Analytical Ultracentrifugation in the Study of Protein Self-association and Heterogeneous Protein-Protein Interactions

Andrea Balbo and Peter Schuck

Protein Biophysics Resource, Division of Bioengineering & Physical Science, ORS, OD, National Institutes of Health, Bethesda, Maryland 20892

This introduction appeared in Cold Spring Harbor Laboratory Press, Protein-Protein Interactions: A Molecular Cloning Manual (E.A.Golemis & P.D. Adams, Edts). 2nd Edition. ISBN 0-87969-722-9, Chapter 14, pages 253-277

Address for Correspondence:
Peter Schuck
NIH/DBEPS
Bldg 13, Rm 3N17
13 South Drive
Bethesda, MD 20892
phone: (301) 435-1950
fax: (301) 480-1242
email: pschuck@helix.nih.gov

INTRODUCTION

Analytical ultracentrifugation (AUC) is a classical method for the characterization of interactions of purified proteins in dilute solutions. Conceptually, it consists simply of the application of a centrifugal force, the real-time observation of the resulting spatial macromolecular redistribution, and first-principle based quantitative analysis of the data. This requires no label or other chemical modification of the proteins, and no interaction with any matrix or surface. A key feature of sedimentation experiments for studying protein interactions is that the faster sedimenting complexes migrate through a solution of the slower sedimenting components. As a consequence, reversibly formed complexes that dissociate can readily re-associate during the experiment, thus permitting the characterization of even weak and transient interactions.

Protein complexes can be characterized with regard to their stoichiometry and the thermodynamic binding constants of complex formation. Importantly, sedimentation techniques can distinguish between multiple coexisting complexes of different stoichiometries and also provide information on self-association properties, and on mixed self- and hetero-association. The latter can be crucial information for the biophysical study of protein-interactions by other techniques. Because the migration due to the centrifugal force is opposed by molecular friction, sedimentation depends additionally on hydrodynamic properties of the molecules, providing information on the low-resolution structure of protein complexes, and enabling the detection of conformational changes.

The quantity of material required for AUC is typically in the order of a few hundred micrograms of greater than 95% purity. For the study of interactions, ordinarily two (or up to three) protein components are mixed. Due to the concentration gradients established during centrifugation, a 10 to 1000fold concentration range is typically observed, and a size-range of three orders of magnitude in molar mass can be covered in a single experiment. Interacting components under study may have sizes ranging from peptides to very large multi-protein complexes. In general, affinities in the range of 10^4 to 10^8/M can be determined, and kinetic dissociation rate constants in the order of ~ 10^{-5} – 10^{-2}/sec.

Historically, AUC was a central technique in the development of biochemistry and molecular biology. It was developed in the 1920s by The Svedberg (Svedberg and Pedersen 1940), and over the last eight decades a wealth of important knowledge has been accumulated regarding the technical implementation of analytical ultracentrifuges (Schachman 1959), the theoretical foundation of ultracentrifugation in thermodynamics and physical chemistry of macromolecules (Tanford 1961), the mathematical analysis of ultracentrifugation experiments (Fujita 1975), and the application of AUC to the study of proteins. While the technique was in decline in the 1970s and 1980s, computational and instrumental advances in the 1990s greatly facilitated the study of protein interactions, and the surging general interest in this topic led to a renaissance of this technique (Schachman 1992).

Today, the mathematical details of advanced centrifugal data interpretation are largely encapsulated in software, requiring only understanding of the experimental concepts and general knowledge in modeling and non-linear regression. This is true for all procedures described here, and a selection of appropriate software resources with help-files and tutorials will be described. AUC is an extremely versatile tool, many variations of ultracentrifugal approaches have been described in the literature, and even an overview of only the most important contributions over the last eight decades is far beyond the scope of this practical introduction. Necessarily, the following is a highly selective description only of the main concepts and the most basic protocols commonly used in our laboratory. In addition to the present text, the reader is advised to download the supplemental material which will contain more detailed descriptions of selected topics. For further

reading, recent reviews that include other approaches or more in-depth descriptions of selected SV and SE methods can be found in (Cox and Dale 1981; Correia and Yphantis 1992; Hansen *et al.* 1994; Demeler *et al.* 1997; Philo 1997b; Lebowitz *et al.* 1998; Arisaka 1999; Laue 1999; Laue and Stafford 1999; Rivas *et al.* 1999b; Philo 2000b; Schuck and Braswell 2000; Lebowitz *et al.* 2002; Schuck 2003; Cole 2004; Vistica *et al.* 2004).

OUTLINE OF PROCEDURE

General Principles

Analytical ultracentrifugation permits the real-time observation of sedimentation behavior of proteins. Two types of experiments exist, sedimentation velocity and sedimentation equilibrium, which are highly complementary.

The analytical ultracentrifuge resembles a conventional preparative centrifuge that is equipped with an optical system for the observation of the protein distribution, in real-time during the centrifugation. This permits detecting the spatial concentration gradients that are caused by the applied gravitational field, and following their evolution with time. Analytical rotors accept specialized assemblies for presenting, typically, 100 – 400 microliters of sample between windows that are optically transparent perpendicular to the plane of rotation. The optical detection system is triggered by the revolution of the rotor, such that data are acquired only during the short intervals when a particular sample assembly (of up to eight) is aligned with the optical light path (Figure 1). Two commercially available optical detection systems, a dual-beam UV/VIS spectrophotometer equipped with a monochromator and a highly sensitive laser interferometer which records the refractive index gradients, are described in more detail below.

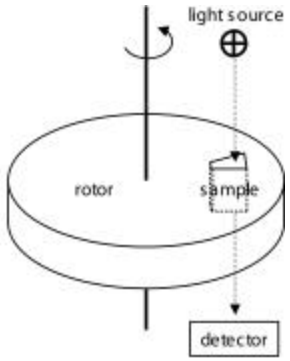

Figure 1: Schematics of the rotor and the optical system

Two basic types of experiments are possible: a) the application of a high centrifugal force and analysis of the time-course of the sedimentation process, termed sedimentation velocity (SV); and b) the application of a low centrifugal force that permits diffusion to balance sedimentation, such that a time-invariant equilibrium gradient can be observed, termed sedimentation equilibrium (SE). Both SE and SV approaches are ideally suited and have unique advantages for the study of protein interactions, and protocols for both will be given, and their relative merits discussed, below. First, as a basis for the analysis of protein interactions, it is necessary to become familiar with the principle of sedimentation for non-interacting proteins.

The sedimentation process is governed by the gravitational force, the buoyancy, and the translational friction. The gravitational force is $F_{sed} = m\omega^2 r$ (with m the protein mass, ω the rotor angular velocity, and r the distance from the center of rotation), and proportional to the square of the rotor speed. As a consequence, adjusting the rotor speed permits the study of a wide range of particle sizes, ranging from small peptides to very large protein complexes (1 kDa to 1 GDa). Following Archimedes principle, the buoyancy force $F_b = -m\bar{v}\rho\omega^2 r$ (with $\bar{v}$ the protein partial-specific volume and ρ the solvent density) opposes the sedimentation and is governed by the mass of the displaced solvent ($-m\bar{v}\rho$). Thus, protein partial-specific volumes are important, and the density effects of protein glycosylation, bound detergent, and preferential hydration may be relevant considerations (see below). Finally, the frictional force is dependent on the translational frictional coefficient, f, as well as the sedimentation velocity $F_f = s(kT/D)\omega^2$ (with the Boltzmann constant k, the absolute temperature T, and the diffusion coefficient D), where the sedimentation coefficient $s = v/\omega^2$ is a molecular constant (with v the absolute migration velocity). The translational frictional properties of molecules are frequently expressed as the ratio of the frictional coefficient to that of a smooth sphere with the protein mass and density, f/f_0. This can permit the measurement of the low-resolution shape of the proteins and their complexes in terms of Stokes radii. The sedimentation coefficient, s, is commonly measured in units of Svedberg, abbreviated S (a capital S for the unit, lower case s for the parameter), with 1 S = 10^{-13} sec. From the balance of these three forces, one can derive the Svedberg equation

$$\frac{s}{D} = \frac{M\left(1-\bar{v}\rho\right)}{RT} \qquad (1)$$

(with M denoting the protein molar mass, and R denoting the gas constant) (Svedberg and Pedersen 1940), which is a fundamental relationship between the three directly measurable quantities for a single protein component: the sedimentation coefficient (obtained from the migration of the sedimentation boundary with time in SV), the diffusion coefficient (obtained from the spread of the sedimentation boundary with time in SV), and the molar mass (obtained from the exponential gradient in SE).

In the following, building on the basic experiments for non-interacting proteins, the concepts how protein interactions can be studied are described. Readers comfortable with mathematical symbols are strongly encouraged to download the Appendix of formulas describing sedimentation, as their knowledge will add clarity and facilitate the important step of mathematical data analysis.

Sedimentation Velocity

Sedimentation velocity observes the separation of proteins due to their different rates of migration in the centrifugal field. This can be partially obscured by diffusion. The starting point in most situations is calculating the sedimentation coefficient distributions c(s), which extracts information on purity, number of species, their relative abundance and low-resolution shapes. It also is a useful tool to identify and study interactions.

Figure 2 shows an example of the evolution of concentration profiles with time in a sedimentation velocity experiment. The most important features are the sedimentation boundaries, sigmoid shaped 'shoulders' that migrate with time to increasing distances from the center of rotation. In Figure 2, several boundaries are visible migrating with different rates, originating from different size protein species. Sedimentation velocity is a hydrodynamic method that leads to a strongly size-

dependent separation, which is usually superior to that of diffusion based methods (such as size-exclusion chromatography or dynamic light scattering). Proteins covering a 1000fold range in molar mass can usually be characterized in a single run.

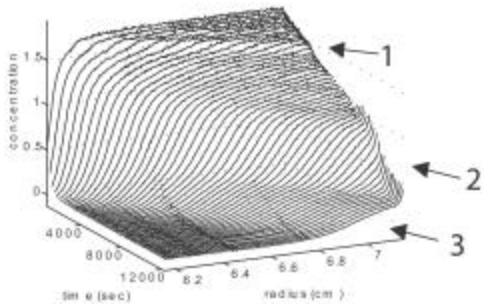

Figure 2: Sedimentation velocity profiles from the sedimentation process of a protein mixture at a high centrifugal field. The three most clearly visible boundaries are marked.

The displacement with time of the boundary midpoints determines the *s*-value, and the increasing spread of the boundaries with time is governed by the diffusion coefficient. More quantitatively, combining the law of diffusion with the centrifugal migration, a theoretical expression can be obtained for the propagation of the sedimentation boundary of a single species, termed the Lamm equation (see Eq. A1) (Lamm 1929). Although experimental data could be modeled with a single species Lamm equation to derive *s*, *D* and therefore according to the Svedberg equation the protein molar mass, *M*, in practice this is frequently not a successful approach, because of impurities and sample heterogeneity leading to excess broadening of the sedimentation boundary, and thus an underestimation of *M* (Dam and Schuck 2004). This may result from not only the superposition of clearly separating boundaries such as shown in Figure 2, but also by sedimentation of similar-sized species that form sedimentation boundaries separated by less than their diffusion broadening. (The ability to model data with a single component Lamm equation solution is a very stringent test for sample purity.)

A more general approach is the description of the sedimentation data as a differential sedimentation coefficient distribution $c(s)$, which describes a superposition of sedimentation boundaries of many species with different *s*-values (Eq. A3). In this technique, the diffusion of all species is approximately taken into account by assuming the same frictional ratio f/f_0 for all sedimenting species (Schuck 2000). This approximation exploits the fact that the frictional ratio, although a hydrodynamic shape factor, is not a strongly shape dependent quantity and is very similar for most folded proteins (ranging usually from 1.2 for very globular to ~1.8 for asymmetric and/or glycosylated proteins). The analysis results in both an estimated weight average value for f/f_0, providing average shape information of the protein ensemble under study, the distribution $c(s)$ (Figure 3), and residuals of the fit (i.e. the magnitudes of deviation of the model from the data points) for assessing if the $c(s)$ model is adequately describing the data. This analysis results in sharp peaks reflecting the abundance of different sedimenting species, resembling in some ways a chromatogram. However, it has to be kept in mind that it is a calculated curve obtained from modeling the raw data (Figure 3). Integration of the $c(s)$ peaks gives the concentrations of the different species as they were present at the start of sedimentation. If the distribution contains a single major peak, it can be transformed to a molar mass distribution $c(M)$ that allows estimation of the molar mass of the main species (Dam and Schuck 2004). The application of this approach to model systems can be found in (Schuck *et al.* 2002; Dam and Schuck 2004).

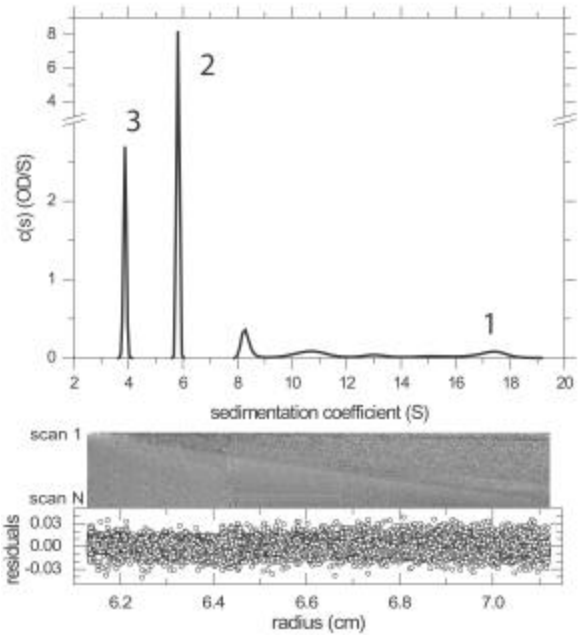

Figure 3: Sedimentation coefficient distribution c(s) from the analysis of the data in Figure 2. The distribution displays two main peaks at 4 and 6 S from two components added to the solution, their oligomers at 8 S, a heterocomplex at 10.5 S and some aggregates. The numbers on the peaks correspond to those in Figure 2. Below are the residuals, shown in two different forms. One is a plot of all radial residuals superimposed in a conventional graph. This shows the magnitude of the maximal residuals (in units of the signal, OD). Above is a bitmap representation, which encodes the magnitude of the residuals in the greyscale of the pixel at different radial values (horizontal lines) for all different scans (assembled vertically starting from scan 1 on top to the last scan on the bottom). This representation exhibits some diagonal features, which reveal systematic deviations of the fit and the data in the shape of the migrating sedimentation boundary. Vertical features and horizontal features are caused by imperfections in the optical configuration. For a good fit, vertical features are tolerable, but very little diagonal features should be observed. Although the c(s) distribution resembles in some ways a chromatogram, it has to be kept in mind that it is a curve calculated using the assumption of similar frictional ratio, and using maximum entropy regularization algorithm that generates the simplest curve consistent with the raw data. While the maximum entropy regularization usually is very successful in suppressing artificial peaks from data with limited information content, it has the property of broadening the peaks and of merging neighboring peaks when working with data of low signal-to-noise ratio.

The *c*(*s*) analysis is strictly applicable for dilute mixtures of proteins and complexes that are stable on the time-scale of sedimentation. It can provide a detailed characterization of the individual protein components with regard to the presence of stable protein complexes and their characterization with regard to size and shape, as well as with regard to the purity of the preparation, which is frequently crucial knowledge for the further characterization. Changes in the *s*-value, for example, induced by binding of a small ligand, are a measure of conformational changes that are accompanied by changes in the low-resolution shape of the protein (which is complementary to conformational changes detected via changes in secondary structural elements by circular dichroism). After determining the number of species, the analysis can frequently be refined by inserting prior knowledge on the molar mass of some species, by modeling the data with a set of Lamm equation solutions for a small number of discrete components, or by global analysis of several experiments with a hybrid approach identifying some of the peaks as discrete species with known or unknown molar mass in combination with variable segments of continuous distributions. Further, the *c*(*s*) analysis can also be used as a tool to detect and rigorously analyze slow or rapidly reversible protein interactions (see below).

Low-Resolution Shape Analysis of Proteins and Protein Complexes: For further interpretation, the observed *s*-values can be transformed to standard conditions (water at 20 ºC), according to

$$s_{20,w} = s_{\exp}\, h_{\exp}\left(1-\bar{v}\,r_{20,w}\right)\Big/\left[h_{20,w}\left(1-\bar{v}\,r_{\exp}\right)\right]$$

and the s_{20w}-value may be compared to the s-values of equivalent compact spheres of the same mass

$$s_{sphere,20w} = 0.012M^{2/3}\left(1-\bar{v}\,r\right)\Big/\bar{v}^{1/3}$$

The *s*-value of a compact sphere is a theoretical upper limit for the sedimentation coefficient. This can help to assign protein species to the observed sedimentation coefficients. If the protein mass is known (or hypothesized for a protein complex), the ratio of the observed and the spherical *s*-value equals the frictional coefficient f/f_0, a measure of shape asymmetry with narrowly bounded values (see above). The s_{20w}-value can also be compared with theoretical *s*-values derived from crystal structures of the proteins, if available (Garcia De La Torre *et al*. 2000).

If the s-values of protein complexes can be determined from the SV experiment, they can also be subjected to low-resolution shape analysis and in some cases provide additional information on the association. However, the determination of the complex s-value requires very high saturation or very high stability of the complex (Figure 6). Likewise, in rapid self-association systems the precise determination of the *s*-value of the smallest species can be far from trivial. In this case, experiments with non-associating mutant proteins or under solvent conditions that weaken the association may be useful.

SV experiments are usually conducted at very high rotor speed to minimize the effects of diffusion and to enhance the hydrodynamic separation (see the SV protocol Step I.6). If a low rotor speed is used, or if small peptides are studied, the diffusion effects will become dominant and the concentration distributions will exhibit broad features that lead to an equilibrium state in which the sedimentation is effectively balanced by diffusion throughout the entire solution column (Figure 4). This state is the subject of the thermodynamic analysis in SE.

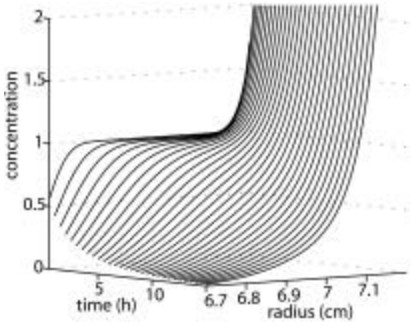

Figure 4: Approach to equilibrium.

Sedimentation Equilibrium

The basic sedimentation equilibrium experiment leads to the analysis of the exponential concentration distribution in steady state of sedimentation and diffusion. In the simplest case, this is governed by the protein molar mass.

SE is established when no change in the concentration distribution of any component is detectable. For a single species under ideal conditions (no repulsive interactions due to volume exclusion or

charge interactions), the concentration profile assumes a Boltzmann distribution (Svedberg and Pedersen 1940) (Eq. A6), which is a single exponential shape with increasing steepness towards the bottom of the cell (Figure 5).

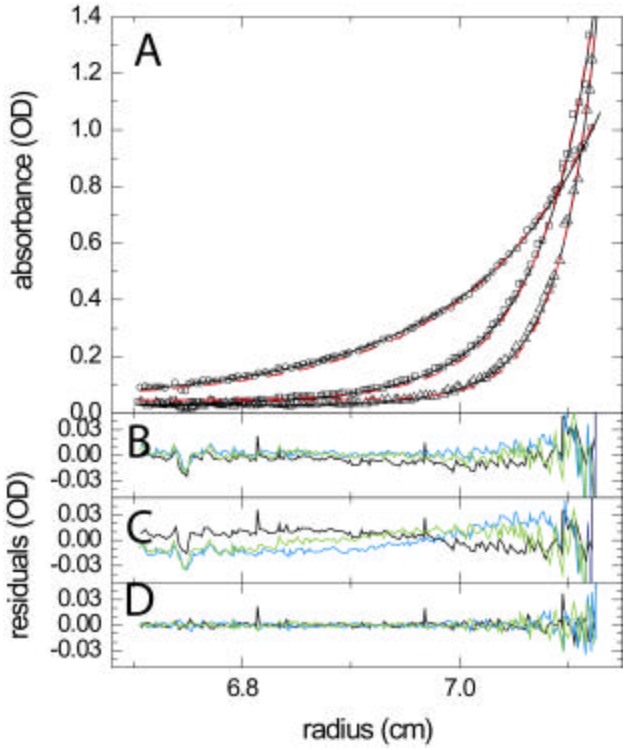

Figure 5: Sedimentation equilibrium data from a self-associating 100 kDa protein in SE at 8,000 rpm (circles), 12,000 rpm (squares), and 15,000 rpm (triangles). The data shown are is subset of a typical experiment, which was conducted at several different loading concentrations and detection wavelengths. The black solid line is a fit of the data with the correct model, with the corresponding residuals for the fits to the different rotor speeds in B. For the judging the residuals, it is very important to review how systematic the deviations are. In Panel B, they are acceptable, although not perfect because of slight systematic errors. For comparison, the residuals C are obtained with an incorrect model (indicated in A as a dashed red line), which shows clearly systematic patterns and must therefore be rejected. An example of nearly perfectly random residuals in is shown in D. SEDFIT also displays the results of a 'runs test' (Straume and Johnson 1992), a statistical measure for the systematic trends in the residuals. However, unfortunately remaining small experimental imperfections make this test too stringent in practice.

The curvature is governed by the square of the rotor speed, and the term $M_b = M(1-\overline{\nu}r)$, also referred to as the buoyant molar mass. Dependent on the rotor speed, the observed concentration range can easily cover a 100fold range, which is limited only by the radial precision and the dynamic range of the optical system used. The best results are obtained in experiments where the same solutions are subjected sequentially to different rotor speeds, as specified in detail in the present SE Protocol. (This configuration generates additional information on baseline parameters, reversibility of interactions, and mass conservation properties which will become apparent – and extremely useful – in the data analysis step, see below.) As a guide, for relatively long columns of 5 mm (the approach outlined in the protocol), and an average molar mass *M* of the protein species of interest, one typically can select a set of three rotor speeds with

$$rpm = \sqrt{\frac{100}{M\,(kDa)}} \times 8{,}000, \quad \sqrt{\frac{100}{M\,(kDa)}} \times 12{,}000, \quad and \sqrt{\frac{100}{M\,(kDa)}} \times 15{,}000 \qquad (2)$$

This generates profiles similar to those shown in Figure 5, which include both shallow and steep regions. The time to attain the first equilibrium is usually 24 – 48 hours, but possibly longer for proteins with very high molar mass, reversible proteins complexes with a small dissociation rate constant, very elongated molecules with unusually low diffusion coefficient, or buffers of high viscosity. (Note that theoretical formulas for the time to reach equilibrium do not take into account chemical reactions which may prolong the SE experiment.) Shorter columns attain equilibrium faster, but require higher rotor speeds and are lower in information content. (Nevertheless, 3 – 4 mm columns may be preferable for proteins with Mw > 100 kDa).

Experiments are conducted with a range of loading concentrations, which may be combined with detection at multiple signals or wavelengths, for example, at 230 nm, 250 nm, and 280 nm. The use of multiple rotor speeds and loading concentrations is particularly important for the analysis of protein interactions. The profiles acquired at all signals, concentrations, and rotor speeds (typically, for example, one AUC run comprising 5 samples at different concentrations, with data acquired at 3 wavelengths sequentially at 3 rotor speeds), are analyzed in a global analysis in which theoretical exponential distributions derived from molecular models of the protein species and their interactions, and are simultaneously fitted to all of the data sets. In these global analyses, the parameters describing the molecular properties (the 'global' parameters) are distinguished from those relating to particular experimental configuration in each experiment (the 'local' parameters). Additional relationships between the local parameters may be established, dependent on the experimental design. Figure 5 shows a subset of a typical analysis, and illustrates how the residuals of the fit are taken as a criterion to accept or reject the model. It is apparent that the data analysis is an important and time-consuming step in the SE study of protein interactions. Therefore, data analysis resources and the use of modeling software are described as part of the Protocol.

Frequently, the protein molar mass, the extinction coefficient, and the partial-specific volume can be predicted from the amino acid sequence. In the absence of protein interactions leading to self- or hetero-association, the sedimentation equilibrium analysis in the form of Eq. A6 is used typically to measure the buoyant molar mass. The experimental determination of this quantity is usually superior in precision to the compositional prediction from tabulated amino acid values, and it can permit the analysis of glycosylated or detergent-solubilized proteins and their interactions.

Protein Interactions in Sedimentation

Protein interactions are frequently not directly visible in the raw data, but will be apparent in characteristic differences to the sedimentation behavior of non-interacting proteins. These are amenable to highly quantitative and first-principle based interpretation in the data analysis step, which consists in the mathematical modeling. Knowledge of the characteristic signature of protein interactions in SV and SE is required for the selection of experimental conditions. The basic familiarity with the physical principles of interactions modulating the sedimentation process and the different concepts of data analysis are indispensable.

General Considerations

Although both sedimentation velocity (SV) and sedimentation equilibrium (SE) are both powerful methods for the study of protein interactions, the theoretical description of reversible protein interactions is conceptually simpler in SE. This is due to the fact that chemical equilibrium and mechanical sedimentation equilibrium can be achieved simultaneously, with the mass action law being fulfilled throughout the solution column. This description requires only equilibrium thermodynamics. In contrast, SV is governed by thermodynamics but also by the reaction kinetics and hydrodynamics (i.e. the shape-dependent frictional properties of the proteins that determine

their translational mobility in solution). These additional aspects can make the theoretical description more complex, but SV also provides a significantly larger data basis and higher resolution of macromolecular species, which may lead to more detailed information on the interaction. When comparing the two approaches in practice, it is also noteworthy that SE can be more sensitive to impurities, in particular degradation produces or low molar mass impurities. These can frequently be resolved in SV. In practice, it is usually best to conduct both SE and SV experiments because they provide complementary information. This is true, in particular, in the absence of knowledge of the interaction scheme.

In either method, it should be kept in mind that for the measurement of binding constants, both techniques require the study of proteins at a range of loading concentrations, which result in detectable populations of all species of unbound proteins and complexes formed. (In this regard, it should be noted that, due to the accumulation of material closer to the bottom of the sample column, the concentration range observed in SE significantly exceeds the loading concentration. In SV, the region of accumulation and 'back-diffusion' at the bottom is much steeper due to the higher centrifugal field, and for medium and large proteins is usually excluded from the analysis.) Also, for hetero-associations the separate characterization of the individual components is essential.

Technical terms and concepts that are important for reversible binding reactions (such as protein interactions) include: (*i*) The reaction scheme, which specifies the reacting species, the product species, and for multi-stage assemblies the pathway (e.g., $A \leftrightarrow A_2$, $A \leftrightarrow A_2 \leftrightarrow A_4$, $A+B \leftrightarrow AB$); (*ii*) Mass action law – a thermodynamic relationship between the equilibrium concentrations of the different species participating in the reaction. For the example of a simple bi-molecular reaction of two different proteins forming a 1:1 complex, this takes the form $[AB]/[A][B] = K_A$, where K_A is the equilibrium association constant ($1/K_A = K_D$, with K_D the equilibrium dissociation constant). For this case, mass action law predicts that for concentrations at K_D, equal concentrations of free and bound species coexist. (*iii*) An isotherm is the dependence of a physical property of a mixture of interacting components on the loading concentrations (keeping all other parameters constant, including the temperature). Some macroscopic observations of a mixture are dependent on the relative concentration of free and complex species, and a set of such measurements in a concentration series generates an isotherm, which can be subject to quantitative analysis for the determination, for example, of the K_D.

Sedimentation Equilibrium

Phenomenologically, interactions in SE lead to an increased steepness of the sedimentation profiles, which is a result of the population of species with higher molar mass. Although a single concentration trace does not contain information whether the formation of complexes is reversible or not, observing the redistribution of components at different rotor speeds and the comparison of populations generated at different loading concentrations does provide this information, which will become apparent in the data analysis.

A model-free phenomenological analysis of SE data of interacting systems can consist in the determination of the weight-average (or signal-average) molar mass as a function of loading concentration or loading composition $M_w(c)$, for self- or hetero-association, respectively. Interactions between proteins will revealed by an average molar mass of the mixture increasing with loading concentration. The analysis of $M_w(c)$ ignores the changes of population of the interacting species along the concentration gradient in SE, approximating the profile with Eq. A6. This is suitable for short column experiments with relatively shallow gradients, where it can be followed up

with an isotherm analysis to determine binding constants in interaction models. However, it provides only a first survey for the long column technique described here. Nevertheless, this estimation of the average molar mass values along the cell can help to identify the reaction scheme, in particular if this information is combined with the results of the initial SV analysis (see below).

If the reaction scheme is known, the SE data can be fitted with specific models for the interaction to determine the binding constants. For a monomer-dimer self-association system, the mass action law takes the form $c_2=K_{12}c_1^2$ (with concentrations and binding constants expressed in molar units), and the combination of the mass action law with the Boltzmann equilibrium distribution (Eq. A6) leads to a double-exponential concentration gradient with coupled amplitudes of the two terms (Eq. A7). Because the sedimentation creates a concentration gradient, a long-column multi-speed SE experiment contains information on an entire binding isotherm. As a consequence of the multi-exponential form of the equilibrium profiles of interacting systems, the data analysis consists of the decomposition of the raw data into the different exponential terms. In this analysis, it is essential that the buoyant molar mass M_b is known, either from prediction based on amino acid sequence and/or mass spectrometry. Alternatively, it may be possible to measure M_b directly under buffer conditions or with mutant proteins where the self-association is absent.

Similar as for self-association, the analysis of heterogeneous interactions can proceed by combination of the mass action law with the mechanical sedimentation equilibrium. For example, the formation of a 1:1 complex follows $c_{AB} = K_a c_A c_B$, and the radial distribution of the total signal follows a tri-exponential shape (Eq. A8), and further generalizations are possible, for example including multiple binding site models with cooperativity. However, in general the limitation exists that each protein species will generate an additional exponential term, and in practice, the identification of more than four terms is very difficult even with advanced global analysis and multi-signal detection.

An important difference between self- and hetero-associations in SE is that for hetero-associations, the buoyant molar mass values $M_{b,A}$ and $M_{b,B}$ can be determined directly by studying the individual protein components separately. On the other hand, if the molar mass values of the components differ by less than ~20% or more than a factor ~ 5, the unbound and/or the bound components cannot be directly distinguished from the shape of the sedimentation profiles due to mathematical correlation of their exponential signal contributions, making a direct analysis of the binding constant impossible. This problem can be circumvented, however, if the protein components exhibit significant different absorption properties (for example, due to extrinsic chromophoric labels), enabling their spectral discrimination in the multi-wavelength analysis. Alternatively, such interactions can be studied by imposing the mathematical constraint on the analysis that the total mass of all components is conserved during sedimentation (Philo 2000b; Vistica *et al.* 2004), which is facilitated if the series of loading concentrations of both components are chosen to be serial dilutions of an equimolar stock mixture, or if the concentration of one component is kept constant in a titration series with the second component (Vistica *et al.* 2004).

Sedimentation Velocity

The analysis of protein interactions proceeds differently in SV, with different approaches dependent on the stability of the complexes formed. Starting point for the analysis is the $c(s)$ distribution obtained at a large range of different loading concentrations of individual protein components, and at different concentrations and/or molar ratios of mixtures for the study of hetero-associations. Although the $c(s)$ model is based on the superposition of non-interacting species (Eq. A3), it is nevertheless an excellent tool to detect interactions, to develop models for the interaction scheme, and to serve as the basis for the second-stage analysis of binding isotherms (Schuck 2003).

Interactions can be detected through the emergence of new peaks at higher concentrations, shifts in the ratios of the peak areas, and/or shifts in the peak positions. Peaks that remain constant

in position but change in the relative area fraction generally indicate the formation of complexes that are stable on the time-scale of sedimentation. (In this case, it can be very important in practice to provide sufficient time to establish chemical equilibrium in the loading mixture by sufficiently long incubation prior to the experiment.) Shifting peak positions indicate more rapid chemical interconversion of species during the sedimentation. Here, the reaction causes the sedimenting system to assume an average sedimentation rate in between those of the reacting species, shifting according to their relative population (Figure 6). The chemical conversion on the time-scale of sedimentation can also cause a broadening of the sedimentation boundary in excess of the usual diffusion broadening. In some cases, this may be detected in the *c*(*s*) modeling through an unrealistically low apparent frictional ratio, and/or a poor quality of fit.

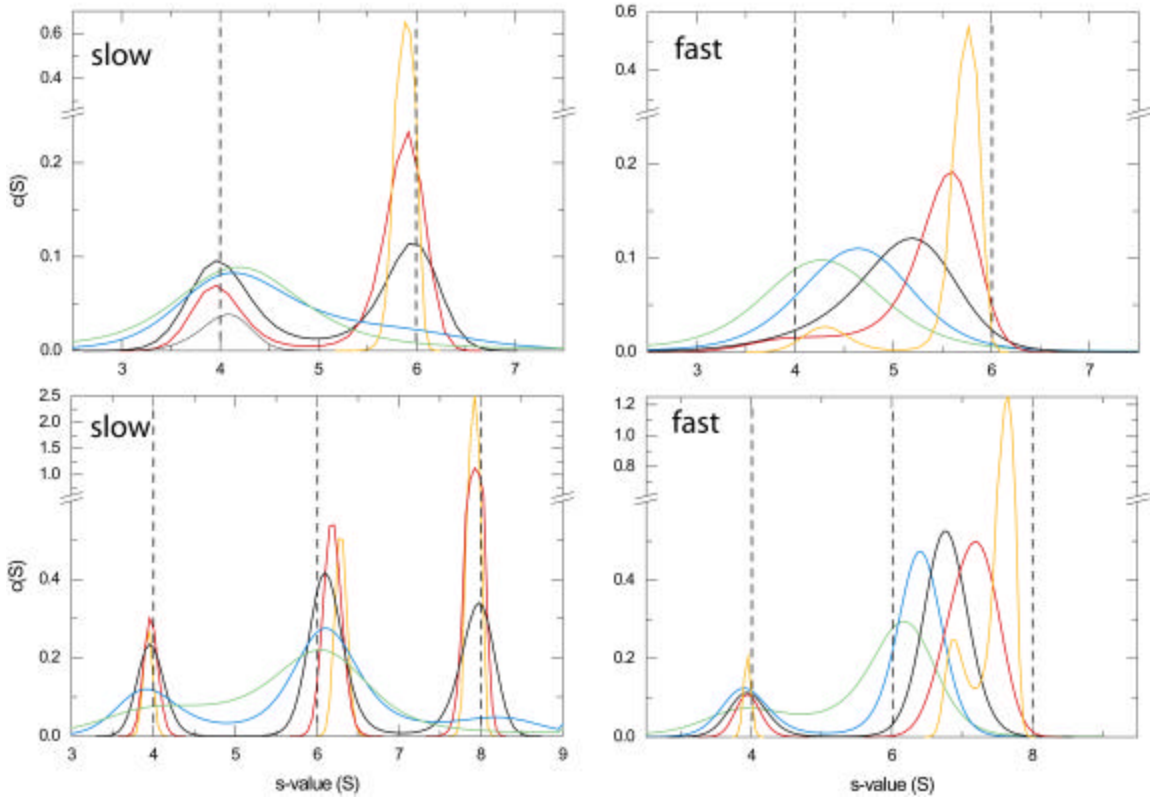

Figure 6: Examples of c(s) distributions for an interacting system. Shown are (scaled) distributions from a self-associating system of a 50 kDa protein sedimenting at 4 S forming a dimer with 6 S (top row) and a heterogeneous association of a 50 kDa (4 S) and a 100 kDa protein (6 S) forming a 1:1 complex with 8 S (bottom row). The influence of the reaction kinetics can be seen in comparison of the slow interaction ($k_{off} = 5\times10^{-5}$/sec, left column) and a fast interaction ($k_{off} = 5\times10^{-3}$/sec, right column) on the time-scale of sedimentation (at 50,000 rpm). For both systems, the protein concentrations were 0.1 (green), 0.3 (blue), 1 (black), 3 (red), and 10-fold (yellow) the K_D. Sedimentation signals were simulated for the interference optical detection system. The distributions at the lowest concentrations are broadened because of the limiting signal/noise ratio, whereas the distributions obtained from high concentrations are sharper. While for slow reactions the peak area changes and the peak positions remain relatively constant, the fast reactions produce peaks with positions at increasing s-values for higher concentrations. Note that even at concentrations 10-fold above K_D, the highest peak does not correspond to the complex s-value. For slow reactions, the peaks reflect approximately the populations of the sedimenting species, while at fast reactions the peaks reflect the effective sedimentation properties of the reacting system.

If the complexes exhibit moderate or higher stability (left panels of Figure 6), the *c*(*s*) profile and the position of the peaks can give valuable insights into the reaction scheme and the stoichiometry (e.g., with assumptions on the frictional ratio of the complex, which may be aided by

a known frictional ratio of the interacting proteins). Prerequisite, however, is that a good fit of the data is achieved with the $c(s)$ model. Suboptimal fits may indicate that the chemical reaction is faster and the complexes are more transient, in which case the boundary broadening originating from dissociation of the complexes during the SV experiment is incorrectly deconvoluted as diffusional broadening, which can generate artificial intermediate peaks in $c(s)$.

Independent of the reaction kinetics (except for very small possible corrections for radial dilution (Schuck 2003)), the integration of the *c(s)* profiles yields a signal-average sedimentation coefficient $s_w(c)$ dependent on the loading composition. This can be subjected to a rigorous analysis with a model for the binding isotherm, from which the binding constants along with the *s*-value of the complex can be determined (Figure 7). The general form of such an isotherm of weight-average or signal-average s-values is given in Eqs. A9 and an example for the reversible formation of a heterogeneous 1:1 complex is given in Eq. A10. (Such models are implemented in the analysis software SEDPHAT.) The physical basis of this approach are mass-balance considerations, also termed 'second-moment methods' in the ultracentrifugation literature. They rely on a high quality fit of the sedimentation boundary, such that the area under the sedimentation profiles and the depletion of total amount of material can be accurately measured, but they are completely independent on the theoretical basis of the boundary model. As a consequence, although the $c(s)$ analysis is theoretically based on the superposition of non-interacting species, it is the most precise method to determine the isotherm of weight-average *s*-values s_w (Schuck 2003). For hetero-associations, the study of $s_w(c_A,c_B)$ can reveal significant information on the stoichiometry of the association. This is illustrated in Figure 7 for a titration series with a constant loading concentration of component A and with a variable loading concentration of component B.

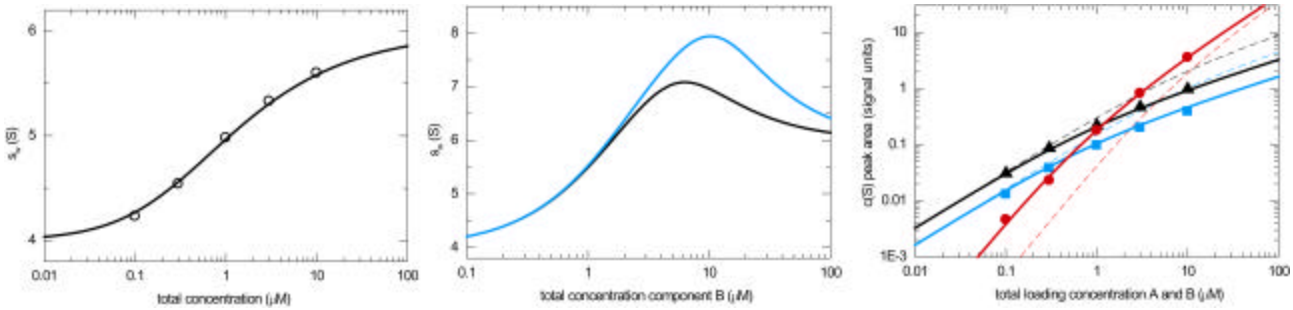

Figure 7: Different variations of isotherm analyses for interacting species. Left: $s_w(c)$ isotherm of the fast self-association shown in Figure 6 (top left), obtained by integration of c(s) over the entire s-range for each loading concentration. Middle: Comparisons of the shape of the $s_w(c)$ isotherm for heterogeneous interactions with a 1:1 complex (black), and for an interaction with two sites forming 1:1 and 1:2 complexes (blue), both for an experiment with constant loading concentration of A (5 micromolar) and a titration series of component B. In the two-site model, equivalent sites were assumed, with the s-value of the 1:2 complex at 9.5 S. Other binding parameters are as in Figure 6 (bottom row), with a K_D of 1 micromolar, s_A = 4 S and s_B = 6 S. Because the weight-average s-values drops if one component is in a molar ratio exceeding the maximal complex stoichiometry, the shape of the isotherm will exhibit a peak. Note that the peak position is different for the 1:1 and 2:1 interactions. Right: Isotherms of the species signals of free A (blue), free B (black), and complex (red) as a function of equimolar loading concentration. The partial species concentrations were obtained from integrating the peak areas of the c(s) curves from the slow heterogeneous association shown in Figure 6 (bottom left). The best-fit K_D is 1.09 micromolar, slightly deviating from the theoretical value of 1 micromolar due to the finite stability of the complex. For comparison, the shape of the isotherms are shown if the K_D would assume a value of 10 micromolar (dotted lines).

For moderately stable or highly stable complexes in which the peak *s*-values of the different species can be resolved and remain at a constant position, the sedimentation coefficient distribution

$c(s)$ can also be analyzed by integration of each peak area for each loading concentration, which results in a set of isotherms for the signals from each individual species. This set can be globally fitted with a mass action law model to determine the binding constant (Figure 7). For this more detailed analysis to be reliable, it is important that the quality of fit of the $c(s)$ analysis is very good. In particular for apparent multi-stage assembly reactions, independent confirmation of either the reaction scheme or the time-scale of the reaction kinetics can be important in order to avoid possible ambiguities resulting from similarities of the boundary shapes to those of unrecognized faster reaction kinetics (see below).

So far, the SV analysis has taken advantage of the $c(s)$ distribution as a primary tool to model the data, which is followed by a second stage analysis taking into account the interaction between the different species. For fast interactions, in particular, this does not completely utilize the information content of the SV profiles. If a model for the interaction scheme has been built, a theoretical description of the sedimentation/diffusion/reaction process can be achieved by solving a set of generalized Lamm equations that include reaction terms describing the interconversion of species during sedimentation (Eq. A5). For chemical reactions with fast interconversion (k_{off} > 0.01/sec), chemical equilibrium can usually be considered instantaneous on the time-scale of sedimentation. The direct modeling of the sedimentation profiles with specific Lamm equation models incorporating the chemical reactions should be performed in a global fit of data at different loading concentrations. This analysis can be more difficult in practice, and has higher requirements for sample purity than the isotherm approach. However, it takes full advantage of the information contained in the shape of the sedimentation profiles on sedimentation coefficients, equilibrium binding constants, and the dissociation rate constant of the complexes (Figure 8). The highest sensitivity for reaction kinetics is in the time-scale of 10^{-3} to 10^{-4}/sec for chemical off-rate constants of the complexes.

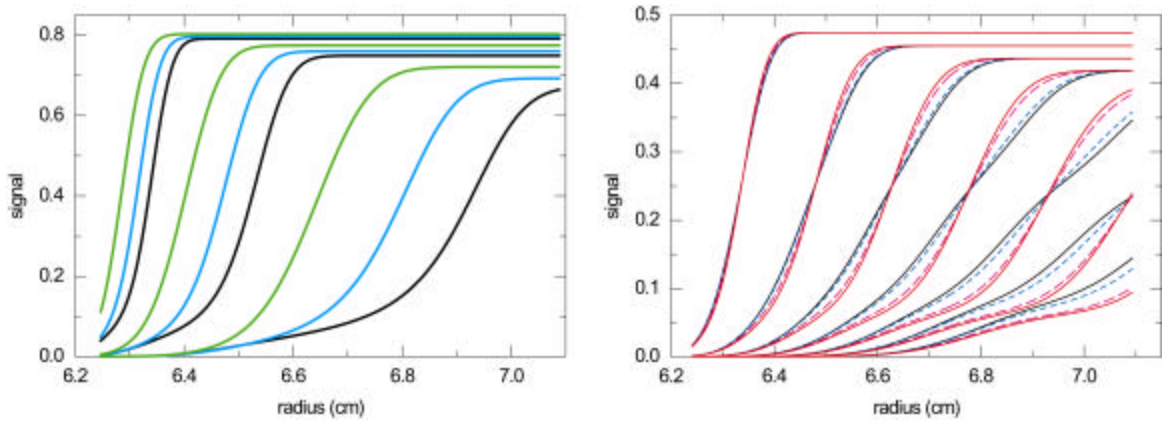

Figure 8: Characteristic shape of sedimentation profiles from interacting systems: Left: Sedimentation boundaries from a rapidly reversible self-association: monomer-dimer (green), monomer-trimer (blue), and monomer-tetramer (black). Three profiles are taken for each at the same time points. Besides the faster migration of the monomer-tetramer system, with higher reaction order the boundary also exhibits an increasingly bimodal shape with a shallow and steeper region. Right: The boundary shapes also differ significantly for systems with different reaction kinetics. The profiles are shown for a heterogeneous association of two proteins at equimolar concentrations at K_D, for chemical off-rate constants of 10^{-2} (red), 10^{-3} (dashed magenta), 10^{-4} (short dashed blue), and 10^{-5}/sec (black). (The other sedimentation parameters are the same as in Figure 6.)

Despite the characteristic information on the reaction kinetics and reaction scheme contained in the boundary shapes of SV, it must be recognized that it may not always be possible to unambiguously determine the correct model. In particular, there can be similarities in the boundary shapes of rapid higher-order complex formation with those from slow multi-stage reactions. This

can be understood considering that the boundary broadening caused by fast chemical interconversion can be similar to the boundary shapes including a more slowly converting reaction intermediate. It can be expected that this ambiguity of the boundary interpretation is more problematic for small proteins, where the hydrodynamic separation and the chemical reaction is masked by higher diffusion. In this case, even the global analysis of SV data from different loading concentrations may not be sufficient to discriminate between the models of slow multi-stage versus fast higher-order association. However, highly valuable information on the association scheme can be added from sedimentation equilibrium data, and from other biophysical techniques that can provide information on either the time-scale of the reaction kinetics, or precise details on the reaction scheme.

Additional Experimental Factors

Both SE and SV analyses can be performed considering populations of incompetent protein fractions (for example, populations of misfolded proteins). Other potential complications are volume changes upon associations, in which case the buoyant molar mass of the complex cannot be calculated as the sum of the buoyant molar masses of the components, leading to pressure dependent association constants. However, except for a few well-known cases, such pressure effects in centrifugation are rare and at the rotor speeds in most ultracentrifugation experiments (even SV) usually not detectable (Schachman 1959; Harrington and Kegeles 1973). Deviations in the sedimentation behavior outlined above can be observed at high protein concentrations, which can lead to hydrodynamic and thermodynamic non-ideality in SV consisting of a concentration-dependence of s and D, and to thermodynamic non-ideality in SE. Non-ideal conditions can also be observed in low ionic strength buffers when electrostatic repulsion of the proteins can alter their sedimentation behavior. Usually, in SE such effects are absent at protein concentrations below 1 mg/ml and in the presence of 10 mM salt, although hydrodynamic non-ideal SV takes place at much lower protein concentrations for highly elongated structures, such as fibers. Theories and data analysis treatments for such systems are available (Dishon *et al.* 1967; Fujita 1975; Chatelier and Minton 1987; Rowe 1992). Further, the theory and experimental practice of SE has been developed for protein interactions under highly crowded conditions (Rivas *et al.* 1999a). Other interactions studied by AUC include those arising from protein-solvent interactions (Costenaro and Ebel 2002), the determination of second viral coefficients in a more general framework of protein-protein interaction (for example, in the study of crystallization conditions (Solovyova *et al.* 2001)), or physical entanglement of protein fibers (for example, of amyloid fibers (MacRaild *et al.* 2003)), or small molecule binding to proteins in drug development (Arkin and Lear 2001).

Optical Systems

The optical system is the key experimental device, besides the centrifuge. Two systems are available with very different properties, and different characteristics for SE and SV. The type of protein interaction to be studied, the extinction properties of the protein components, the concentration range needed to populate the complex, as well as the buffer requirements dictate the choice of the optical detection system. In turn, this can impose specific requirements for the experimental setup.

The two standard optical systems that are part of the commercial analytical ultracentrifuge (Beckman Coulter) are a UV/VIS absorption optical scanner (abbreviated ABS in the following) (Hanlon *et al.* 1962), and refractive index sensitive laser interferometric (IF) imaging system (Richards and Schachman 1959). They differ considerably in the properties of the optical signal, the required sample preparation, and the technical details of conducting the experiments, but both

are highly useful in SE and SV AUC. Generally, the choice of the optical signal and the protein extinction determines the detection limits and the range of binding constants that can be characterized, which is governed by the necessity to detect both the free and complex species. It typically ranges from low nM to mM.

The ABS system consists of a flash lamp that is pulsed synchronously with the rotor revolution, a monochromator to select the detection wavelength, a slit assembly to scan the radial positions, and a photomultiplier tube. It functions very similarly to a double-beam spectrophotometer with a linear range of 0 – 1.5 OD. The wavelength range is from the far ultraviolet (UV) to the visible (VIS) (190 – 800 nm), and typical wavelengths used are: most commonly 280 nm for the tryptophan and tyrosine extinction of proteins; 230 nm to take advantage of the several fold higher protein extinction from the peptide backbone while also utilizing a strong emission peak of the light source; the minimum of the extinction at ~250 nm to permit the detection of (2 – 3 fold) higher protein concentrations without compromising the linearity of the signal (which would be more problematic in regions of steep extinction profiles, such as in the 290 – 300 nm range); and VIS absorption to take advantage of chromophoric labels, for example, fluorescein or rhodamine based tags. Sample loading concentrations are commonly 0.1 – 0.5 OD for SE experiments, or 1 – 1.5 OD for SV experiments, although this can be quite flexible. The ABS system requires windows with good UV transmission, usually quartz. The noise structure is governed by statistical noise, in the presence of small systematic radial-dependent offsets.

The most important advantage provided by the absorption system is its selectivity. Proteins can be detected separately from buffer components (with some exceptions), and ordinarily no precise match of the reference buffer is needed. A large concentration range can be covered through the choice of wavelengths with different protein extinctions. Further, proteins with different extinction properties can be distinguished, in particular those with extrinsic or intrinsic chromophores in the VIS range. Multi-wavelength analyses in SE routinely take advantage of this to enhance the precision of the measured binding constants, or to distinguish the number and stoichiometry of heterogeneous complexes (Servillo *et al.* 1981; Lewis *et al.* 1993; Schuck 1994). The absorption system has a limited linearity, which excludes the use of buffer components that produce a high baseline absorption offset (for example, nucleotides such as ATP, GDP). HEPES, DTT, or β-mercaptoethanol also absorb in the UV, but are tolerable at low to moderate concentrations for 280 nm detection.

The Rayleigh laser interferometer (IF) images optical path length differences caused by changes in the local refractive index due to macromolecular redistribution. Although the refractive index increment of proteins in water is not very large, the imaging system has an exquisite sensitivity, permitting a minimum detection limit for proteins of a few micrograms/ml, typically comparable to that provided for by far-UV absorption at 230 nm. The recorded quantity is the displacement of interference fringes arising from superposition of two laser beams projected to a camera. It should be noted that the refractive index increment of proteins is largely independent of amino acid composition, but will change with glycosylation or bound detergent. Commercial instruments are configured so that a protein concentration of 1 mg/ml in 12 mm centerpieces will generate a signal of 3.3 fringes (this corresponds to a molar signal increment analogous to the extinction coefficients of $2.75 \times M_w$ (fringes/Mcm)). In comparison, the typical instrumental noise is ~0.005 fringes. Besides the high sensitivity and signal-to-noise ratio, the independence from the protein extinction properties, other important advantages of the IF system are an unlimited linearity, the ability to resolve steeper gradients, and a high speed of data acquisition enabled by the imaging system (in contrast to the more than 10-fold slower scanning mode of the ABS system). This

enables the study of high molar mass complexes at higher rotor speed (e.g., 1 MDa particles at 50 or 60,000 rpm).

As a consequence of the high sensitivity of the IF system, it has the drawback that any mechanical imperfection that causes shifts in the optical pathlengths in the nm range or larger will generate a signal. This includes the surface roughness and imperfections of all optical elements, and even very small vibrations of the rotor or any optical component. These imperfections result in a characteristic noise structure of a systematic radial-dependent but time-invariant (TI) baseline offset, combined with a radial-invariant (RI) but time-dependent baseline offset. Fortunately, these signals can be readily distinguished from the signals arising from macromolecular redistributions, either computationally by using algebraic noise decomposition techniques in SV (Schuck and Demeler 1999) and SE (Vistica *et al.* 2004), through time-difference techniques (Stafford 1992), or experimentally by using water blanks in SE (Ansevin *et al.* 1970). However, all approaches requires the absence of mechanical deformations of the windows and the cell assemblies during centrifugation, which can be achieved by using sapphire windows (and 'aging' the cell assembly for SE; see the protocol) (Ansevin *et al.* 1970). In practice, it may not always be possible to assess visually the protein distributions directly until the systematic noise contributions have been removed, but this represents no drawback in practice.

Another kind of signal that is superimposed on that of macromolecular redistribution is caused by the refractive index changes from the sedimentation of buffer components. For example, the sedimentation of 100 mM NaCl can generate a characteristic signal that is usually an order of magnitude larger than that of the proteins under study. These signals can be eliminated if the buffer composition in the reference and the sample chambers are precisely matched. This can be ensured by equilibrium dialysis, or by gel permeation chromatography. In addition, for the buffer components to generate signals that exactly match in the reference and sample sector, it is also very important to precisely match the reference and sample volumes.

In summary, the ABS system has a substantial advantage of selectivity, but is generally limited in linearity and the speed of scanning, while the IF system is very fast and has unlimited linearity but is affected by more technical considerations. The selection of the best optical system will depend on the type of experiment, the required loading concentrations, and the buffer composition. Variations of the protocol for either choice are given below. In practice, the IF system is used more frequently for SV and the ABS system for SE experiments. More detailed considerations guiding the selection and the technical aspects for executing SE and SV experiments with both the IF and ABS system can be found in Tables 1 and 2.

Buoyancy and Partial-Specific Volumes

The partial-specific volume of the proteins under study can be an important factor in the sedimentation behavior. The proper consideration of protein buoyancy can be crucial for selecting buffer conditions and interpreting the results of the AUC experiment. Proteins with components other than amino acids, such as enzymes with prosthetic groups, glycoproteins or solubilized membrane proteins require special considerations, as do buffer components that increase solvent density, including glycerol, sucrose, high salt.

Sedimentation is measured in solution and requires the consideration of buoyancy. Buoyancy describes the effect that the effective sedimenting mass is reduced by the mass of the displaced solvent. In some exceptional cases, this can even lead to flotation rather then sedimentation. Therefore, the partial-specific volume $\bar{v}$ or its inverse, the density ρ, of the protein, including the

solvation shell, should be known before the ultracentrifugation experiment, in addition to the solvent density. The precision of $\bar{v}$ is important because with typical values for protein of 0.70 – 0.75 ml/g, the buoyancy term $(1-\bar{v}r)$ will amplify the errors of $\bar{v}$ by a factor 3 – 4.

For proteins that do not self-associate, frequently the molar mass is known from the amino acid sequence, and the buoyant molar mass (and consequently the partial-specific volume) can be directly measured in sedimentation experiments. For the study of heterogeneous associations, this experimental determination is usually sufficient, although it should be compared with theoretical expectations. The determination by sedimentation is not easily possible, however, for most self-associating systems.

A first, and frequently sufficient approximation for $\bar{v}$ is the theoretical prediction based on the amino acid composition. This calculation along with the prediction of the solvent density can be conveniently performed with the software SEDNTERP, which has implemented the required tabulated data (Laue *et al.* 1992). There are two important situations where this compositional approach is not sufficient. One is for proteins or protein complexes that contain components other than amino acids, such as glycosylation, prosthetic groups, PEG, detergent micelles, or bound lipids. These can have profound effects on the buoyancy and need to be taken into account. The second possible problem with the compositional determination occurs when protein solvation becomes important for the buoyancy via preferential solvation effects. This can potentially be the case when using buffers that have a density significantly different from water (e.g., caused stabilizing agents such as a low percentage sucrose or glycerol), in very high concentrations of salt or other small molecule components, or with mixed solvents. How to proceed in these cases is described in more detail in the Appendix 2. Treatments for glycoproteins and detergent solubilized membrane proteins are established.

Experimental Approaches

There are many different variations of AUC experiments, some of which provide specific advantages for certain types of interacting protein systems. In their choice, a key factor to consider is how much characteristic information on the interaction can be extracted in the data analysis.

There are many different experimental configurations for both SE and SV. For SE, this presentation focuses on long-column SE, which provides the highest information content and is suitable to multi-wavelength analysis. However, the disadvantage of this method is the relatively long time required to attain equilibrium, in particular when conducting SE at several rotor speeds. For proteins that do not have the required stability, short-column SE can be used (Correia and Yphantis 1992). Here, only a single weight-average molar mass is extracted from the sedimentation gradients, and the information on the interaction is derived the isotherm of the weight average molar mass as a function of loading composition. Another more rapid method for measuring molar mass is the Archibald method (Schuck and Millar 1998), which uses a long column but measures the slope close to the meniscus and the bottom which, throughout the entire experiment, contains direct information on the molar mass. For SV experiments, alternative experimental configurations include analytical zone centrifugation (Lebowitz *et al.* 1998), isopycnic density gradient sedimentation (Ifft 1976), and sedimentation velocity with pseudo-absorbance detection (Kar *et al.* 2000) (which allows doubling the number of samples per run).

Obviously, which technique to choose depends on the protein under study, its stability, buffer requirements, available concentrations, the interaction mode, the anticipated binding constants, etc. For a researcher with a solid background in physical biochemistry, the complexity of AUC does not necessarily lie in the depth of the required knowledge. Further, the mathematical approaches described in the Appendix are largely encapsulated in the modern analysis software. The biggest

challenge is rather the quickly diverging decision tree for choosing the optimal experimental setups, each requiring special technical considerations. It is clearly a huge benefit that the method has evolved over many decades and generated powerful tools for studying many aspects of protein interactions in solution, but it also makes the required detailed knowledge very broad. A second difficulty compounding the first is that the early decisions of experimental setup are closely linked to the likelihood of success in the mathematical modeling step of the final data analysis.

Tables 1, 2, and 3 may serve as a first guide for which technique to choose, but these are clearly incomplete, and it should be noted that some of the recommendations and comments may not apply to some particular protein system. Despite these problems, it is the consensus of many laboratories to start with a SV experiment. This can verify the purity of the preparation, and yields valuable information on the association state and number of species even in a first superficial analysis. It can also lead to the identification of the binding scheme, and frequently at least the qualitative characterization of the binding kinetics. For interacting systems of suitable purity, the more time-consuming SE is conducted after the initial characterization by SV. In the following, two basic protocols are presented, one for a typical SV, and one for SE with the goal of characterizing a heterogeneous interaction.

Both protocols also contain an overview of the workflow for the data analysis, which is an integral – and frequently the most time-consuming – step of the experiment. The protocols are based on the software package SEDFIT and SEDPHAT. Other analysis programs are also available and well established, for example for SV those focusing on different mathematical approaches such as the *dc/dt* transformation (DCDT+) (Stafford 1992; Philo 2000a), the van Holde-Weischet method (ULTRASCAN) (van Holde and Weischet 1978; Demeler *et al.* 1997), or an alternative implementation of boundary modeling of reactive systems in the time-difference mode (SEDANAL) (Stafford and Sherwood 2004) based on finite element solutions of the Lamm equation (Cox and Dale 1981), and the direct boundary modeling softwares LAMM (Behlke and Ristau 2002) and SVEDBERG (Philo 1997a). For SE, WINNONLIN (Johnson and Straume 1994) is an alternative for the analysis of self-associating systems. The presentation here is confined to SEDFIT/SEDPHAT which is preferred in our laboratory, partly because it incorporates the largest set of available analytical methods in a single user interface for both SV and SE (Schuck 2003; Vistica *et al.* 2004). This includes fast finite element Lamm equation solutions (Schuck 1998; Schuck *et al.* 1998), the diffusion deconvoluted $c(s)$ distribution (Schuck 2000; Schuck *et al.* 2002) and a least-squares $g^*(s)$ distribution (Schuck and Rossmanith 2000), boundary modeling for kinetically controlled or fast reactive systems (Cox and Dale 1981), systematic noise analysis in SV (Schuck and Demeler 1999) and SE (Vistica *et al.* 2004), and general mass conservation models for self- and hetero-associating systems for SE (Vistica *et al.* 2004). SEDPHAT also permits the global analysis of different data types. Besides SV and SE this also includes $s_w(c)$ and $M_w(c)$ isotherms and autocorrelation functions from dynamic light scattering. SEDFIT/SEDPHAT both have large help-systems on the internet at the site www.analyticalultracentrifugation.com, including tutorials for more detailed aspects of their use; workshops for sedimentation analysis with SEDFIT/SEDPHAT are currently being held at the National Institutes of Health, Bethesda.

A very powerful refinement of the data analysis involves the incorporation of prior knowledge. For different systems, prior knowledge can take many different forms, ranging from the known molar masses of the interacting components (for example, a small ligand binding to a very large protein), known binding kinetics or binding constants, known number of species, or additional available data from other sources, such as dynamic light scattering. Many options are available in SEDFIT/SEDPHAT. It is beyond the scope of this general protocol to discuss the many special cases of data analysis problems, and the sometimes modified experimental configurations that may follow from such knowledge.

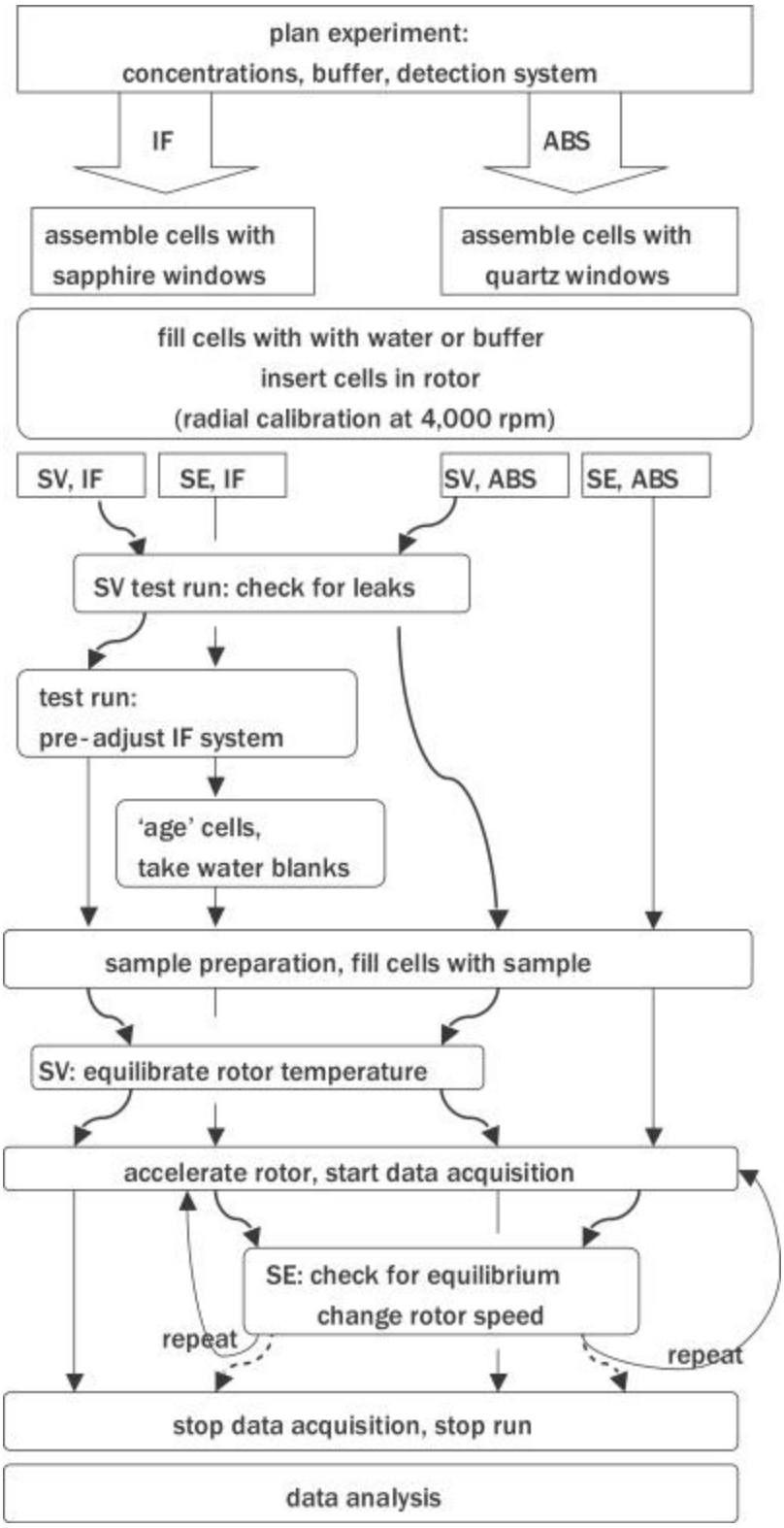

Figure 9: Flowchart of the analytical ultracentrifugation experiment. For different requirements of SV and SE experiments in IF and ABS detection, also consult Table 1.

SUMMARY

Sedimentation techniques allow the detailed characterization of protein interactions. This includes the determination of binding constants, at least semi-quantitative information of binding kinetics, and the measurement of the hydrodynamic shape of the complexes. Analytical ultracentrifugation techniques are firmly based on first principles. There is no interaction with a surface, no requirement to label or immobilize the protein, as the measurement takes place in free solution. In contrast to many other techniques, protein self-association and self-association coupled to heterogeneous protein-protein associations can be studied. A key advantage is that many co-existing complexes can be detected and characterized. This is true, in particular, for sedimentation velocity.

REFERENCES

Ansevin AT, Roark DE, Yphantis DA. 1970. Improved ultracentrifuge cells for high-speed sedimentation equilibrium studies with interference optics. Anal. Biochem. 34:237-261.

Arisaka F. 1999. Applications and future perspectives of analytical ultracentrifugation. Tanpakushitsu Kakusan Koso 44:82-91.

Arkin M, Lear JD. 2001. A New Data Analysis Method to Determine Binding Constants of Small Molecules Using Equilibrium Analytical Ultracentrifugation with Absorption Optics. Anal. Biochem. 299:98-107.

Behlke J, Ristau O. 2002. A new approximate whole boundary solution of the Lamm differential equation for the analysis of sedimentation velocity experiments. Biophys Chem 95(1):59-68.

Bevington PR, Robinson DK. 1992. Data Reduction and Error Analysis for the Physical Sciences. New York: Mc-Graw-Hill.

Chatelier RC, Minton AP. 1987. Sedimentation equilibrium in macromolecular solutions of arbitrary concentration. II. Two protein components. Biopolymers 26:1097-1113.

Cole JL. 2004. Analysis of heterogeneous interactions. Methods Enzymol 384:212-232.

Correia JJ, Yphantis DA. 1992. Equilibrium sedimentation in short solution columns. In: Harding SE, Rowe AJ, Horton JC, editors. Analytical ultracentrifugation in biochemistry and polymer science. Cambridge, U.K.: The Royal Society of Chemistry. p 231-252.

Costenaro L, Ebel C. 2002. Thermodynamic relationships between protein-solvent and protein-protein interactions. Acta Crystallogr D Biol Crystallogr 58(Pt 10 Pt 1):1554-9.

Cox DJ, Dale RS. 1981. Simulation of transport experiments for interacting systems. In: Frieden C, Nichol LW, editors. Protein-protein interactions. New York: Wiley.

Dam J, Schuck P. 2004. Calculating sedimentation coefficient distributions by direct modeling of sedimentation velocity profiles. Methods Enzymol 384:185-212.

Demeler B, Saber H, Hansen JC. 1997. Identification and interpretation of complexity in sedimentation velocity boundaries. Biophys J 72(1):397-407.

Dishon M, Weiss GH, Yphantis DA. 1967. Numerical simulations of the Lamm equation: III. Velocity centrifugation. Biopolymers 5:697-713.

Fujita H. 1975. Foundations of ultracentrifugal analysis. New York: John Wiley & Sons.

Garcia De La Torre J, Huertas ML, Carrasco B. 2000. Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. Biophys J 78(2):719-730.

Hanlon S, Lamers K, Lauterbach G, Johnson R, Schachman HK. 1962. Ultracentrifuge studies with absorption optics. I. An automatic photoelectric scanning absorption system. Arch. Biochem. Biophys. 99:157-174.

Hansen JC, Lebowitz J, Demeler B. 1994. Analytical ultracentrifugation of complex macromolecular systems. Biochemistry 33:13155-13163.

Harrington WF, Kegeles G. 1973. Pressure effects in ultracentrifugation of interacting systems. Methods Enzymology 27:106-345.

Ifft JB. 1976. Sedimentation equilibrium of proteins in density gradients. Biophys Chem 5:137-157.

Johnson ML, Straume M. 1994. Comments on the analysis of sedimentation equilibrium experiments. In: Schuster TM, Laue TM, editors. Modern Analytical Ultracentrifugation. Boston: Birkhäuser. p 37-65.

Kar SR, Kinsbury JS, Lewis MS, Laue TM, Schuck P. 2000. Analysis of transport experiment using pseudo-absorbance data. Anal. Biochem. 285:135-142.

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

Laue TM. 1999. Analytical centrifugation: equilibrium approach. Current Protocols in Protein Science:20.3.1-20.3.13.

Laue TM, Shah BD, Ridgeway TM, Pelletier SL. 1992. Computer-aided interpretation of analytical sedimentation data for proteins. In: Harding SE, Rowe AJ, Horton JC, editors. Analytical Ultracentrifugation in Biochemistry and Polymer Science. Cambridge: The Royal Society of Chemistry. p 90-125.

Laue TM, Stafford WFI. 1999. Modern applications of analytical ultracentrifugation. Annu. Rev. Biophys. Biomol. Struct. 28:75-100.

Lebowitz J, Lewis MS, Schuck P. 2002. Modern analytical ultracentrifugation in protein science: a tutorial review. Protein Sci 11(9):2067-79.

Lebowitz J, Teale M, Schuck P. 1998. Analytical band centrifugation of proteins and protein complexes. Biochem. Soc. Transact. 26:745-749.

Lewis MS, Shrager RI, Kim S-J. 1993. Analysis of protein-nucleic acid and protein-protein interactions using multi-wavelength scans from the XL-A analytical ultracentrifuge. In: Schuster TM, Laue TM, editors. Modern Analytical Ultracentrifugation. Boston: Birkhäuser. p 94-115.

MacRaild CA, Hatters DM, Lawrence LJ, Howlett GJ. 2003. Sedimentation velocity analysis of flexible macromolecules: self-association and tangling of amyloid fibrils. Biophys J 84(4):2562-9.

Philo JS. 1997a. An improved function for fitting sedimentation velocity data for low molecular weight solutes. Biophys. J. 72:435-444.

Philo JS. Probing receptor-ligand interactions by sedimentation equilibrium; 1997b; San Jose, CA. SPIE. p 170-177.

Philo JS. 2000a. A method for directly fitting the time derivative of sedimentation velocity data and an alternative algorithm for calculating sedimentation coefficient distribution functions. Anal. Biochem. 279:151-163.

Philo JS. 2000b. Sedimentation equilibrium analysis of mixed associations using numerical constraints to impose mass or signal conservation. Methods in Enzymology 321:100-120.

Richards EG, Schachman HK. 1959. Ultracentrifuge studies with Rayleigh interference optics. I. General applications. J. Phys. Chem. 63:1578-1591.

Rivas G, Fernandez JA, Minton AP. 1999a. Direct observation of the self-association of dilute proteins in the presence of inert macromolecules at high concentration via tracer sedimentation equilibrium: theory, experiment, and biological significance. Biochemistry 38(29):9379-88.

Rivas G, Stafford W, Minton AP. 1999b. Characterization of heterologous protein-protein interactions via analytical ultracentrifugation. Methods: A Companion to Methods in Enzymology 19:194-212.

Rowe AJ. 1992. The Concentration Dependence of Sedimentation. In: Harding SE, Rowe AJ, Horton JC, editors. Analytical Ultracentrifugation in Biochemistry and Polymer Science. Cambridge: Royal Society of Chemistry. p 394-406.

Schachman HK. 1959. Ultracentrifugation in Biochemistry. New York: Academic Press.

Schachman HK. 1992. Is There a Future for the Ultracentrifuge? In: Harding SE, Rowe AJ, Horton JC, editors. Analytical Ultracentrifugation in Biochemistry and Polymer Science. Cambridge: Royal Society of Chemistry. p 3-15.

Schuck P. 1994. Simultaneous radial and wavelength analysis with the Optima XL-A analytical ultracentrifuge. Progr. Colloid. Polym. Sci. 94:1-13.

Schuck P. 1998. Sedimentation analysis of noninteracting and self-associating solutes using numerical solutions to the Lamm equation. Biophys. J. 75:1503-1512.

Schuck P. 2000. Size distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. Biophys. J. 78:1606-1619.

Schuck P. 2003. On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation. Anal. Biochem. 320:104-124.

Schuck P, Braswell EH. 2000. Measurement of protein interactions by equilibrium ultracentrifugation. In: Coligan JE, Kruisbeek AM, Margulies DH, Shevach EM, Strober W, editors. Current Protocols in Immunology. New York: John Wiley & Sons. p 18.8.1-18.8.22.

Schuck P, Demeler B. 1999. Direct sedimentation analysis of interference optical data in analytical ultracentrifugation. Biophys. J. 76:2288-2296.

Schuck P, MacPhee CE, Howlett GJ. 1998. Determination of sedimentation coefficients for small peptides. Biophys. J. 74:466-474.

Schuck P, Millar DB. 1998. Rapid determination of molar mass in modified Archibald experiments using direct fitting of the Lamm equation. Anal. Biochem. 259:48-53.

Schuck P, Perugini MA, Gonzales NR, Howlett GJ, Schubert D. 2002. Size-distribution analysis of proteins by analytical ultracentrifugation: strategies and application to model systems. Biophys J 82(2):1096-1111.

Schuck P, Radu CG, Ward ES. 1999. Sedimentation equilibrium analysis of recombinant mouse FcRn with murine IgG1 and Fc fragment. Mol. Immunol. 36:1117-1125.

Schuck P, Rossmanith P. 2000. Determination of the sedimentation coefficient distribution by least-squares boundary modeling. Biopolymers 54:328-341.

Servillo L, Brewer HB, Osborne JC. 1981. Evaluation of the mixed interaction between apolipoproteins A-II and C-I by equilibrium sedimentation. Biophys. Chem. 13:29-38.

Solovyova A, Schuck P, Costenaro L, Ebel C. 2001. Non-ideality by sedimentation velocity of halophilic malate dehydrogenase in complex solvents. Biophysical Journal 81(4):1868-80.

Stafford WF. 1992. Boundary analysis in sedimentation transport experiments: a procedure for obtaining sedimentation coefficient distributions using the time derivative of the concentration profile. Anal. Biochem. 203:295-301.

Stafford WF, Sherwood PJ. 2004. Analysis of heterologous interacting systems by sedimentation velocicty: curve fitting algorithms for estimation of sedimentation coefficients, equilibrium and kinetic constants. Biophys Chem 108:231-243.

Straume M, Johnson ML. 1992. Analysis of residuals: criteria for determining goodness-of-fit. Methods Enzymol. 210:87-105.

Svedberg T, Pedersen KO. 1940. The ultracentrifuge. London: Oxford University Press.

Tanford C. 1961. Physical chemistry of macromolecules. New York: Wiley.

van Holde KE, Weischet WO. 1978. Boundary analysis of sedimentation velocity experiments with monodisperse and paucidisperse solutes. Biopolymers 17:1387-1403.

Vistica J, Dam J, Balbo A, Yikilmaz E, Mariuzza RA, Rouault TA, Schuck P. 2004. Sedimentation equilibrium analysis of protein interactions with global implicit mass conservation constraints and systematic noise decomposition. Anal. Biochem. 326:234-256.

Table 1

	absorbance optics (ABS)	**interference optics (IF)**
selectivity:	• selective detection (e.g., in the presence of non-absorbing components)	• not selective: sensitive to all solution components (including buffer salts)
linearity and concentration range:	• linear to ~ 1.5 OD, a large concentration range may be achieved by the use of multiple wavelengths	• unlimited linearity, 10^4-fold concentration range
buffer considerations	• buffer cannot contain large amounts of DTT, TRIS, HEPES, other absorbing components for use in far UV (e.g. 230 nm)	• advantageous in the presence of strongly absorbing components (e.g., nucleotides, nucleic acids), **but** requires an exact chemical match of reference buffer volume and composition (through dialysis or gel filtration)
baselines	• small time-invariant (TI) radial baseline profile[9]	• generates significant time-invariant radial-dependent (TI) and radial-invariant time-dependent (RI) baselines, unproblematic in SV, but not trivial in SE
maximum signal/noise ratio	• ~ 300	• > 3000
data acquisition	• ~ minutes/scan, may be limiting rotor speed in SV, depends on scanning mode	• few seconds/scan
windows	quartz windows	• sapphire windows
conditions for velocity sedimentation (SV) • high speed, single speed typical sample requirements: • stability for 3 hours [3] • several cells with a range of loading concentrations; for example, stock solution with serial dilutions [4]	• volume 400 microliters (as low as 150 microliters) • rotor speed high: 40 – 60,000 rpm[1] • optimal loading absorbance: 0.5 – 1.3 OD • typical minimal desirable loading absorbance ~ 0.05 OD[5] • requires thorough temperature equilibration • controlled start from 0 rpm[6] • constant baseline usually with small radial-dependent features[7] • scan settings for fast scans (continuous mode, 0.003 cm radial increment)	• volume 400 microliters (as low as 150 microliters), sample/reference precisely matched • rotor speed usually 50 – 60,000 rpm[1] • optimal loading concentration: > 0.1 mg/ml (> 0.3 fringes) • typical minimal desirable loading concentration: ~ 0.05 mg/ml[5] • requires thorough temperature equilibration • controlled start from 0 rpm, may need pre-adjustment of optics[6] • generates radial baseline profile and radial-invariant offsets in each scan, which can be computationally eliminated after modeling[8]
conditions for equilibrium sedimentation (SE) • low speed, multiple speeds typical sample requirements: • stability for 2 – 5 days [3] • use gel-filtration to remove small Mw contaminants • several cells with a range of loading concentrations; for example, stock solution with serial dilutions [4]	• volume 180 microliters sample and 190 microliters reference (150 microliters sample for Mw > 100 kDa) • two or three rotor speeds, lowest at c(b)/c(m) ~ 3, highest generating meniscus depletion c(m) ~ 0 [2] • optimal loading absorbance: 0.2 to 0.5 OD[4] • typically scan at multiple wavelengths: 280 nm, 230 nm, 250 nm • usually no prior temperature equilibration required • constant baseline usually with small radial-dependent features[9] • scan settings for slow, precise scans (step mode, 0.001 cm radial increment)	• volume 180 microliters, sample/reference precisely matched (150 microliters sample for Mw > 100 kDa) • two or three rotor speeds, lowest at c(b)/c(m) ~ 3, highest meniscus depletion c(m) ~ 0; can tolerate steeper gradients leading to higher sample concentration[2] • optimal loading concentration: > 0.1 mg/ml (> 0.3 fringes) • require 'aging' of cell assemblies, water blanks • usually no prior temperature equilibration required • radial baseline profile and radial-invariant offsets, requires water blanks or TI noise elimination from global analysis of different rotor speeds[10]

(1) Choice of rotor speed: generally as fast as possible but dependent on protein size and optical system; the acquisition of at least 5 – 10 scans during the complete sedimentation process is desirable in SV; for molar mass determination slightly lower rotor speeds may be desirable (2) The ratio of concentration at the bottom relative to the meniscus, c(b)/c(m), can be theoretically predicted by simulating the approach to equilibrium with SEDFIT. This also provides a lower limit for the time to attain equilibrium and allows assessing the concentration profiles and gradients in equilibrium; (3) Stability may depend on temperature – SV and SE can be run at 4 °C; sedimentation equilibrium can be shortened by reducing column volume. [4] Concentration choice will depend on the purpose of the experiment [5] Lower values are possible, but with deteriorating level of detail due to limiting signal/noise ratio. [6] Controlled start from 0 rpm excludes the use of a low-speed (typically 3,000 rpm) phase for adjustment of optical and scan settings or temperature equilibration prior to high-speed acceleration. [7] Ideally exhibits a constant flat baseline, but ordinarily shows some time-invariant features from imperfections in the windows, which can be computationally eliminated after data analysis [8] Computational elimination is usually unproblematic in conjunction with modeling the time-course of sedimentation. [9] Baseline may shift at different wavelengths or when using buffer components with unstable absorbance, such as DTT (may by substituted by TCEP). Radial-dependent features may be eliminated computationally in the global analysis of equilibrium at a sufficient range of rotor speeds. [10] Computational treatment of TI noise in sedimentation equilibrium depends on the use of a sufficiently large range of rotor speeds, but may be improved by global multi-signal analysis in conjunction with absorbance data.

Table 2

	detection	details and recommendation
ionic strength you need sufficient ions to screen protein charges and prevent long-range electrostatic interactions from affecting protein sedimentation	both IF and ABS	▪ *always use > 10 mM NaCl or other salt*
absorption for ABS detection, the absorption of the buffer at the detection wavelength should not exceed 0.2 OD ▪ *select appropriate detection system* ▪ *if in doubt, measure the absorption spectrum of the buffer*	ABS	▪ *always use > 10 mM NaCl or other salt* ▪ HEPES and TRIS buffers, as well as EDTA and EGTA, absorb in the far UV. At low concentrations they can be tolerable for 280 nm ABS detection (e.g., 10 mM TRIS), but this may not permit the 230 nm detection. ▪ β-mercaptoethanol or DTT at low mM concentrations are compatible with IF and ABS detection at 280 nm, but they will generate an absorbance signal which may change with time. (For the ABS detection system, this requires a radial-invariant baseline ('RI-noise') to be considered in the data analysis) ▪ the presence of nucleotides at > 50 μM usually prohibits the use of the ABS system
refractive index the IF system detects the sedimentation of buffer salts (and any other buffer component) and will be sensitive to even very small mismatches in the concentration between the sample and reference buffer	IF	▪ *use size-exclusion chromatography or equilibrium dialysis to change buffer, if necessary* ▪ high concentrations > 1M of buffer components with large refractive indexsignal, e.g. guanidine hydrochloride, CsCl, glycerol, and others, are very difficult to match optically between sample and reference. In this case, either use the ABS system, or use pure H_2O as a reference buffer and explicitly treat the buffer as a sedimenting component in the data analysis ▪ if detergents are required, if possible, non-absorbing detergents in conjunction with the ABS system is usually advantageous over the IF system
density buffer components that raise the density of the solution may create density gradients at high rotor speed	both	▪ *glycerol or sucrose should be absent, if possible* ▪ solutions of higher density decrease the sedimentation velocity and increase the time required to attain SE ▪ self-forming density gradients may be a concern for mixed solvents and high concentrations of buffer components (such as CsCl or sucrose)
viscosity buffer components that lead to increased viscosity extend the experimental time	both	▪ if glycerol cannot be avoided, multiply the time-intervals for establishing SE of 6 h with the relative viscosity of the solution. ▪ for SV experiments, use larger time intervals between scans and let the experiment continue until the protein is depleted from the solution column.
preferential hydration when using buffer components that significantly increase the solvent density, preferential binding or exclusion of water from the protein solvation shell can lead to changes in the protein buoyancy and in its sedimentation behavior	both	▪ this is usually not a concern for buffers with density close to water (1.02 g/ml), for which the density between the hydration shell and the buffer is nearly matched ▪ the effect of buffer components strongly interacting with the protein, such as chaotropic agents or detergents, needs to be considered with regard to the altered partial-specific volume of the protein (see Appendix 2). ▪ the characterization of heterogeneous interactions between proteins is usually insensitive of hydration effects, as the effective buoyant molar mass the solvated protein can be determined by sedimentation for each protein separately (see Appendix 2).

Standard phosphate buffered saline works well with regard to all necessary considerations. For differences between IF and ABS detection, see also Table 1. Information on chemical compatibility with solvents can be found at http://camis.sr.unh.edu/AUC/cell.html.

Table 3:

problem under study:	sedimentation equilibrium	sedimentation velocity
determine the molar mass of a tight protein complex (assume known partial-specific volume)	single gradient possible, but desirable are 2 – 3 rotor speeds, 2 – 3 concentrations; Model with a single exponential to extract Mw **advantage:** direct measurement, usually reliable estimate, error < 5% **disadvantage:** resolves contaminants only poorly, experiment takes up to a few days	single SV experiment is possible, but desirable is a dilution series of concentrations; Modeling with the Lamm equation is in theory possible but usually very sensitive to heterogeneity and only gives a lower limit of M; c(M) can be advantageous; optimal is a hybrid discrete continuous model describing impurities with continuous sections, floating Mw of main discrete component **advantage:** relatively tolerant of impurities outside the size-range of interest (they will be resolved), takes several hours **disadvantage:** frequently lower precision (~ 10 %)
determine the oligomeric state of a membrane protein in detergent solution	density compensation, or Edelstein-Schachman technique {Edelstein, 1967 #186}	
determine the purity of the sample, detect protein aggregates	results can be highly variable dependent on the nature of the sample	long-column, high-speed SV with c(s) analysis is the method of choice, very sensitive for detection of higher oligomers, stable complexes
determine the number of species	(variable resolution)	long-column SV with c(s) gives better resolution
presence of self-association	dilution series, powerful negative control: can all data be fit with a single species model?	dilution series, test: are c(s) peak position concentration dependent?
kinetics of association	no information possible	diagnostics: are c(s) peak positions concentration dependent, or shift only the peak heights with concentration? global modeling of sedimentation boundaries
determine the stoichiometry of a weak protein complex	dilution series, model SE globally with different stoichiometries, compare	dilution series With c(s) analysis, analyze global model isotherm of $s_w(c)$ with models of different stoichiometries **advantage:** may work better than SE when complexes cannot be populated well, shape of c(s) may give hint of complexes formed. Tolerant of some impurities. Alternative: global modeling of sedimentation boundaries; **advantage:** boundary shape can exhibit a characteristic shape for higher-order oligomerization, but is very dependent on association kinetics
determine the association constant of self-association	dilution series, model SE globally **advantage:** very direct, use prior molar mass information complication: need good estimate of partial-specific volume, in particular for weak self-association	dilution series with c(s) analysis, global model of isotherms sw(c) **disadvantage:** need to span a very large range of concentrations, since s(1) and s(n) are not known a priori (in contrast to M(1) and M(n) for a given association scheme)
determine association constant of heterogeneous and mixed protein interaction	sediment and completely characterize the sedimentation behavior of both components separately, then use dilution series of mixture, model SE globally **advantage:** very direct, use prior molar mass information; no need to know partial-specific volume	sediment and completely characterize the sedimentation behavior of both components separately, then use dilution series of mixture With c(s) analysis, global model of isotherm of sw(c) Alternative: global modeling of sedimentation boundaries; **disadvantage:** dependent on association kinetics
hydrodynamic shapes of complexes, ligand-induced conformational change	no information possible	populate complex near saturation, best in concentration series to verify limiting s-value of peak in c(s); followed by hydrodynamic modeling.

Appendix 1: Mathematical formulas in the theory of sedimentation and the modeling of data

Sedimentation Velocity

The evolution with time of the concentration distribution of an ideally sedimenting single protein species $\chi(r,t)$ is described by the Lamm equation (Lamm 1929)

$$\frac{\partial c}{\partial t} = \frac{1}{r}\frac{\partial}{\partial r}\left[rD\frac{\partial c}{\partial r} - sw^2 r^2 c \right] \tag{A1}$$

where r denotes the distance from the center of rotation, t denotes the time, ω the angular velocity of the rotor, and s and D the sedimentation and diffusion coefficient, respectively.
Although experimental data could be modeled with Eq. A1 to derive s, D and therefore M according to the Svedberg equation

$$\frac{s}{D} = \frac{M\left(1-\bar{v}r\right)}{RT} \tag{A2}$$

(with R the gas constant, T the temperature, $\bar{v}$ the protein partial-specific volume and $\rho>$ the solvent density), in practice this is frequently not a successful approach, because of impurities and sample heterogeneity leading to excess broadening of the sedimentation boundary. Unrecognized excess broadening of the boundary would lead to an overestimation of D, and thus an underestimation of M (Dam and Schuck 2004). In practice, modeling data with the Lamm equation can only give a lower limit for the molar mass.

A more general approach which takes into account possible heterogeneity of the sample is to model the sedimentation data as a differential sedimentation coefficient distribution $c(s)$

$$a(r,t) \cong \int_{s_{\min}}^{s_{\max}} c(s)\, c_1(s,F,r,t)\, ds \tag{A3}$$

where $\chi_1(s,F,r,t)$ denotes the sedimentation profiles of a species at unit concentration, initially uniformly loaded, with a sedimentation coefficient s and a hydrodynamic frictional ratio $F = (f/f_0)$, from which the diffusion coefficient is determined according to the hydrodynamic relationship

$$D(s) = \frac{\sqrt{2}}{18p} kT\, s^{-1/2} \left(hF\right)^{-3/2} \left(\left(1-\bar{v}r\right)/\bar{v}\right)^{-1/2} \tag{A4}$$

(with $\eta>$ the solvent viscosity) (Schuck 2000). This method can deconvoluted the effects of diffusion, and distinguish sedimenting species that migrate at s-values causing separation less than the diffusional broadening of the boundary. In practice, the value F will be determined by non-linear regression during the data analysis, such that F represents the weight-average frictional ratio of the samples under study.

It is possible to convert the sedimentation coefficient distribution $c(s)$ into a molar mass distribution $c(M)$, because for each value of s, with the given value of F, the diffusion coefficient D is determined, and thus the Svedberg equation A2 gives the molar mass M. This analysis $c(M)$ can only be interpreted with caution. Although the detailed shape of $c(s)$ is largely independent on the particular value of F, and the assumption that all sedimenting species can be described with one single value of F has little impact on $c(s)$, this is not true with $c(M)$. For example, different values of F will leave the peak positions of $c(s)$ invariant, but they will lead to different molar mass values

in *c*(*M*). Nevertheless, if the *c*(*s*) distribution shows a single major peak, the transformation to *c*(*M*) is usually a very good approximation for the molar mass (frequently within 10% or better). This precision can be sufficient to assign the oligomeric state.

More details of the c(s) analysis can be found in (Dam and Schuck 2004). A tutorial on size-distribution analysis can be found on the internet at www.analyticalultracentrifugation.com/sizedistributions.htm.

For interacting proteins, a more general form of the Lamm equation is available that takes explicitly into account the chemical reaction during the sedimentation process (Fujita 1975). For a 1:1 hetero-association, this can take the form

$$\begin{aligned} \frac{\partial c_i}{\partial t} &= \frac{1}{r}\frac{\partial}{\partial r}\left[rD_i\frac{\partial c_i}{\partial r} - s_i w^2 r^2 c_i\right] + s_i q \qquad i = 1,2,3 \\ q &= \frac{k_{off}}{K_d}c_1 c_2 - k_{off}c_3 \qquad\qquad s_1 = s_2 = -s_3 = -1 \end{aligned} \tag{A5}$$

with the indices 1 and 2 describing the free and 3 the bound species, respectively. q represents the chemical reaction flux, characterized with the equilibrium dissociation constant K_D, and the chemical off-rate constant k_{off}. Modeling of data with Eq. A5 can be performed with SEDPHAT, as described in more detail at the website www.analyticalultracentrifugation.com/sedphat/sedimentation_velocity_for_reactive_systems.htm

Sedimentation Equilibrium

The distribution of proteins in sedimentation equilibrium can be derived from equilibrium thermodynamics. It can be shown that proteins in sedimentation equilibrium assume Boltzmann distributions (Svedberg and Pedersen 1940)

$$a_l(r) = c(b)e_l d\exp\left[M(1-\bar{\nu}r)\frac{w^2}{2RT}(r^2-b^2)\right] \tag{A6}$$

where $a_\lambda(r)$ is the signal measured at radius r, ε_λ the molar extinction coefficient, d the optical pathlength (usually 1.2 cm), and c(b) is the concentration at the bottom of the sample column. The curvature of this exponential shape is governed by the square of the rotor speed, and the term $M_b = M(1-\bar{\nu}r)$, also referred to as the buoyant molar mass.

For a protein in a reversible monomer-dimer self-association equilibrium, mass action law $c_2 = c_1^2 K_{12}$ is fulfilled simultaneously in all positions within the solution column, leading to the distribution

$$\begin{aligned} a_l(r,t) = {} & c_1(b)e_l d\exp\left[M_b\frac{w^2}{2RT}(r^2-b^2)\right] \\ & + K_{12}c_1(b)^2 2e_l d\exp\left[2M_b\frac{w^2}{2RT}(r^2-b^2)\right] \end{aligned} \tag{A7}$$

here, M_b denotes the monomer buoyant molar mass $M(1-\bar{\nu}r)$, and K_{12} the equilibrium association constant for dimer formation. Similarly, heterogeneous interactions are characterized by local mass action law, which leads to the form

$$a_l(r,t) = c_A(b)e_{l,A}d\exp\left[M_{b,A}\frac{w^2}{2RT}(r^2-b^2)\right] + c_B(b)e_{l,B}d\exp\left[M_{b,B}\frac{w^2}{2RT}(r^2-b^2)\right] + K_a c_A(b)c_B(b)\left(e_{l,A}+e_{l,B}\right)d\exp\left[\left(M_{b,A}+M_{b,B}\right)\frac{w^2}{2RT}(r^2-b^2)\right] \quad \text{(A8)}$$

where K_a denotes the equilibrium association constant.

The equilibrium expressions Eq. A6, A7, and A8, as well as others considering two-site binding and mixed self- and hetero-association are available in the software SEDPHAT for global modeling of experimental sedimentation equilibrium data.

Isotherm Analysis

For the analysis of protein interactions, isotherms are useful tools that relate the concentration dependence of a measured average quantity to those of the individual protein species and the concentration dependent shift in their population. Isotherms of weight-average sedimentation coefficients have particular importance, because the weight-average sedimentation coefficient can be rigorously determined without reference to a model for a particular reaction scheme (Schuck 2003).

The isotherm of weight-average sedimentation coefficients takes the general form

$$s_w(c^{tot}) = \frac{1}{c_{tot}}\sum_i c_i s_i \quad \text{(A9)}$$

(with s_w denoting the weight-average s-value, c^{tot} the total protein concentration, i and s_i denoting all molecular species and their sedimentation coefficient, respectively). For the example of a heterogeneous interaction of proteins A and B forming a reversible 1:1 complex, the isotherm takes the following form: For a series of loading concentrations c_A^{tot} and c_B^{tot}, the weight-average sedimentation coefficient detected at a signal λ is

$$s_w^l(c_A^{tot},c_B^{tot}) = \frac{e_A c_A^{free} s_A + e_B c_B^{free} s_B + (e_A+e_B)Kc_A^{free}c_B^{free}s_{AB}}{e_A c_A^{tot} + e_B c_B^{tot}}$$
$$c_A^{tot} = c_A^{free} + Kc_A^{free}c_B^{free} \quad \text{(A10)}$$
$$c_B^{tot} = c_B^{free} + Kc_A^{free}c_B^{free}$$

(with the equilibrium association constant K, and the extinction coefficients ε). This equation implies mass action law and mass conservation. This can be used for the study of interacting systems in the following way: In a series of experiments at different total loading concentrations of A and B, the weight-average sedimentation coefficient $s_w^l(c_A^{tot},c_B^{tot})$ is determined. This data can then be modeled with Eq. A10. This and other isotherm models are implemented in SEDPHAT.

References:

Dam J, Schuck P. 2004. Calculating sedimentation coefficient distributions by direct modeling of sedimentation velocity profiles. Methods Enzymol 384:185-212.

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

Schuck P. 2000. Size distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. Biophys. J. 78:1606-1619.

Svedberg T, Pedersen KO. 1940. The ultracentrifuge. London: Oxford University Press.

Appendix 2: Protein Partial-Specific Volumes for Sedimentation Analysis

Sedimentation is measured in solution and according the principle of Archimedes, will be affected by buoyancy, i.e., the mass of the solvent displaced by the sedimenting particle. The sedimenting particle in this sense comprises the protein with the solvation shell. The net sedimenting mass is referred to as the buoyant molar mass, M_b.

Thermodynamically, this effect can be described most rigorously by the density increment $d\rho/dc$ (Eisenberg 1976) (with ρ the solution density and c the protein concentration): $M_b = M\, d\rho/dc$. Frequently, this is replaced by analogy to the physical picture of a sedimenting particle with an effective partial-specific volume ϕ', and M_b is written as $M(1-\phi'\rho)$. Ignoring the effect of protein solvation on the sedimentation, this can be simplified to $M(1-\bar{v}\rho)$, where $\bar{v}$ is the partial-specific volume of the protein. The value of $\bar{v}$ can have significant influence on the sedimentation behavior. Typical values for proteins are 0.70 – 0.75 ml/g, and accordingly, M_b is usually 25 – 30% of the true molar mass. Any inaccuracies in the predetermined $\bar{v}$ will be amplified in the buoyancy term $(1-\bar{v}r)$ by a factor 3 – 4.

The sedimentation analysis can proceed in two possible ways: Either the partial-specific volume is determined independently (from prediction or measurement, see below) and the molar mass is measured in the sedimentation experiment, or the molar mass is known (from the amino acid sequence or from mass spectroscopy) and the partial-specific volume can be rigorously determined from the sedimentation experiment. Unfortunately, the latter approach can be difficult in the presence of self-association. Frequently a mixed approach is used, in which $\bar{v}$ is determined from sedimentation based on the known molar mass, but compared with the predicted $\bar{v}$. The experimental and theoretical values can be regarded consistent if they are within the typical error of 1%. If they differ more, possible reasons can be self-association of the protein, or significant effects of preferential hydration (see below). For this reason, the independent prediction or measurement of $\bar{v}$ can be very valuable.

For most proteins that do not contain glycosylation (e.g. if they are expressed in *E. coli*), the prediction of $\bar{v}$ can be based on the composition of amino acids. This calculation can be performed with the software SEDNTERP. When working in dilute buffers with densities < 1.02, such as PBS, effects of preferential hydration are minimal and can be safely ignored, as the hydration shell has nearly the same density of the bulk solvent, and is therefore gravitationally neutral (Lebowitz *et al.* 2003). This may no longer be true if any buffer component is present at a very high concentration, or if any buffer component binds significantly to the protein. This can be the case, for example, in the presence of a few percent glycerol, sucrose, and molar quantities of salt or chaotropic agents. It can also be true if the protein is highly charged, so that a significant number of counter-ions are bound to the protein (Ebel *et al.* 2000). At high concentrations of co-solutes, it should be noted that the sedimentation of the co-solute may generate a significant dynamic density gradient affecting the macromolecular sedimentation (Schuck 2004).

Large deviations can also occur if the protein contains prosthetic groups, glycosylation, or is embedded in a detergent micelle. The $\bar{v}$ contributions of prosthetic groups may be calculated by Traube's rule (Hoiland 1986). An approach to account for glycosylation is the determination of the carbohydrate mass by mass spectrometry (assuming a known polypeptide mass), and the estimation of the average carbohydrate partial-specific volume, $\bar{v}_c$, from composition (Shire 1992; Fairman *et al.* 1999; Lewis and Junghans 2000). For example, the most frequent glycosylation types in Chinese hamster ovary cells for HIV gp120 has been found to be of the high-mannose and complex type, which have $\bar{v}_c$ values of ~ 0.63 ml/g (Leonard *et al.* 1990). If the glycosylated protein mass

measured by mass spectroscopy is M_P and the amino acid mass is M_A, the carbohydrate mass follows as $M_C = M_P - M_A$, and the partial specific volume of the glycosylated protein is

$$\bar{v} = \frac{1}{M_P}\left(M_A\bar{v}_A + (M_P - M_A)\bar{v}_c\right) \qquad \text{(A11)}$$

The consideration of bound detergent is more complicated due to the frequently unknown ratio of protein/detergent in the solubilized protein (Tanford and Reynolds 1976; Schubert and Schuck 1991; Howlett 1992; le Maire *et al.* 2000; Noy *et al.* 2003). Depending on the properties of the protein and the detergent, a possible approach is density matching (increasing the solvent density to make the detergent neutrally buoyant). It should be noted that the latter can be complicated by H/D exchange if D_2O/H_2O mixtures are used, or by preferential solvation of the protein and detergent micelles if a high concentration of co-solute is used to raise the solvent density. Nevertheless, maximal estimated errors from preferential solvation can be sufficiently small to measure the protein oligomeric state (Center *et al.* 2001).

A method for the theoretically rigorous experimental determination of the density increment $d\rho/dc$ is densitometry on solutions of proteins in dialysis equilibrium with the solvent. Unfortunately, to reach sufficient precision in $\bar{v}$, this method requires 5 – 20 mg of protein (Durchschlag 1986; Lebowitz *et al.* 2003). An alternative approach for particles with densities close to that of water are experiments in solvents of different densities, for example, in H_2O and D_2O, the global analysis of which permit calculating $\bar{v}$ of the macromolecule (Edelstein and Schachman 1967). For typical unmodified proteins, however, this can lead to relatively large errors because only a small relative density variation can be experimentally achieved with D_2O.

In this context, it is very useful to consider the solution properties of the protein and the nature of the interaction to be studied. As mentioned above, for non-self-associating proteins consisting only of a polypeptide chain of known sequence molar mass, the buoyant molar mass obtained in the sedimentation equilibrium experiment itself provides an accurate measure of the protein $\bar{v}$. Such determination can be used in the analysis of the hetero-association. The $\bar{v}$ of the complexes can be calculated from the weight average of the components. This assumes that the volume change of the interaction is negligible for the purpose of sedimentation studies; this appears to be true in the overwhelming majority (although not all) of all protein interactions studied, and can be verified by the independence of the binding constants of pressure in experiments with different solution columns and rotor speeds.

For self-associating proteins, unless there are known conditions that abolish the association, the determination of $\bar{v}$ from the known sequence molar mass is more difficult, since the buoyant molar mass of the monomer is typically highly correlated with the binding constants of the interaction. While the assumption of the theoretical sequence-based value of $\bar{v}$ does not lead to very large errors for moderate to strong affinity constants, the error increases significantly for weak interactions. In this case, error bars should reflect the uncertainty of the sequence based prediction of the $\bar{v}$ value.

References:

Center RJ, Schuck P, Leapman RD, Arthur LO, Earl PL, Moss B, Lebowitz J. 2001. Oligomeric Structure of Virion-Associated and Soluble Forms of the Simian Immunodeficiency Virus Envelope Protein in the Pre-Fusion Activated Conformation. Proc. Natl. Acad. Sci. U S A 98:14877-14882.

Durchschlag H. 1986. Specific volumes of biological macromolecules and some other molecules of biological interest. In: Hinz H-J, editor. Thermodynamic data for biochemistry and biotechnology. Berlin: Springer. p 45-128.

Ebel C, Eisenberg H, Ghirlando R. 2000. Probing protein-sugar interactions. Biophys J 78(1):385-93.

Edelstein J, Schachman H. 1967. Simultaneous determination of partial specific volumes and molecular weights wtih microgram quantities. J. Biol. Chem. 242:306-311.

Eisenberg H. 1976. Biological macromolecules and polyelectrolytes in solution. Oxford: Clarendon Press.
Fairman R, Fenderson W, Hail ME, Wu Y, Shaw S-Y. 1999. Molecular weights of CTLA-4 and CD80 by sedimentation equilibrium ultracentrifugation. Anal. Biochem. 270:286-295.
Hoiland H. 1986. In: Hinz H-J, editor. Thermodynamic data for biochemistry and biotechnology. Berlin: Springer. p 17-44.
Howlett GJ. 1992. Sedimentation analysis of membrane proteins. In: Harding SE, Rowe AJ, Horton JC, editors. Analytical Ultracentrifugation in Biochemistry and Polymer Science. Cambridge, U.K.: The Royal Society of Chemistry. p 470-483.
le Maire M, Champeil P, Moller JV. 2000. Interaction of membrane proteins and lipids with solubilizing detergents. Biochim. Biophys. Acta 1508:86-111.
Lebowitz J, Lewis MS, Schuck P. 2003. Back to the future: A rebuttal to Henryk Eisenberg. Protein Sci 12:2649-2650.
Leonard CK, Spellman MW, Riddle L, Harris RJ, Thomas JN, Gregory TJ. 1990. Assignment of intrachain disulfide bonds and characterization of potential glycosylation sites of the type 1 recombinant human immunodeficiency virus envelope glycoprotein (gp120) expressed in Chinese hamster ovary cells. J Biol Chem 265:10373-10382.
Lewis MS, Junghans RP. 2000. Ultracentrifugal analysis of the molecular mass of glycoproteins of unknown or ill-defined carbohydrate composition. Methods in Enzymol. 321:136-149.
Noy D, Calhoun JR, Lear JD. 2003. Direct analysis of protein sedimentation equilibrium in detergent solutions without density matching. Anal. Biochem. 320:185-192.
Schubert D, Schuck P. 1991. Analytical ultracentrifugation as a tool for studying membrane proteins. Progr. Colloid Polym. Sci. 86:12-22.
Schuck P. 2004. A model for sedimentation in inhomogeneous media. I. Dynamic density gradients from sedimenting co-solutes. Biophys. Chem. 108:187-200.
Shire S. 1992. Determination of molecular weight of glycoproteins by analytical ultracentrifugation. Palo Alto, CA: Beckman Instruments.
Tanford C, Reynolds JA. 1976. Characterization of membrane proteins in detergent solutions. Biochim. Biophys. Acta 457:133-170.