, which is probably more than was known about the system before the analysis.

10. MODEL TO SYNTHETIC DATA

Using software programs that simulate a variety of ideal associative systems can be an invaluable aid to modeling. As well as providing examples of fits to simple ideal systems, a simulation program can also serve as a platform for confirming actual fits encountered in more complicated multicomponent systems. An equilibrium simulation program is included as part of the XL-A Data Analysis Package; simeq.exe is located on the XL-A Program Disk.

11. THE MODEL AS A DIAGNOSTIC TOOL

The advantage of using a model as a diagnostic tool is the variety of information that can be revealed about the homogeneity and ideality of the system. For example, a single ideal species should yield the same $M_{w,app}$ regardless of initial concentration or speed (see Figures 4 and 6). However, as more complicated systems are studied, diagnosing their behavior may not always be as simple. The following demonstrates some diagnostic manipulations that can be used in evaluating sample behavior under optimum conditions. This material is also presented as a flow chart in Section 13. Included as an addendum are some examples of how diagnosing a system under more realistic conditions can be more challenging.

Optimum Diagnostic Conditions:

1. Evaluating an ideal associating system.

An ideal associating system should yield either a constant or increasing $M_{w,app}$ with increasing initial concentrations, as read from a M_w *vs.* concentration plot of a mass action association (see Figure 4). With increasing speeds, the apparent molecular weight and the association constant should remain constant. (Note: an increase in $M_{w,app}$ with increasing concentration indicates the system is still assembling.) A K_a that is independent of either speed or concentration indicates a reversible associating system. Most protein-protein associations studied at low to moderate concentrations (<1 mg/mL) behave nearly ideally.

2. Evaluating a heterogeneous noninteracting system.

Heterogeneous, noninteracting behavior in a single ideal or associating system can result from either a competing irreversible equilibrium or contamination by aggregates (material present as a percentage of a mixture). Aggregates are heterogeneous with respect to molecular weight and can often be removed by size exclusion chromatography. Irreversible equilibria are heterogeneous with respect to K_a and may not fractionate as easily. For either condition, the apparent

molecular weight and association constant decrease with both increasing speeds and concentrations (see Figures 4 and 6). The reason for this is that the higher order components or aggregates are pelleted and no longer contribute to a fit.

For an associating system containing higher order aggregates, the stoichiometry, *n*, should converge to a higher value (at a given speed or concentration) when allowed to vary.

3. Evaluating for nonideality.

As a general rule, there should be a good reason for including the nonideality term, *B*, in a fit. Casually including *B* in a model to see if a fit is improved is not the way to approach this parameter. Since legitimate parameter values are typically small (10^{-3}-10^{-4}), indicating a fair amount of nonideality, forcing *B* into a fit can corrupt its thermodynamic significance. Conditions that can warrant the inclusion of *B* are: a weight-average molecular weight dropping with increasing concentration, or a sample suspected to be highly charged or asymmetric. It should be noted that systems will begin to exhibit nonideality effects when pushed to higher concentrations. For nonideal systems behaving as a single species, *B* is typically given an initial guess of 10^{-4} and converges to a positive number during a fit.

Including *B* in an associating system can be considerably more difficult. The easiest case is if K_a and *B* are both large, *e.g.*, $K_a = 10^5$ M^{-1} and $B = 0.01$ mL/g (with an approximate extinction coefficient of 1 at A_{280}). It also helps if the equilibrium is a finite one and the final assembly is complete at relatively low concentrations. This allows any nonideality to be observed in near isolation at high concentrations. For systems of this type, both *B* and K_a may be solved simultaneously.

For systems where the nonideality term is very weak, *B* may be obscured by the association. For systems of this type, it may be necessary to determine each parameter separately. One method that has proved successful (Laue *et al.*, 1984) is to neglect nonideality when fitting for K_a, then to go back with a fixed estimate of *B* and compare the fits. For situations of this type, values for *B* are based on the size, shape, and estimated charge of the molecules using equations given in Tanford (1961).

Realistic Diagnostic Conditions

A variety of competing situations may exist that can make evaluation of sample behavior very difficult. Below are two scenarios.

1. Association and nonideality in the same system.

For a nonideal associating system, the sample may exhibit behavior consistent with an ideal noninteracting model. The reason is that the increase in apparent molecular weight with increasing concentration for an associating system is offset by a decrease in the molecular weight for a nonideal system. The net effect may be a constant apparent molecular weight with increasing concentration. If this situation is suspected, one solution may be to minimize the effect of nonideality. If the nonideality is suspected to occur from crowding or excluded volume effects, the sample may be run under more dilute conditions. If the nonideality is suspected to result from charge effects, the sample may be run in a higher ionic strength buffer.

2. Association and heterogeneity in the same system.

If an ideal associating system contains a competing irreversible equilibrium, a condition similar to the above may occur. Since increases in speed or concentration have opposite effects on the direction of change in the apparent molecular weight, it may be difficult to identify any of the associative states.

12. TEST THE FIT AND THE MODEL

Before accepting a fit, it is a good idea to test it. One technique involves making guesses on either side of a fitted parameter to determine whether the algorithm converges back to the same value. Another technique involves allowing all the fitted parameters in question to vary in a final iterative step to see if they all return to their respective values. (Note: this last technique may not work for parameters that are too highly correlated, *i.e.*, with correlation coefficients > 0.95.) If, after this step, the fit is still in question (based on fit values, statistics, or residuals), an additional term may need to be considered, such as the nonideality coefficient, or the presence of higher order aggregates that may be throwing off the fit. Or perhaps the wrong model was selected. Nagging doubts about a fit or a model can often be dispelled by repeating a set of experiments with a fresh preparation.

Comparisons with other nonlinear fitting routines may yield slightly different values. This can be due to differences in how algorithms converge on a least-squares minimum or to rounding errors between algorithms.

13. FLOW CHART

The flow chart on page 39 is provided as an aid to adapting a rational approach to modeling. The order of analysis appears on the left and follows a vertical stepwise approach. The center columns show the

affected parameters, while the column to the far right identifies the associative state being tested.

14. LIMITATIONS

Modeling is not always straightforward, and there may appear, at times, a certain level of ambiguity in a fit. Competing models can appear indistinguishable, and certain fit criteria can run contrary to common sense. In sedimentation analysis, there are a variety of limitations or sources for error that can lead to this uncertainty. While some conditions leading to ambiguous fits can be prevented through careful experimental design, others cannot, *e.g.*, problems inherent to an iterative least-squares process.

The most obvious condition affecting a fit concerns sample preparation. There is no substitute for having a clean sample. Heterogeneous aggregates present in a sample can wreak havoc during modeling. Purification through size exclusion chromatography is suggested as the method of choice (Laue, 1992). Dialysis can be effective if the contaminants are very small.

Other conditions that can affect a fit range from physical aberrations, such as dirty windows, to experimental conditions (*e.g.*, inappropriate speed selection), to data collection (*e.g.*, not enough data points, averages, or data sets), to algorithm limitations (*e.g.*, convergence on a local rather than a global least-squares minimum), to inequities in modeling (*e.g.*, improper model selection).

It is important to emphasize that there are systems too complicated for this approach, *or any approach*. Systems containing multiple species, for example, can provide at best only a rough order of approximation in their interpretation. This should not be construed as a limitation of sedimentation analysis or modeling, but rather a reflection of the level of complexity present in nature.

Order of Analysis
(single and multiple data files)

1. Fit to ideal model
   - vary: $\ln(A_0)$, $M_{w,app}$
     - i) constrain baseline (single files)
     - ii) vary baseline (multiple files)

2. Fit to nonideal model
   - vary: $\ln(A_0)$, $M_{w,app}$ (for nonassociative); then B (not all at once)
     - i) fit to nonassociative model; constrain baseline (single files); vary baseline (multiple files)
     - ii) fit to associative model (fix M); constrain baseline (single files); vary baseline (multiple files)

3. Fit to homogeneous associative model
   - vary: $\ln(A_0)$, $M_{w,app}$
     - i) constrain M, n
     - ii) constrain baseline (single files); vary baseline (multiple files)
     - iii) vary n (if high, vary n + 1)

Affected Parameters						Associative State
INCREASED SPEED		INCREASED CONCENTRATION		B	n	
$M_{w,app}$	K_a	$M_{w,app}$	K_a			
0		0				▶ Homogeneous ideal species
–		–				▶ Heterogeneity
		-		+		▶ Nonideal nonassociative or associative
0	0	+ or 0	0		>	▶ Homogeneous associative monomer–n-mer (reversible)
–	–	–	–			▶ Heterogeneous noninteracting aggregation

Key to symbols:

$M_{w,app}$	an increase (+), decrease (-), or little or no change (0); estimated by fitting to an ideal model
K_a	an increase (+), decrease (-), or little or no change (0)
B	positive value (+)
n	a greater-than-expected value (>)

APPENDIX B: PARTIAL SPECIFIC VOLUME

The partial specific volume ($\bar{v}$) is defined as the change in volume (in mL) of the solution per gram of added solute. Typically, for proteins, an error in this parameter is magnified threefold in calculating the molecular weight, so it is important to obtain accurate values for $\bar{v}$. An extensive discussion and tabulation of values for partial specific volumes can be found in a review by Durschlag (1986). Two methods are generally used for the experimental measurement of $\bar{v}$: densitometry or analytical ultracentrifugation in solvents with differing isotope composition. Densitometry is the most accurate, but requires expensive instrumentation and large amounts of solute. Use of solvents containing D_2O vs. H_2O in parallel experiments in the ultracentrifuge allows one to solve for $\bar{v}$ in the equations:

Equation 14A

$$M_{H_2O} = M(1 - \bar{v}\rho_1)$$

Equation 14B

$$M_{D_2O} = M(1 - \bar{v}\rho_2)$$

where M_{H_2O} and M_{D_2O} are the buoyant molecular weights in H_2O and D_2O, respectively, and ρ_1 and ρ_2 are the densities of the two solvents.

An alternative method is to estimate $\bar{v}$ based on the sample's composition. Comparisons with experimentally derived values indicate that estimates based on protein amino acid compositions are typically good to within 1-2%. Estimates for conjugated proteins, with carbohydrate moieties for example, can lead to greater errors. Hydration of molecules is another source of error that must be considered in the calculation.

Partial specific volume is usually estimated from composition using the method of Cohn and Edsall (1943):

Equation 15

$$\bar{v}_c = \frac{\sum_i W_i \bar{v}_i}{\sum_i W_i} = \frac{\sum_i N_i M_i \bar{v}_i}{\sum_i N_i M_i}$$

where $\bar{v}_c$ is the calculated partial specific volume, W_i is the weight percent of the *i*th component, N_i is the number of residues, M_i is the molecular weight of the corresponding component (for amino

acids, the residue molecular weight minus 18) and $\bar{v}_i$ is the partial specific volume of the component. Tables 1-4 show some representative values as reported by Durschlag (1986). This calculation is for 25°C and can be adjusted for a temperature range of 4-45°C using the equation:

Equation 16

$$\bar{v}_T = \bar{v}_{25} + [4.25 \times 10^{-4}(T - 298.15)]$$

where $\bar{v}_T$ is the partial specific volume at temperature T (in Kelvin) and $\bar{v}_{25}$ is the partial specific volume calculated from equation 15 at 25°C (Laue, 1992).

Table 1. Amino Acids[1] at 25°C

AMINO ACID	M_r	$\bar{v}$ (mL/g)	HYDRATION	
			mol-H_2O mol-aa	mol-H_2O mol-aa
			(pH 6-8)[2]	(pH 4)[3]
Ala	71.1	0.74	1.5	1.5
Arg	156.2	0.70	3	3
Asn	114.1	0.62	2	2
Asp	115.1	0.60	6	2
Asx[4]	114.6	0.61	4	2
Cys	103.2	0.63[5]	1	1
2Cys	204.3	0.63	-	-
Gln	128.1	0.67	2	2
Glu	129.1	0.66	7.5	2
Glx[4]	128.6	0.665	4.8	2
Gly	57.1	0.64	1	1
His	137.2	0.67	4	4
Ile	113.2	0.90	1	1
Leu	113.2	0.90	1	1
Lys	128.2	0.82	4.5	4.5
Met	131.2	0.75	1	1
Phe	147.2	0.77	0	0
Pro	97.1	0.76	3	3
Ser	87.1	0.63	2	2
Thr	101.1	0.70	2	2
Trp	186.2	0.74	2	2
Tyr	163.2	0.71	3	3
Unk[4]	119.0	0.72	2.4	2
Val	99.1	0.86	1	1

[1] Reprinted from Laue *et al.* (1992) with permission of Royal Society of Chemistry; based on values from Cohn and Edsall as presented in Durschlag (1986).
[2] Used for calculation of ϕ in 6 M guanidine-HCl.
[3] Used for calculation of ϕ in 8 M urea.
[4] Values for Asx and Glx are averages of acid and amide forms. Values for Unk are average of all amino acids.
[5] Based on value from Zmyatnin as presented in Durschlag (1986).

Table 2. Carbohydrates at 20°C[1]

CARBOHYDRATE	M_r	$\bar{v}$ (mL/g)
Fructose	180	0.614
Fucose	164	0.671
Galactose	180	0.622
AGlucose (calculated)	180	0.622
(0.5 M)	180	0.623
(3 M)	180	0.638
Hexose	180	0.613
Hexosamine	179	0.666
Sucrose (0.05 M)	342	0.613
(0.1-0.2 M)	342	0.616
(1 M)	342	0.620
Lactose (0.1 M)	342	0.606
(0.4 M)	342	0.610
Mannose	180	0.607
Methyl-pentose	165	0.678
N-Acetyl-galactosamine	221	0.684
N-Acetyl-glucosamine	221	0.684
N-Acetyl-hexosamine	221	0.666
N-Acetyl-neuraminic acid	308	0.584
Sialic acid	308	0.584

[1] Excerpted from Laue *et al.* (1992); values of $\bar{v}$ excerpted from Durschlag (1986).

Table 3. Denaturants at 20°C[1]

DENATURANT	M_r	$\bar{v}$ (mL/g)
Guanidine HCl (lim c→0)	96	0.70
(1 M)	96	0.732
(2 M)	96	0.747
(6 M)	96	0.763
Urea (lim c→0)	60	0.735
(1 M)	60	0.745
(8 M)	60	0.763
DOC sodium deoxycholate		
(below and above cmc2)		0.779
SDS sodium dodecyl sulfate		
(below cmc)	60	0.814
(above cmc)[3]	60	~0.86
(above cmc in H_2O)[3]	60	0.854
(above cmc 0.1 M NaCl)[3]	60	0.863
Triton X-100 (above cmc)		0.913
Tween-80 (above cmc)[2]		0.896

[1] Excerpted from Laue *et al.* (1992); values of $\bar{v}$ excerpted from Durschlag (1986).
[2] cmc = critical micelle concentration.
[3] Determined at 25°C.

Table 4. Miscellaneous at 20°C[1]

SUBSTANCE	$\bar{v}$ (mL/g)
Acetyl-CoA	0.638
ATP	0.44
CTP	0.44
Ethanol	1.18
Glycerol (10%)	0.767
(20%)	0.768
(30%)	0.770
(40%)	0.772

[1] Excerpted from Laue *et al.* (1992); values of $\bar{v}$ excerpted from Durschlag (1986).

The effects of pH on partial specific volume are usually minor and are generally ignored in most calculations. However, in instances where changes in structure such as unfolding may occur, large errors in $\bar{v}$ may be observed. These structural changes can be monitored by methods such as circular dichroism to ensure the correct calculation for $\bar{v}$.

Experiments are often run in the presence of denaturants such as urea or guanidine-HCl. These compounds affect partial specific volume through preferential binding of the denaturant to the protein. Calculations, in this case, must take into account the binding of denaturant as well as the effect it has on the hydration of the molecule (Durschlag, 1986; Lee and Timasheff, 1974, 1979; Prakash and Timasheff, 1981). In these cases, $\bar{v}$ must be replaced by Φ, the apparent isopotential partial specific volume. Generally, increasing ionic strength results in a linear increase in Φ. The hydration effects of urea and guanidine-HCl result in a nonlinear relationship, however. In these cases, Φ replaces $\bar{v}$ and can be determined by:

Equation 17

$$\varphi = \bar{v} - \left(\frac{1}{\rho} - \bar{v}_3\right)\left(A_3 - g_3 A_1\right)$$

where $\bar{v}$ is from equation 15, ρ is the solvent density, $\bar{v}_3$ is the partial specific volume of the denaturant, A_3 is the number of grams of denaturant bound to the protein, g_3 is the number of grams of denaturant per gram of water and A_1 is the hydration in grams of water per gram of protein. (Note: this equation, as printed in Laue *et al.*, 1992, contains an error.)

Table 5 shows the values for 6 M guanidine-HCl and 8 M urea at 20°C (Lee and Timasheff, 1974, 1979; Prakash and Timasheff, 1981).

Table 5. ϕ for Proteins in 6 M Guanidine-HCl[1] and 8 M Urea[2]

PROTEIN	$\bar{v}$ (native)	ϕ (6 M guanidine-HCl)	ϕ (8 M urea)
Lima bean			
trypsin inhibitor	0.732	0.698	0.691
Ribonuclease A	0.696	0.694	0.695
α-Lactalbumin	0.704	0.698	0.699
Lysozyme	0.702	0.694	0.700
β-Lactoglobulin	0.751	0.719	0.719
Chymotrypsinogen A	0.733	0.712	0.720
α-Chymotrypsin	0.738	0.713	0.714

[1] Values excerpted from Lee and Timasheff (1974).
[2] Values excerpted from Prakash *et al.* (1981).

A_3 is calculated for both urea and guanidine-HCl assuming one molecule of denaturant bound to every pair of peptide bonds plus one for every aromatic side chain (including histidine) using the following equations:

Equation 18

$$A = \frac{\text{moles of denaturant}}{\text{moles of protein}} \times \frac{M_{\text{w,denaturant}}}{M_{\text{w,protein}}}$$

Equation 19

$$\frac{\text{moles of denaturant}}{\text{moles of protein}} = \frac{N_t - 1}{2} + N_{\text{aromatic}}$$

where N_t is the total number of amino acids, and N_{aromatic} is the number of aromatic amino acids. For 6 M guanidine-HCl at 20°C, ρ = 1.1418 g/mL, $\bar{v}_3$ = 0.763 mL/g and g_3 = 1.007 g guanidine-HCl per g H_2O. For 8 M urea at 20°C, ρ = 1.1152 g/mL, $\bar{v}_3$ = 0.763 mL/g and g_3 = 0.752 g urea per g H_2O.

The assumptions in these calculations have been tested in the determination of $\bar{v}$ from the amino acid composition of a protein. The methods are assumed to hold true for most non-amino-acid constituents as well, but have not been tested to as great an extent. If there is any question about the validity of a calculation, one might consider confirming $\bar{v}$ with an experimental method since an accurate value is critical for further analysis.

APPENDIX C: SOLVENT DENSITY

Density (ρ) is simply the mass (in g) of one mL of solvent and is dependent on temperature and composition. As with $\bar{v}$, ρ can be a source of error in molecular weight calculations, so that a value accurate to 4-5 decimal places is usually desirable. Solvent density can be measured with a pycnometer, but extensive published data permit accurate calculations that account for both temperature and buffer composition (Johnson and Frazier, 1985; Wolf *et al.*, 1986).

The solvent density can be corrected to the experimental temperature using the equation of Kell for pure water (ρ_T) as modified by Laue *et al.* (1992) to yield cgs units:

Equation 20

$$\rho_T = 1.000028 \times 10^{-3}\left[\frac{999.83952 + 16.945176T}{1 + (16.879850 \times 10^{-3}\,T)}\right]$$
$$-1.000028 \times 10^{-3}\left[\frac{(7.9870401 \times 10^{-3}\,T^2) + (46.170461 \times 10^{-6}\,T^3)}{1 + (16.879850 \times 10^{-3}\,T)}\right]$$
$$+1.000028 \times 10^{-3}\left[\frac{(105.56302 \times 10^{-9}\,T^4) - (280.54253 \times 10^{-12}\,T^5)}{1 + (16.879850 \times 10^{-3}\,T)}\right]$$

where T is the temperature in °C and the factor, 1.000028×10^{-3}, converts the units from kg/m^3 to g/mL. Densities calculated in this manner are good to at least five decimal places from 0-100°C. Temperature corrections for isotopes of water are different (Laue, 1992; Steckel and Szapiro, 1963) and will not be discussed here.

The density of a buffer or any other solution can be estimated by summing the density increments calculated for each component (Svedberg and Pedersen, 1940). Density increments for each component can be calculated using the polynomial function:

Equation 21

$$\Delta\rho_{c_i} = (A_i + B_i c_i^{1/2} + C_i c_i + D_i c_i^2 + E_i c_i^3 + F_i c_i^4) - 0.998234$$

where Δρ is the density increment at molar concentration c_i. The coefficients A_i to F_i are fitted parameters using tabulated values of ρ as a function of c_i, and are determined using least-squares procedures (Johnson and Frazier, 1985). Values of A-F excerpted from Laue (1992) for some common buffer components are shown in Table 6.

Table 6. Coefficients for the Power Series Approximation of the Density[1]
$\rho = A + (B \times 10^{-3} c^{1/2}) + (C \times 10^{-2} c) + (D \times 10^{-3} c^2) + (E \times 10^{-3} c^3) + (F \times 10^{-6} c^4)$

SOLUTE	Form	A	B	C	D	E	F	[c]max (M)	10^4 × Max. Error g/mL
Acetate	acid	0.9990	-4.04	1.023	-0.9156	0.67	-2.403	16.0	18
	trichloro	0.98774	-	7.6173	2.8618	-19.437	239.421	4.0	6
Ammonium	chloride	0.99787	1.3717	1.5743	-3.8795	-	-	4.7	0.7
	sulfate	0.99828	-	7.835	-59.751	4.0581	-	3.8	1
Calcium	chloride	0.698	0.698	0.698	0.698	0.698	0.698	0.698	0.698
Cesium	chloride	0.694	0.694	0.694	0.694	0.694	0.694	0.694	0.694
Magnesium	chloride	0.99801	-	7.7423	-2.8939	2.1536	-	4.0	4
	sulfate	0.99763	3.9236	11.547	-4.8526	3.0613	-	2.9	0.6
Phosphoric	acid	0.99834	-	5.3345	-3.0937	-	-	0.6	0.6
Potassium	bromide	0.99793	1.6329	8.303	-0.5438	0.1187	-	4.6	0.6
	chloride	0.99791	1.5586	4.5573	-0.7041	-0.2539	10.4034	3.9	0.6
	hydroxide	0.99842	-	4.7917	-1.1602	0.2968	-	12.0	6
	phosphate monobasic	0.99815	-	9.7583	-7.4352	-	-	0.8	0.8
	phosphate dibasic	0.99837	-	14.5429	-7.3137	-	-	0.5	0.8

Sodium	acetate	0.99623	12.4776	2.2804	9.2092	-27.6174	267.596	4.2	0.7
	chloride	0.99828	-	4.1261	-1.0580	0.6097	-	6.2	1
	citrate tribasic	0.99497	1.5782	-	-27.741	-	-	1.8	7
	hydroxide	0.99624	7.0838	3.8735	-0.8969	0.1253	-	14.0	3.5
	phosphate monobasic	0.99826	-	8.838	-3.8893	5.4158	-41.9615	5.0	0.9
	phosphate dibasic	0.99794	3.4466	13.197	-7.0728	-	-	0.4	0.8
	phosphate tribasic	0.99843	-	18.6772	-7.5401	-	-	0.5	1
Other	EDTA	0.99824	-	17.8212	-16.271	-	-	0.18	0.6
	Sucrose	0.99822	-0.0346	13.2207	-1.8117	-	-24.855	4.0	1.2
	Tris buffer	0.99846	-1.3079	3.0305	0.2876	-0.6279	-	3.5	0.6
	Urea	0.99835	-0.4476	1.528	0.5425	-1.449	9.824	9.0	0.6
	Guanidine-HCl	0.99831	-	2.557	-2.7133	-	-	6.2	3.8

[1] Excerpted from Laue *et al.* (1992).

The values for A-F fit well to density data for simple salts. For organic salts and alcohols, however, some systematic deviations are apparent in the fit to determine A-F, and even though errors may be small, the polynomial model may not be adequate for this calculation.

Using density increments, the solution density at temperature *T* is:

Equation 22

$$\rho_{T,b} = \frac{\left(0.998234 + \sum \Delta\rho_{c_i}\right)\rho_T}{0.998234}$$

where ρT,b is the calculated density, $\rho_T/0.998234$ is the temperature correction factor, and $\Sigma\Delta\rho_{c_i}$ is the sum of density increments for buffer components.

All density correction calculations neglect the contribution of macromolecular solutes. Both high concentration and redistribution of the macrosolute can affect these values. Also, redistribution of the solvent, while usually negligible, should be examined. As with $\bar{v}$ calculations, ρ calculations are, for the most part, accurate. If, however, any doubt exists, one should use experimental methods for density measurement.

REFERENCES

Bard, Y. *Nonlinear Parameter Estimation*. San Diego, Academic Press,1974.

Bates, D. M., Watts, D. G. *Nonlinear Regression Analysis and Its Applications*. New York, J. Wiley, 1988.

Becerra, S. P., Kumar, A., Lewis, M. S., Widen, S. G., Abbotts, J., Karawya, E. M., Hughes, S. H., Shiloach, J., Wilson, S. H. Protein-protein interactions of HIV-1 reverse transcriptase: implication of central and C-terminal regions in subunit binding. Appendix: Ultracentrifugal analysis of mixed association, by M. S. Lewis. *Biochemistry 30*, 11707-11719 (1991)

Bevington, P. R., Robinson, D. K. *Data Reduction and Error Analysis for the Physical Sciences*. 2nd ed. New York, McGraw-Hill, Inc., 1992.

Chervenka, C. H. *A Manual of Methods for the Analytical Ultracentrifuge*. Spinco Division, Beckman Instruments, Palo Alto, 1969.

Cohn, E. J., Edsall, J. T. *Proteins, Amino Acids and Peptides as Ions and Dipolar Ions*. New York, Reinhold, 1943.

Correia, J. J., Yphantis, D. A. Equilibrium sedimentation in short solution columns. *Analytical Ultracentrifugation in Biochemistry and Polymer Science*, pp. 231-252. Edited by S. E. Harding, A. J. Rowe and J. C. Horton Cambridge, Royal Society of Chemistry, 1992.

Dierckx, P. An algorithm for smoothing, differentiation and integration of experimental data using spline functions. *J. Computat. Appl. Math. 1*, 165-184 (1975)

Durschschlag, H. Specific volumes of biological macromolecules and some other molecules of biological interest. *Thermodynamic Data for Biochemistry and Biotechnology*, pp. 45-128. Edited by H.-J. Hinz. Berlin, Springer-Verlag, 1986.

Formisano, S., Brewer, H. B., Jr., Osborne, J. C., Jr. Effect of pressure and ionic strength on the self-association of Apo-A-I from the human high density lipoprotein complex. Appendix: evaluation of volume changes in associating systems by sedimentation equilibrium. *J. Biol. Chem. 253*, 354-360 (1978)

Haschemeyer, R. H., Bowers, W. F. Exponential analysis of concentration or concentration difference data for discrete molecular weight distributions in sedimentation equilibrium. *Biochemistry 9*, 435-445 (1970)

Holladay, L. A., Sophianopoulos, A. J. Nonideal associating systems. I. Documentation of a new method for determining the parameters from sedimentation equilibrium data. *J. Biol. Chem. 247*, 427-439 (1972)

Holladay, L. A., Sophianopoulos, A. J. Statistical evaluation of ways to analyze nonideal systems by sedimentation equilibrium. *Anal. Biochem. 57*, 506-528 (1974)

Johnson, M. L. Analysis of ligand-binding data with experimental uncertainties in independent variables. *Methods in Enzymology*, Vol. 210, pp. 106-117. Edited by L. Brand and M. L. Johnson. San Diego, Academic Press, 1992.

Johnson, M. L., Faunt, L. M. Parameter estimation by least-squares methods. *Methods in Enzymology*, Vol. 210, pp. 1-37. Edited by L. Brand and M. L. Johnson. San Diego, Academic Press, 1992.

Johnson, M. L. Frazier, S. G. Nonlinear least-squares analysis. *Methods in Enzymology*, Vol. 117, pp. 301-342. Edited by C. H. W. Hirs and S. N. Timasheff. San Diego, Academic Press, 1985.

Johnson, M. L., Correia, J. J., Yphantis, D. A., Halvorson, H. R. Analysis of data from the analytical ultracentrifuge by nonlinear least-squares techniques. *Biophys. J. 36*, 575-588 (1981)

Lamm, O. Die Differentialgleichung der Ultrazentrifugierung. *Ark. Mat., Astron. Fys. 21B*:2, 1-4 (1929)

Laue, T. M. Short column sedimentation equilibrium analysis for rapid characterization of macromolecules in solution. *Technical Information DS-835*. Palo Alto, Calif., Spinco Business Unit, 1992.

Laue, T. M., Johnson, A. E., Esmon, C. T., Yphantis, D. A. Structure of bovine blood coagulation factor Va. Determination of the subunit associations, molecular weights, and asymmetries by analytical ultracentrifugation. *Biochemistry 23*, 1339-1348 (1984)

Laue, T. M., Shah, B. D., Ridgeway, T. M., Pelletier, S. L. Computeraided interpretation of analytical sedimentation data for proteins. *Analytical Ultracentrifugation in Biochemistry and Polymer Science*, pp. 90-125. Edited by S. E. Harding, A. J. Rowe and J. C. Horton. Cambridge, Royal Society of Chemistry, 1992.

Lee, J. C., Timasheff, S. N. The calculation of partial specific volumes of proteins in guanidine hydrochloride. *Arch. Biochem. Biophys. 165*, 268-273 (1974)

Lee, J. C., Timasheff, S. N. The calculation of partial specific volumes of proteins in 6 M guanidine hydrochloride. *Methods in Enzymology*, Vol. 61, pp. 49-57. Edited by C. H. W. Hirs and S. N. Timasheff. New York, Academic Press, 1979.

Marquardt, D. W. An algorithm for least-squares estimation of nonlinear parameters. *J. Soc. Ind. Appl. Math. 11*, 431-441 (1963)

Nelder, J. A., Mead, R. A simplex method for function minimization. *Comput. J. 7*, 308-313 (1965)

Prakash, V., Timasheff, S. N. The calculation of partial specific volumes of proteins in 8 M urea solution. *Anal. Biochem. 117*, 330-335 (1981)

Ross, P. D., Howard, F. B., Lewis, M. S. Thermodynamics of antiparallel hairpin-double helix equilibria in DNA oligonucleotides from equilibrium ultracentrifugation. *Biochemistry 30*, 6269-6275 (1991)

Steckel, F., Szapiro, S. Physical properties of heavy-oxygen water. I. Density and thermal expansion. *Trans. Faraday Soc. 59*, 331-343 (1963)

Straume, M., Johnson, M. L. Analysis of residuals: criteria for determining goodness-of-fit. *Methods in Enzymology*, Vol. 210, pp. 87-105. Edited by L. Brand and M. L. Johnson. San Diego, Academic Press, 1992a.

Straume, M., Johnson, M. L. Monte Carlo method for determining complete confidence probability distributions of estimated model parameters. *Methods in Enzymology*, Vol. 210, pp. 117-129. Edited by L. Brand and M. L. Johnson. San Diego, Academic Press, 1992b.

Svedberg, T., Pedersen, K. O. *The Ultracentrifuge*. Oxford, Clarendon Press, 1940.

Tanford, C. *Physical Chemistry of Macromolecules*. New York, J. Wiley, 1961.

Wolf, A. V., Brown, M. G., Prentiss, P. G. Concentrative properties of aqueous solutions: conversion tables. *Handbook of Chemistry and Physics*, pp. D219-D271. Editor-in-Chief: R. C. Weast. 67th ed. Boca Raton, CRC Press, 1986.

For Beckman Coulter's worldwide office locations and phone numbers, please visit "Contact Us" at **beckmancoulter.com**

CENT-1070TCH11.15-A