

CHAPTER 1

A Brief Introduction to the Analytical Ultracentrifugation of Proteins for Beginners

DAVID J. SCOTT AND PETER SCHUCK

1 Introduction

Analytical ultracentrifugation (AUC) is one of the classical methods for the characterization of purified proteins in dilute solutions, and as such there is a large body of theoretical and practical studies in the literature going back eight decades, covering aspects including the technical implementation of AUC,¹ the theoretical foundation of ultracentrifugation in thermodynamics and physical chemistry of macromolecules,² and the mathematical analysis of ultracentrifugation experiments.³ Developed in the 1920s by The Svedberg,⁴ and redesigned into a commercial instrument by Edward Pickels,⁵ it became a central technique in the development of biochemistry and molecular biology. While the technique was in decline in the 1970s and 1980s, new instrumentation and numerical analysis in the 1990s stimulated renewed interest in AUC, in particular for the study of protein interactions.⁶

For the novice, however, the highly developed technical aspects in conjunction with the multitude of analytical approaches can be daunting. Therefore, this introduction to the book is unashamedly aimed at the novice who has come into contact with the methodology of the analytical ultracentrifuge and wishes to use the instrument for the characterization of globular protein samples in solution. Such a novice can be a molecular biologist, a protein crystallographer or NMR spectroscopist, or working as part of a high-through-put proteomics project, with interest in the characterization of the aggregation state, heterogeneity and thermodynamic characterization of reversible interactions of proteins. The aim of the present introduction is to provide help in where to start, or how to design and analyse a successful AUC experiment. Necessarily, this requires a very selective presentation, and for more in-depth information and descriptions of selected sedimentation velocity (SV) and sedimentation equilibrium (SE) methods, the reader is referred to recent reviews, monographs, practical protocols, and websites,^{1,7–25} and, of course, subsequent chapters in this book. Fortunately, with regard to the data interpretation, the mathematical

details of advanced analysis methods are largely encapsulated in software. However, in addition to general knowledge in modelling of data and non-linear regression, their use requires understanding of the basic concepts behind them.

2 What Can I Do with My Protein Sample?

At its most basic level, AUC simply consists of the application of a centrifugal force with the simultaneous real-time observation of the resulting redistribution of the macromolecule. Analysis can be performed from first principles, given quantitative and rigorous data on a particular sample. This requires no label or other chemical modification of the proteins; such as would occur in cross-linking or many fluorescence experiments; and no interaction with any matrix or surface, as would be required for gel filtration or surface plasmon resonance. As proteins are studied in solution, the experimentalist has direct access to their solution properties: a key strength of the technique. A central feature of sedimentation experiments for the study of protein interactions is that faster sedimenting complexes are transported through a solution of the slower sedimenting components. Consequently, reversibly formed complexes that dissociate can readily re-associate during the experiment, thus permitting the hydrodynamic and thermodynamic characterization of even weak and transient interactions.

Analytical ultracentrifugation of non-interacting proteins can reveal the molar mass, gross shape, and the heterogeneity of the sample. The latter includes the detection of even trace quantities of oligomers and aggregates, which can be of interest in biotechnology applications or aid in the interpretation of biosensor experiments. For interacting systems, protein complexes can be characterized with regard to their stoichiometry and the thermodynamic and kinetic constants of complex formation. Importantly, sedimentation techniques can distinguish between multiple coexisting complexes of different stoichiometries and also provide information on self-association properties, on mixed self- and hetero-association. The latter can be a crucial information for the biophysical study of protein interactions with other techniques, such as isothermal titration calorimetry. Hydrodynamic separation also yields information on the low-resolution structure of protein complexes and can enable the detection of conformational changes.

Although a large number of specialized centrifugation techniques have been developed for a variety of studies, such as analytical zone sedimentation,¹¹ difference sedimentation,²⁶ synthetic boundary measurements,²⁷ density gradient^{28,29} or fractionating^{30,31} and short-column techniques,³² the vast majority of ultracentrifugation experiments for the characterization of proteins are conducted by either conventional loading SV or long-column SE, which will be described in the following sections.

Ideally, both SV and SE techniques should be carried out on your sample, starting with SV to characterize the purity of the material. It is possible to characterise effectively a sample with SV using the latest analysis methods, however, both methods are highly complementary. The essential piece of information needed to plan an AUC experiment is to make sure at least three different concentrations of the sample are analysed, covering as wide a concentration range as possible. This is because concentration-dependent behaviour of the sample provides an information-rich data set that is highly effective in characterizing the solution properties of proteins, in particular,

their interactions. This applies equally for SV and for SE. For equilibrium experiments there is the additional requirement that data should be obtained, where possible, in a sequence of experiments at three different rotor speeds.

The quantity of material required is typically of the order of a few hundred micrograms. Owing to the concentration gradients established during centrifugation, a 10- to 1000-fold concentration range is typically observed in a single cell, and a size range of three decades in molar mass can be covered in a single experiment. Interacting components under study may have sizes ranging from peptides to very large multiprotein complexes. In general, affinities in the range of 10^4 – 10^8 M^{-1} can be determined, and kinetic dissociation rate constants of the order of $\sim 10^{-5}$ – 10^{-2} s^{-1} can be distinguished.

3 General Principles

Conceptually, the analytical ultracentrifuge can be thought of as a conventional preparative centrifuge that is equipped with an optical system for the observation of the protein distribution in real time during the centrifugation. The acquired data report on the spatial gradients that result from the application of the centrifugal field, and their evolution with time. Analytical rotors accept specialized assemblies for containing, typically, 100–400 μL of sample between windows that are optically transparent and perpendicular to the plane of rotation. The optical detection system is synchronized with the rotor revolution, such that data are acquired only while the sample assembly is in the light path (Figure 1).

The most commonly used optical detection systems are a dual-beam UV/VIS spectrophotometer equipped with monochromator (ABS) and a highly sensitive laser interferometer which records the refractive index gradients (IF). The ABS system requires typical loading concentrations between 0.1 and 1.5 OD, dependent on the

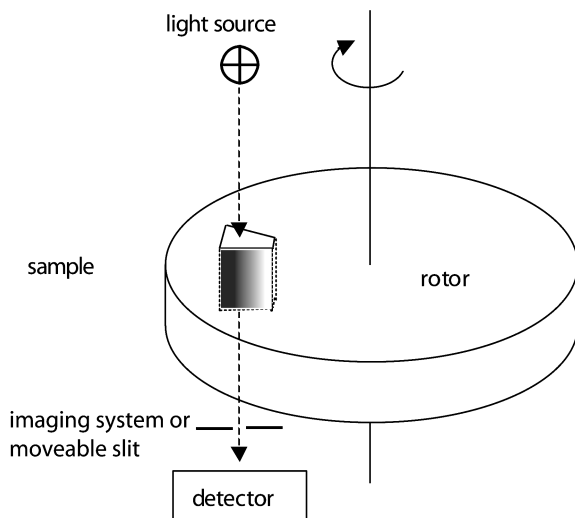


Figure 1 Schematics of the analytical ultracentrifugation detection

type of experiment, and has the advantage of being able to selectively detect the protein, for example at 280 nm (for the aromatic amino acids), 230 nm (peptide backbone), or for characteristic chromophores in the VIS, if present. The IF system is not selective, but completely linear in concentration and offers high signal-to-noise ratios and rapid data acquisition (which can be particularly valuable for observing the time course of sedimentation). Drawbacks of the IF system are the experimental requirement to have a precise match of the composition of the reference buffer, and systematic time-invariant signal offsets to be accounted for in the data analysis (see below). More practical information with regard to the selection of the optical system can be found in refs. 14, 16 and 33.

Two basic types of experiments are possible: (a) the application of a high centrifugal force and the analysis of the time course of the sedimentation process, termed sedimentation velocity (SV); and (b) the application of a low centrifugal force that permits the diffusion to balance the sedimentation such that a time-invariant equilibrium gradient can be observed, termed sedimentation equilibrium (SE). Both SV and SE approaches are uniquely suited for the study of protein interactions. 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 three factors – the gravitational force, the buoyancy and the hydrodynamic friction. The gravitational force is $F_{\text{sed}} = m\omega^2 r$ (with m the protein mass, ω^2 the rotor angular velocity, and r the distance from the centre of rotation). Since it is proportional to the square of the rotor speed, 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). The buoyancy force $F_{\text{b}} = -m\bar{v}\rho\omega^2$ (with \bar{v} the effective protein partial-specific volume and ρ the solvent density) opposes the sedimentation (following Archimedes principle) and is governed by the mass of the displaced solvent. Thus, protein partial-specific volumes are important, and the density effects of protein glycosylation, bound detergent and preferential hydration may be relevant considerations. Finally, the frictional force is governed by the hydrodynamic translational frictional coefficient as well as the migration velocity, and can be expressed as $F_{\text{f}} = s(kT/D)\omega^2$ (with k the Boltzmann constant, T the absolute temperature, and D the diffusion constant), where the sedimentation coefficient $s = v/\omega^2$ is a molecular constant (with v the absolute migration velocity). This permits the measurement of the low-resolution shape of the proteins and their complexes in terms of Stokes radii. A key parameter is the sedimentation coefficient s (measured in units of Svedberg, with $1 \text{ S} = 10^{-13} \text{ s}$). From the balance of these three forces, one can derive the Svedberg equation

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

(with M denoting the protein molar mass, and R denoting the gas constant),⁴ which provides 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).

4 Sedimentation Velocity

The principle behind SV is quite straightforward: at a high rotor speed, typically 40 000 rpm or above, the vast majority of proteins will sediment to the bottom of the cell. This deceptively simple process is highly information rich, as the sedimentation depends both on the size and shape of the protein. Hence a 50-kDa protein may sediment slower than a 40-kDa protein due to the heavier protein being highly elongated. If a protein self-associates with concentration then its apparent rate of sedimentation will change when either raising or lowering the loading concentration. SV will also give information on solution heterogeneity.

4.1 Setting Up a Simple Sedimentation Velocity Experiment

When first characterizing a sample by AUC, a typical amount of protein required would be 1 mL at 1 OD₂₈₀, assuming the use of the ABS detection system, or at 0.5–1 mg mL⁻¹, assuming the IF detection system. This will be enough for SV with three cells at three-fold dilutions, and have some sample left to perform an SE run. The three velocity cells should be loaded with 400 μL of sample: this will be a sample volume that fills *ca.* 85% of the length of the cell. Typically, concentrations can be at 1:1, 1:3 and 1:10 of the stock concentration. Generally, the concentration range should be as wide as possible, with the constraint not to exceed 1 mg mL⁻¹ (to prevent non-ideal sedimentation), and to remain well within the detection limits (*e.g.*, 0.05 mg mL⁻¹ < *c* for the IF system, and 0.05 OD < *c* < 1.5 OD for the ABS system). The reason for the concentration series is simple: for a self-associating system, lowering the concentration sufficiently will cause oligomers to dissociate into smaller species, which will sediment slower. However, if the protein is present in a variety of oligomers that do not interconvert, then dilution will have no effect whatsoever on the proportion of oligomers. Hence, a dilution series is essential to characterizing any protein in solution. Indeed, one SV experiment (of three dilutions) will very quickly tell the experimentalist in a matter of hours whether their protein sample undergoes a complex set of self-associations of exquisite biological necessity, or that the protein is an aggregated mess, and another sample preparation is needed.

The protein purity should be >95%, and it is recommended to perform size-exclusion chromatography as the last preparative step. With regard to the buffer requirements, it is useful to include at least 20 mM salts to suppress electrostatic interactions. Obviously, no solvent component should interfere with the optical detection, and reference buffer precisely matched in volume and composition is needed for experiments with the IF system. For most proteins under common experimental conditions, the effective partial-specific volume can be predicted with sufficient precision from the amino acid sequence,^{19,34} for example, using the software SEDNTERP, which also permits to calculate the solvent density and viscosity from tabulated data.³⁵ Greater care must be used with proteins containing non-amino acid components (*e.g.*, glycoproteins, proteins with prosthetic groups, detergent-solubilized proteins), or buffer conditions containing glycerol, sucrose or other components increasing the density and potentially leading to preferential solvation.^{23,34,36}

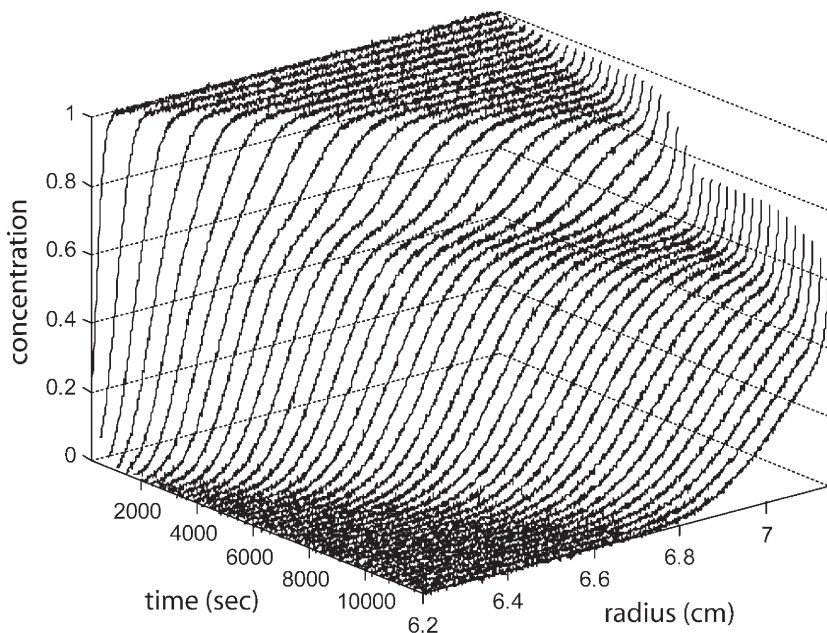


Figure 2 Sedimentation profiles of two species of 50 kDa, 4 S, and 150 kDa, 8 S, sedimenting at 50 000 rpm

In the absence of other considerations (and unless the protein complexes under study exceed several 100 kDa in size), the highest possible rotor speed (50 000 or 60 000 rpm, dependent on the rotor type used) is recommended for the highest resolution. A practical aspect crucial for the SV experiment is a thorough temperature equilibration to prevent convection in the solution column to occur. Very detailed practical instructions for planning and starting the SV run can be found in ref. 33.

Typical sedimentation data may look like as shown in Figure 2, showing initially the solution plateau, the evolution of the sedimentation boundaries of a non-interacting two-component mixture and the solvent plateau. Owing to the sector-shaped solution column, the solution plateau concentration decreases with time as the proteins move toward larger radial position with wider cross-sections. This amounts typically to 10–20% during the SV experiment.

To harness this powerful method of analysis, a suitable method of interpretation is needed. This is the concern of the next section.

4.2 Sedimentation Velocity Analysis

As indicated above, the major parameter derived from SV experiments is the sedimentation coefficient (s), which is the velocity per unit centrifugal force:

$$s = \frac{1}{\omega^2 r} \frac{dr}{dt} = \frac{m}{f} \quad (2)$$

where s is the sedimentation coefficient, $\omega^2 r$ the centrifugal field, r the radius, t time of sedimentation, m the buoyant molecular mass and f the frictional coefficient. Although the sedimentation coefficient (or s value) is a molecular constant, the experimentally observed sedimentation coefficients will vary with temperature and solution density, and are always converted into a standard state of 20 °C in water at infinite dilution (denoted $s_{20,w}^0$). This can be achieved, for example, using the software SEDNTERP, written by Dr. John Philo.³⁵ A comparison of the $s_{20,w}^0$ value with that predicted theoretically for a smooth compact sphere of the same mass and density gives the frictional ratio, f/f_0 , which reports on the shape of the molecule (by conversion into hydrodynamically equivalent ellipsoids with the software SEDNTERP, or by comparison with the theoretically predicted s value derived by hydrodynamic modelling from the crystal structure using the software HYDRO).³⁷ Because the $s_{20,w}^0$ value cannot exceed that of a smooth compact sphere of the same mass and density, the $s_{20,w}^0$ value can be used in some cases to deduce the oligomeric state of the protein.¹⁹

The simplest analytical method to derive the sedimentation coefficient from experimental data consists in reading the boundary displacement with time, and application of Equation (2). Although this illustrates conceptually the source of information, it does not provide the most precise estimate for the s value, and neglects a significant amount of information to be extracted from the sedimentation data. As can be seen in Figure 1, in addition to a displacement of the sedimentation boundary, diffusional broadening can be observed, which can be rich in information.

4.2.1 Modelling with Discrete Lamm Equation Solutions

In the other extreme, it is possible to base the analysis on the detailed sedimentation–diffusion equation, the Lamm equation³⁸

$$\left(\frac{\partial c}{\partial t}\right)_r = -\frac{1}{r} \left\{ \frac{\partial}{\partial r} \left[s\omega^2 r^2 c - D r \left(\frac{\partial c}{\partial r} \right) \right] \right\}_t \quad (3)$$

which describes the evolution of the complete concentration profiles during the SV experiment in the sector-shaped ultracentrifugal solution column. Several software packages allow fitting Equation (3) to experimental data, in order to determine both s and D (and consequently the molar mass M ; see Equation (1)) for one or several species. These include BPCFIT, LAMM, SEDANAL, SEDFIT, SEDPHAT, SVEDBERG and ULTRASCAN. Modern approximate analytical and numerical algorithms for solving Equation (3) are described in refs. 39–44, and adaptations to account for non-ideal solutions, solvent and pressure gradients have been developed.^{45–47} Although rigorous in theory, in practice, this approach is not easily applicable in many cases, due to the need to know in advance the number of protein species. This can be difficult given the exquisite sensitivity of SV to heterogeneity, including trace amounts of degradation products or impurity. Further, as illustrated in ref. 48, if species are present with slightly different s value (such that the difference in migration due to the different s values is less than the diffusional broadening), unrecognized heterogeneity will lead to an overestimation of D , and consequently an underestimate of the molar mass. Therefore, in practice, direct modelling with

Equation (3) will frequently give only a lower limit of the molar mass. If, on the other hand, Equation (3) can be fitted to the experimental data with residuals within the noise of the data acquisition with the correct molar mass, this provides a stringent demonstration of sample homogeneity.

Several strategies have been developed to determine sedimentation coefficient distributions, which can report on the heterogeneity of the sample, and in some cases may be used to derive molar mass information. Some are outlined in the following in chronological order, making increasing use of the computational power of desktop computers to apply more mathematically complex approaches embedded in the software packages. They also differ in the ability and strategy used to account for the systematic radial-dependent baseline profiles that occur in data acquired with the IF system.

4.2.2 The van Holde–Weischet Distribution

The integral sedimentation coefficient distribution $G(s)$ introduced in 1978 by van Holde and Weischet⁴⁹ basically exploits the observation that the displacement from diffusion proceeds proportionally to the square root of time, while the sedimentation is linear with time. Technically, this is accomplished by dividing the sedimentation boundaries into fractions (horizontal slices), reflecting the radii R_i of certain intervals of the fractional plateau concentration, followed by a transformation of the radius coordinate into apparent sedimentation coefficients s^* with

$$s^* = \frac{1}{\omega^2 t} \ln \frac{r}{r_m} \quad (4)$$

s^* simply reflects the velocity at which an ideal particle would have to migrate in order to travel from the meniscus to the position r within the time t . The s_i^* values from the corresponding boundary fractions of the different scans are then extrapolated to infinite time according to the formula

$$s_i^*(t) = s - \frac{2\sqrt{D}}{\omega^2 r_m} \Phi^{-1} \left(1 - \frac{2i}{N} \right) \times \frac{1}{\sqrt{t}} \quad (5)$$

with Φ^{-1} denoting the inverse error function, which is based on an Faxén-type approximate solution of the Lamm equation for a single species.^{3,49} A drawback of this method is that it requires the selection of the subset of experimental scans for the analysis, such that the scans to be analysed exhibit clear solution and solvent plateaus permitting the boundary division. For peptides and small proteins, this is not always possible. Also, it is difficult to apply to data from the IF detection exhibiting systematic radial-dependent baseline offsets.⁵⁰

The method can be found in the software ULTRASCAN^{8,51} and SEDFIT.⁵⁰ The application is illustrated in Figure 3. For mixtures of sedimenting species, the $G(s)$ distribution ideally shows piecewise vertical segments corresponding in length to the relative amount of species with certain s values. Because the underlying Equation (5) is valid for a single species only, $G(s)$ cannot resolve species that do not show clearly separating boundaries (*i.e.* if the difference in migration due to the different s values is less than the diffusional broadening).⁵⁰ In this case, the $G(s)$ transformations will exhibit

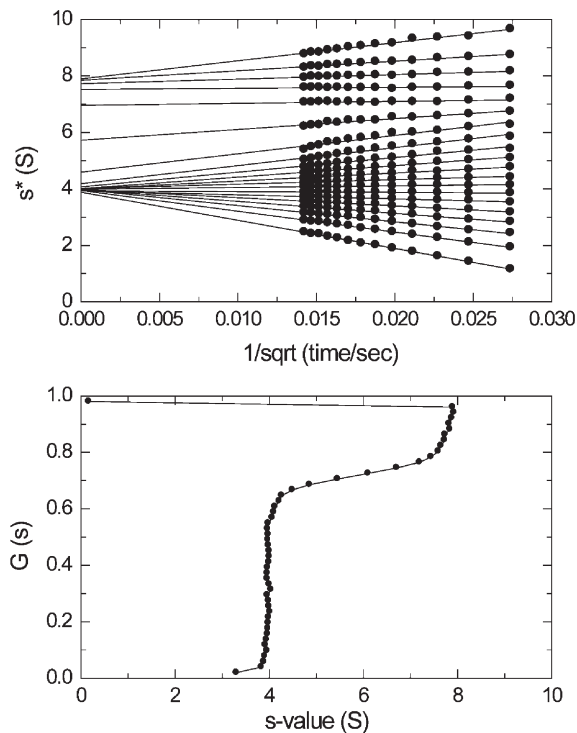


Figure 3 Van Holde–Weischet analysis of the data shown in Figure 1. Upper panel: the s_i^* values from each boundary (vertical columns of symbols), and the extrapolation of the corresponding boundary fractions in an inverse square root of time-scale (lines). Lower panel: the extrapolated s values for each fraction, which compose the integral distribution $G(s)$

diagonal lines with positive slopes within the range of s values of the sedimenting species. This highlights the particular strength of the method, which is the diagnosis if the sample is homogeneous, heterogeneous, or if it exhibits concentration-dependent repulsive interactions leading to non-ideal sedimentation. (The latter is characterized by a decreasing sedimentation coefficient with increasing concentrations.) In summary, for homogeneous samples, $G(s)$ will show a vertical line, for heterogeneous or associating systems it will be a diagonal line with positive slope, and for non-ideal sedimentation it will be a diagonal line with negative slope. Demeler discusses this methodology in greater depth in Chapter 11.

4.2.3 Apparent Sedimentation Coefficient Distributions $g^*(s)$

A different strategy to analyse experimental SV data is the differential sedimentation coefficient distribution $g^*(s)$. The asterisk on the g emphasizes that it is a distribution of apparent sedimentation coefficients, meaning the distribution of s values of hypothetical non-diffusing particles. By design, the area under the $g^*(s)$ distribution is equal to the amount of sedimenting material with that sedimentation coefficient.

Two different techniques to derive $g^*(s)$ are currently in use. The first is a data transformation, first introduced in 1992 by Stafford⁵² and further developed by Philo,⁵³ based on a time derivative dc/dt of the data and termed $g(s^*)$. The other is based on a direct least-squares fit of the data with an integral equation for the distribution, termed $ls-g^*(s)$, described in 2000 by Schuck and Rossmanith.⁵⁴

The dc/dt -based method was very important because it provided the first approach, short of subtracting experimentally measured water blanks, that was specifically adapted to account for the systematic radial-dependent baseline offsets of the IF data acquisition system. The time-dependent offsets are reduced by a vertical alignment procedure prior to the data analysis, and the radial-dependent offsets vanish by taking pair-wise differences of scans. The approximation is made that the pair wise difference $\Delta c/\Delta t$ is taken for the time derivative dc/dt . This is used, in turn, in the transformation

$$g(s^*) = \left(\frac{\partial c}{\partial t} \right) \frac{1}{c_0} \left(\frac{\omega^2 t^2}{\ln(r/r_m)} \right) \left(\frac{r}{r_m} \right)^2 \quad (6)$$

(with c_0 denoting the initial loading concentration) which is based on the equations for the sedimentation of non-diffusing species, and combined in an iterative procedure to account for radial dilution effects.⁵² Details on the computational approach can be found in ref. 55. The distribution from the dc/dt approach has been termed $g(s^*)$. A limitation of this approach is the validity of the approximation $\Delta c/\Delta t \sim dc/dt$, which can cause artificial broadening of the $g(s^*)$ curves and skewing at small s values,^{53,54} to an extent that is increasing with larger boundary displacement between scans and with larger time interval covered by the set of scans considered in the analysis. The dc/dt method can be found in the software DCDT+, and in this implementation, the maximum number of scans is recommended for which the errors remain within an acceptable tolerance. An approach to quantitatively account for effects of the approximation $\Delta c/\Delta t \sim dc/dt$ in the modelling of $g(s^*)$ curves is described in ref. 53. Figure 4 illustrates the result of this method.

The second approach to calculate the apparent sedimentation coefficient distribution is based on the definition of $g^*(s)$ as a distribution of non-diffusing species, each sedimenting according to a step-function $U(s, r, t)$

$$U(s, r, t) = e^{-2\omega^2 st} \begin{cases} 0 & \text{for } r < r_m e^{\omega^2 st} \\ 1 & \text{else} \end{cases} \quad (7)$$

which together produce a signal

$$a(r, t) \cong \int g^*(s) U(s, r, t) ds \quad (8)$$

Taking advantage of the increased computational power, it is possible to directly fit Equation (8) to the experimental data.⁵⁴ Adjustments accounting for solvent compressibility can be made in Equation (7).⁴⁶ Because this method is based on the least-squares fit of the data (hence the designation $ls-g^*(s)$), the usual goodness of fit criteria, such as r.m.s. deviation and residuals, are available. To avoid noise amplification in solving Equation (8), regularization is applied to calculate the simplest distribution consistent with the raw data.

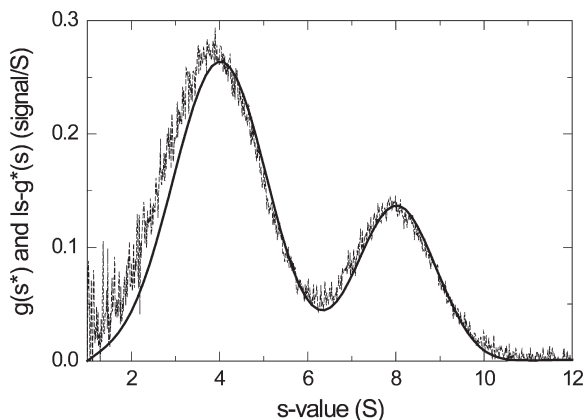


Figure 4 Application of $g(s^*)$ from the dc/dt method (thin dotted line), and the $ls-g^*(s)$ method (bold solid line) to the data shown in Figure 2

The systematic offsets of the IF systems are taken into account by explicitly calculating the best-fit estimates of the radial-dependent baseline profiles as well as the time-dependent offset from the data to be analysed, taking advantage of a new algebraic approach that can replace the alignment and differentiation procedure with minimal amplification of noise.⁵⁶ This can be used to remove the systematic noise contribution for inspection of the fit in the original data space. A reconstruction of the sedimentation model from the dc/dt -based $g(s^*)$ distribution has been described, which can allow to inspect the goodness-of-fit of the $g(s^*)$ representation.²⁰ An overview of computational details can be found in ref. 48. Although in theory the $ls-g^*(s)$ distribution is equivalent to the $g(s^*)$ distribution from dc/dt approach, because of the absence of the $\Delta c/\Delta t \sim dc/dt$ approximation, it does not have the stringent requirements for scan subset selection, and is therefore more widely applicable,⁵⁴ and more suitable, in particular, to data obtained at higher rotor speed or from the ABS optical detection system with the typical large boundary displacement between successive scans. This approach is implemented in SEDFIT, and results are also shown in Figure 4.

The limitation of its application is the underlying idea of an apparent sedimentation coefficient distribution, which is that the sedimentation–diffusion process under observation can be modelled by an apparent distribution of non-diffusing species, such that the leading and trailing parts of the diffusionally broadened boundary can be described by populations of species apparently sedimenting faster or slower, respectively. In the direct boundary model, the quality of the approximation of the sedimentation boundary as a superposition of non-diffusing species can be tested by inspecting the quality of fit. As in any data analysis problem, if the model does not adequately describe the data, no trust can be put in the parameters of the model, or in this case, the sedimentation coefficient distribution. For example, in the $ls-g^*(s)$ distribution, a failure of fitting the original data well may result in false peaks, and requires shortening the time interval of scans considered for analysis.

It can be shown in theory – at least for data that cover a narrow time interval – that the shape of the $g^*(s)$ distribution is approximately Gaussian, with a half-width

related to the diffusion coefficient of the protein. As a consequence, it is possible in a second stage of the SV analysis to fit the $g(s^*)$ distributions with Gaussians, and to use the Svedberg equation (Equation (1)) to derive the molar mass.⁵³ For several distinct non-interacting species it is possible to fit multiple Gaussian peaks where the number of peaks correspond to the number of species. Although this technique has the advantage of not being biased by sedimenting species at s values outside the diffusion-broadened Gaussian peak, in theory it is not as accurate as the approach of solving the Lamm equation, Equation (3), due to the theoretical approximations made.⁵³ Further, if there is unrecognized heterogeneity of species within one $g^*(s)$ peak, as outlined above, it will only provide a lower limit to the molar mass.

A conceptual connection exist between the $g^*(s)$ and the integral $G(s)$ distributions by van Holde–Weischet, which can be obtained via extrapolation of $g^*(s)$ to infinite time,⁵⁰ and between $g^*(s)$ and $c(s)$, which is described in the following section and can be considered an extension of $g^*(s)$ from an apparent distribution of non-diffusing particles to the distribution of diffusing particles.

4.2.4 Sedimentation Coefficient Distributions $c(s)$

The goal of the $c(s)$ distribution is to calculate a distribution of sedimenting species taking into account their diffusion. This is accomplished by an extension of Equation (8), in which the full solution of the Lamm equation for sedimentation and diffusion is considered for each sedimenting species:

$$a(r, t) \cong \int c(s) \chi_1(s, D(s), r, t) ds \quad (9)$$

where $\chi_1(s, D, r, t)$ denotes the normalized Lamm equation solution (Equation (3)).⁵⁷ Essentially, analysing the experimental data with Equation (9) is equivalent to finding the best combination of a large set of Lamm equation solutions that fits the data best. The problem of estimating the extent of diffusion $D(s)$ for each species with a certain s value is approached with a hydrodynamic scaling law that assumes that all particles have a uniform frictional ratio, ff_0 , taking advantage of the weak size dependence of $D \sim s^{-1/2}$ and the weak dependence of the frictional ratio on macromolecular shape. The parameter ff_0 is adjusted to a best-fit estimate in the fitting process. A comparison of this method with other approaches can be found in ref. 50, and the detailed procedure is described in ref. 48. Like the $ls-g^*(s)$ distribution, the systematic noise contributions from IF detection system are calculated explicitly and can be subtracted from the raw sedimentation profiles, and the $c(s)$ distribution produces a detailed model of the original sedimentation profiles which can be assessed with conventional measures of the goodness-of-fit. This approach is implemented in the software SEDFIT and SEDPHAT. The results of the application to the data in Figure 2 are shown in Figure 5.

Because diffusion is accounted for by using Lamm equation solutions as model functions in Equation (9), the resulting sedimentation coefficient distribution has sharp features that are not diffusively broadened. The information on boundary spreading can be recovered by applying the Svedberg equation (Equation (1)) for

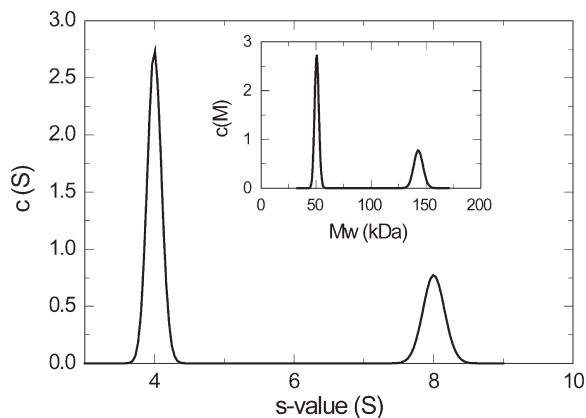


Figure 5 Application of $c(s)$ to the data shown in Figure 2. The inset shows the corresponding $c(M)$ distribution from the same fit

each pair of s and D values to transformation $c(s)$ into a molar mass distribution $c(M)$. Because the relationship between s and D for each species can be adjusted to specific problems, several variants of $c(s)$ are available for studies of protein conformational changes, for bimodal sedimentation coefficient distributions, for sedimentation of macromolecules in the presence of small molar mass species and for global analysis of multiple data sets with combinations of discrete and piecewise continuous distributions. A generalization to take advantage of multisignal detection for the study of heterogeneous protein–protein interactions has been described.⁵⁸

Beyond the high resolution and the sensitivity to trace components, the $c(s)$ method has the property of being able to be employed over the whole data set of SV profiles from the very beginning of sedimentation to the complete depletion of visible material, a distinct advantage over the $g^*(s)$ and van Holde–Weischet methods. The $c(s)$ method can also be used to analyse data from very small peptides, where the plateau region has collapsed before the meniscus has cleared. Such samples are dominated by back-diffusion from the bottom of the cell, and have in the past been very difficult to analyse, however, the $c(s)$ method copes as well with these as with fast sedimenting samples. A disadvantage of the method is introduced by the possible ambiguity of solving Equation (9) with slightly different distributions that may fit the data equally well. Maximum entropy regularization is embedded in the algorithm in order to produce always the simplest distribution that fits the data with a quality statistically indistinguishably from the best-fit, but this can have a tendency to merge neighbouring peaks. Further, caution must be used in the interpretation of the $c(M)$ distribution, which is valid only if there is a single major peak, otherwise it is limited to the case where all species have the same frictional ratio ff_0 , which may not be known *a priori*, or models for bimodal distributions must be used. Further, $c(s)$ can provide a significant amount of detail on the sample, but a pre-requisite of a reliable interpretation is that a very good fit of the raw sedimentation data can be achieved. This method is covered in detail in Chapter 2.

4.3 Analysis of Interacting Systems by Sedimentation Velocity

The experimental data basis for the study of protein interactions should include SV data obtained over a large range of different loading concentrations of the individual protein components. Then, for hetero-associations it is necessary to obtain data at different reactant concentrations and molar ratios, ideally covering a range from far below to far above the K_D of the interaction. A good starting point for the analysis of interacting systems with SV is to derive the $c(s)$ or $g^*(s)$ distributions. While the van Holde–Weischet method is excellent to diagnose qualitatively the presence of an interaction, by displaying a concentration-dependent increase in the sedimentation coefficients and positive sloped $G(s)$ distributions, no theory is currently known to allow rigorous thermodynamic or hydrodynamic parameters of reversibly interacting proteins to be derived from the quantitative analysis of $G(s)$. For clarity in the following sections, the diagnosis and analysis of interaction is illustrated as it appears in the $c(s)$ method, which is described in more detail in Chapter 2. However, Chapter 3 discusses the implementation of the time-difference modelling in some depth, while more detailed information on the van Holde–Weischet method can be found in Chapter 11.

The presence of an interaction changes the sedimentation profiles in SV in three different ways: (1) owing to the molar mass of the complex being higher than that of the individual components, the average sedimentation coefficient is increased in a concentration-dependent manner, reflecting the concentration dependent population of the complex; (2) because of the altered sedimentation process, which is now influenced also by the chemical reaction between the sedimenting species, the boundary shapes have different features; and (3) owing to the (even transient) formation of complexes, boundary components can appear with co-sedimentation of the individual protein components and therefore the boundary features may not necessarily reflect the properties of the separate species, but rather reflect the sedimenting system. Correspondingly, different analysis strategies can be applied that exploit these different aspects.

First, the weight-average sedimentation coefficient s_w can be obtained by integration of $g^*(s)$ or $c(s)$ distributions and plotted as a function of loading concentration.²⁰ Although it may seem counter-intuitive at first due to the hydrodynamic and kinetic processes observed in SV, this isotherm $s_w(c)$ permits a rigorous thermodynamic analysis of the interaction, because s_w can be strictly obtained from applying mass balance consideration to the sedimentation boundaries.^{1,59} Irrespective of the boundary shape or the peak structure in $c(s)$ or $g^*(s)$, from inspecting the concentration dependence of s_w the presence or absence of an interaction can be discerned. The quantitative analysis of $s_w(c)$ can proceed by standard techniques to derive s values of the individual species and complexes and the equilibrium binding constant.⁵⁹ For a 1:1 hetero-association, the isotherm takes the form

$$s_w(c_A^{\text{tot}}, c_B^{\text{tot}}) = \frac{\epsilon_A c_A^{\text{free}} s_A + \epsilon_B c_B^{\text{free}} s_B + (\epsilon_A + \epsilon_B) K c_A^{\text{free}} c_B^{\text{free}} s_{AB}}{\epsilon_A c_A^{\text{tot}} + \epsilon_B c_B^{\text{tot}}} \quad (10)$$

$$c_A^{\text{tot}} = c_A^{\text{free}} + K c_A^{\text{free}} c_B^{\text{free}} \quad c_B^{\text{tot}} = c_B^{\text{free}} + K c_A^{\text{free}} c_B^{\text{free}}$$

(with ϵ denoting the extinction coefficients and K the equilibrium association constant). This approach was reviewed recently in refs. 20 and 60. For convenience, this

and other isotherm models described below for sedimentation analysis have been implemented for global modelling in the software SEDPHAT.

The effect of the boundary shapes can be observed in different ways. With the $c(s)$ methodology, interactions are detected either by the emergence of new peaks or by shifts in the peak position as the concentration of one or both reactant is altered. Peaks that remain constant in position but change in relative area are indicative of complex formation that is slow on the time-scale of sedimentation. A shift of peak positions indicates more rapid chemical interconversion of species during sedimentation, such that the sedimenting system assumes an average sedimentation rate between those of the reacting species. This is illustrated in Figure 6 for hetero-associating systems (analogous illustrations for self-association can be found in refs. 23 and 33). For slow interactions, the relative areas of the $c(s)$ peaks as a function of loading concentration can be analysed in the form of isotherms of species populations to determine K_D . For rapid interactions, Gilbert and Jenkins have theoretically analysed the sedimentation of the system. For the case of a heterogeneous interactions of a two-component mixture, they predicted the formation of two boundaries – one reflecting the undisturbed sedimentation of one free species, and the second reflecting the co-sedimentation of free species of both components together with the complex. These boundaries occur in distinct proportions and at well-defined sedimentation coefficients in between the s values of the free component and the complex.^{61,62} Accordingly, integration of $c(s)$ peaks of fast reactions can give isotherms of the amplitude and s value of the reaction boundary that can be modelled on the basis of Gilbert–Jenkins theory,⁶³ in combination with the isotherms of weight-average s -values, to derive the binding constant of the interaction.

The analyses of isotherms derived from the weight-average s values, the species populations, and the reaction boundary can also give valuable insights to define the reaction scheme and the stoichiometry. Such an analysis is discussed in more detail by Schuck in Chapter 2, and in the refs. 20, 63 and 64.

Finally, a method to obtain more direct insights in the stoichiometry of the complexes formed is the multisignal sedimentation coefficient distribution $c_k(s)$, available in the software SEDPHAT and described in ref. 58. Conceptually, it exploits spectral differences of proteins to analyse the composition of the different $c(s)$ peaks, which can be advantageous, in particular, in the presence of extended associations forming multiple complexes.

4.4 Reaction Kinetics in Sedimentation Velocity

While AUC is not normally a technique associated with the determination of reaction rate information, as indicated in Figure 6, it is sensitive to the rate of chemical conversion relative to the time-scale of sedimentation. For SV experiments with currently accessible rotor speeds, reactions can be considered fast, or essentially instantaneous, on the time-scale of sedimentation if the chemical off-rate constant k_{off} is larger than 10^{-2} s^{-1} . On the other extreme, complexes can be considered essentially stable species if the off-rate constants are lower than 10^{-5} s^{-1} . Therefore, inspection of $c(s)$ curves can permit the qualitative diagnostics of the reaction kinetics. The range of kinetic rate constants where the complex formation is comparable in time scale to the sedimentation, in particular reactions with k_{off} between 10^{-3} and 10^{-4} s^{-1} , falls in between those indicated for slow and fast reactions in Figure 6.

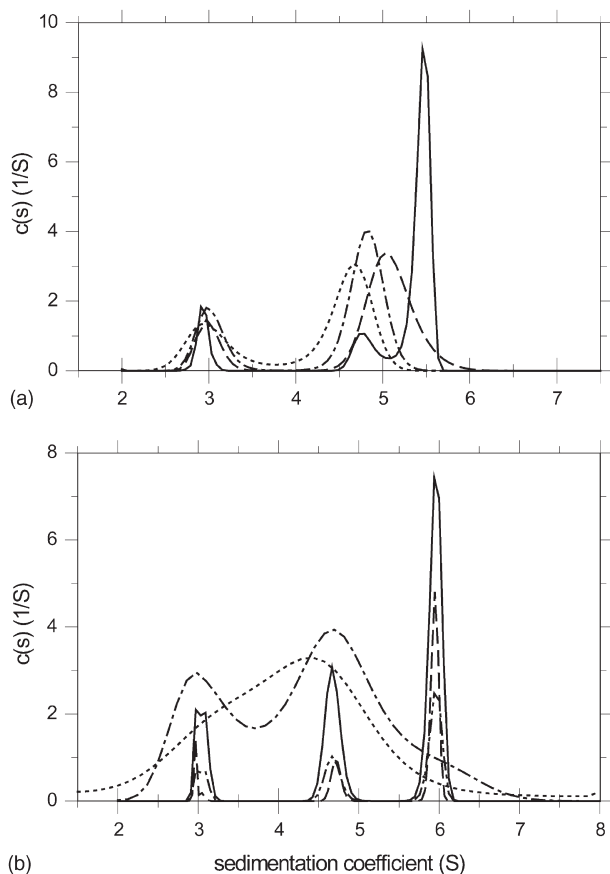


Figure 6 Examples of $c(s)$ distributions for interacting systems. Shown are (scaled) distributions for the hetero-association of a 30-kDa protein and 3 S with a second species of 50 kDa and 4.5 S forming a 6 S complex of 1:1 stoichiometry. The influence of the reaction kinetics can be seen in comparison of the fast interaction (a) and a slow interaction (b) on the time-scale of sedimentation. Data were simulated mimicking interference optical detection at a rotor speed of 50 000 rpm, with the usual signal-to-noise ratio. While for rapid reactions the peak position of the reaction boundary changes with concentration, slow reactions result in peak positions that are independent of concentrations. However, it must also be considered that the distributions at the lowest concentrations are broadened because of the limiting signal to noise ratio, whereas the distributions obtained from high concentrations are sharper. For rapid reactions 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. (a) Rapid interaction with $k_{\text{off}} = 1 \times 10^{-2} \text{ s}^{-1}$ and an equilibrium constant of $K_D = 10 \mu\text{M}$, with equimolar loading concentrations of 1 (dotted line), 3 (dash-dotted line), 10 (dashed line), and 30 μM (solid line). (b) Slow interaction with $k_{\text{off}} = 3.2 \times 10^{-5} \text{ s}^{-1}$, $K_D = 3.2 \mu\text{M}$, and equimolar loading concentrations at 0.3 (dotted line), 1 (dash-dotted line), 3 (dash-dot-dotted line), 10 (dashed line), and 30 μM (solid line)

Although the isotherm approach from integration of $c(s)$ peaks, in particular the isotherm of weight-average s values $s_w(c)$ can still provide a valid and highly useful approach to characterize the interaction, the direct modelling of the data with Lamm equations incorporating reaction terms can provide a more complete description. An example of such a Lamm equation is that for a 1:1 hetero-association:

$$\frac{\partial c_i}{\partial t} = \frac{1}{r} \frac{\partial}{\partial r} \left[r D_i \frac{\partial c_i}{\partial r} - s_i \omega^2 r^2 c_i \right] + \sigma_i q, \quad i = 1, 2, 3 \quad (11)$$

where

$$q = \frac{k_{\text{off}}}{K_d} c_1 c_2 - k_{\text{off}} c_3 \quad \text{and} \quad \sigma_1 = \sigma_2 = -\sigma_3 = -1 \quad (12)$$

q represents the chemical flux, and the indices 1 and 2 describe the individual free reactants, and 3 the complex. Algorithms for solving Equation (11) have been reviewed in 1981 by Cox and Dale,⁷ but it has become practical only recently to perform these calculations efficiently enough to model globally this equation to data obtained from different loading concentrations and molar ratios, and thereby estimate the equilibrium and chemical rate constant. This exploits the specific influence of the reaction kinetics on the boundary shape, as illustrated in Figure 7. This

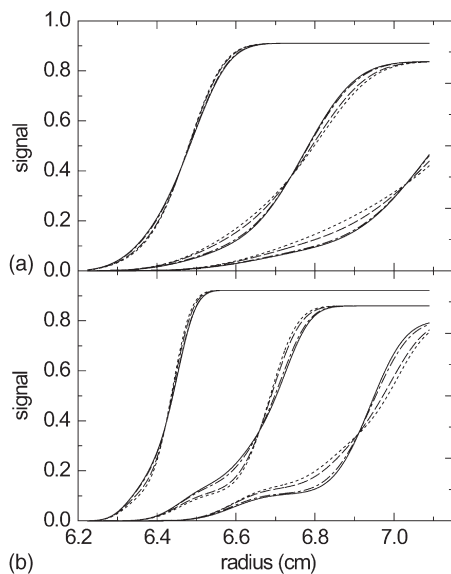


Figure 7 Boundary shapes differ for systems with different reaction kinetics. Characteristic shapes are more pronounced for larger species with correspondingly smaller diffusion coefficient. (a) Concentration profiles at 3 000, 9 000, and 12 000 s for the sedimentation of the system shown in Figure 6, at a concentration three-fold above K_D . Kinetic rate constants are $k_{\text{off}} = 10^{-2}$ (dotted line), 10^{-3} (dashed line), 10^{-4} (dash-dotted line), and 10^{-5} (solid line). (b) Sedimentation under similar conditions but with three-fold larger protein species, at times of 500, 1000, and 1500 s

approach has been implemented for modelling time-difference data in SEDANAL,⁶⁵ as outlined in more detail in Chapter 3, and for direct modelling of the raw data with explicit models for the systematic noise components in the software SEDPHAT.^{64,66} The advantage of this approach is that it permits to fully exploit the details of the boundary shapes, as opposed to the weight-average sedimentation coefficient and the amplitudes and s values of the $c(s)$ peaks of the reaction boundaries only. The disadvantage is that these analyses are significantly more difficult and time-consuming in practice, have a far higher requirement for sample purity and a relatively narrow range of kinetic rate constants for which the analysis is well determined.

It should be strongly emphasized that it is not always possible to assign unambiguously the correct model using SV alone, as there can be a similarity in the boundary shapes of a rapid higher-order complex formation with those from a slow multi-stage reactions. Ambiguity of the boundary interpretation is more problematic for small proteins, where the hydrodynamic separation and the chemical reaction is masked by a higher diffusion coefficient. In this case, the combination with information from other biophysical techniques, in particular those reporting on the time-scale of reaction or on the potential association scheme, can be highly valuable and provide the necessary information required for correct model assignment.

Whenever possible the interaction should also be studied by SE, which is described in the following. SE is independent of reaction kinetics and hydrodynamic parameters, but has intrinsically a lower size-dependent resolution and therefore benefits very significantly from a prior characterization of sample purity by SV. SV and SE are highly complementary, and both approaches should give consistent results.^{67,68}

5 Sedimentation Equilibrium

At lower rotor speeds, the transport of sample down the centrifuge cell is balanced by its desire to diffuse back up the cell due to the creation of a concentration gradient. SE is established when no change in the concentration distribution of any component is detectable (Figure 8). The time required to attain SE is proportional to the square of the solution column height, and, for proteins <200 kDa, is usually on the order of 1–2 days for sample volumes of 100–180 μL .

For a single species under ideal conditions, *i.e.* in the absence of repulsive interactions due to volume exclusion or charge interactions, the concentration distribution can be described by

$$a(r, t) = c(r_0)\epsilon d \exp\left[M(1 - \bar{v}\rho)\frac{\omega^2}{2RT}(r^2 - r_0^2)\right] \quad (13)$$

where $c(r_0)$ is the concentration at a reference radius, ϵ the molar extinction coefficient and d the optical pathlength (usually 1.2 cm).⁴ Ideally, the observed concentration range can easily be varied over 100-fold for one sample; using several absorbance wavelengths and interference optics means that 1 to 2000-fold concentration range can easily be probed. This should be combined with SE experiments conducted at several different loading concentrations and molar ratios of mixtures, as well as the acquisition of SE data consecutively at two to three different rotor speeds.

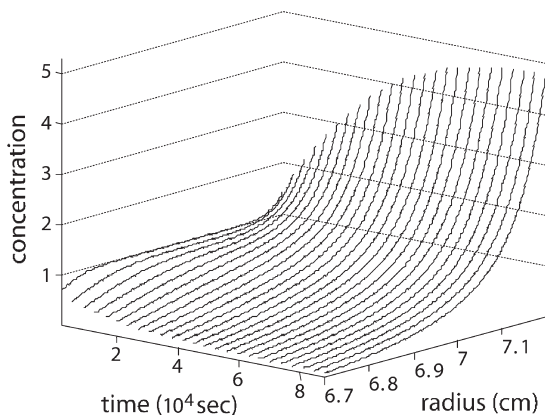


Figure 8 Concentration distributions during the approach to equilibrium. Sedimentation of the protein mixture from Figure 2 is shown, at a shorter solution column and at a rotor speed of 12 000 rpm. Only the final equilibrium scans will be subjected to data analysis with thermodynamic models of sedimentation equilibrium

Therefore, SE is extraordinarily powerful at assaying self-associating systems, where different oligomeric states are populated at different concentration ranges.

5.1 Setting Up a Sedimentation Equilibrium Experiment

When performing a SE experiment, the one parameter most likely to cause the novice the most difficulty is the correct selection of rotor speed. Too slow a rotor speed and there is not enough curvature in the data for adequate curve fitting; too fast a rotor speed and the majority of the sample will pellet to the bottom of the cell, making measurement impossible. As a rough guide, when working with 180 μL samples, for an average molar mass M of the protein species of interest, typically one can select a set of three rotor speeds with

$$\text{rpm} = \sqrt{\frac{100}{M(\text{kDa})}} \times 8000, \quad \sqrt{\frac{100}{M(\text{kDa})}} \times 12\,000, \quad \text{and} \quad \sqrt{\frac{100}{M(\text{kDa})}} \times 15\,000 \quad (14)$$

Typical SE profiles are shown in Figure 9, showing a typical ‘low-speed’ profile⁶⁹ at the lowest rotor speed with concentration ratios of $\sim 5:1$ from meniscus to bottom, and the so-called ‘meniscus-depletion’ condition⁷⁰ at the highest rotor speed. The SE experiment will start with the approach to equilibrium at the lowest rotor speed. When SE has been attained and the data acquired, the rotor speed is increased to the second, and, again after attainment of SE and data acquisition to the third (the highest).

The time required for attaining equilibrium with the assumed 180 μL samples ($\sim 5\text{-mm}$ columns) at the first rotor speed is typically of the order of 24–48 h, and equilibria at the higher rotor speed are approached in slightly less time. For 3 mm column length, 16–20 h are typical, but higher rotor speeds than those in Equation (14) have to be chosen to achieve sufficient curvature, and still shorter times are possible

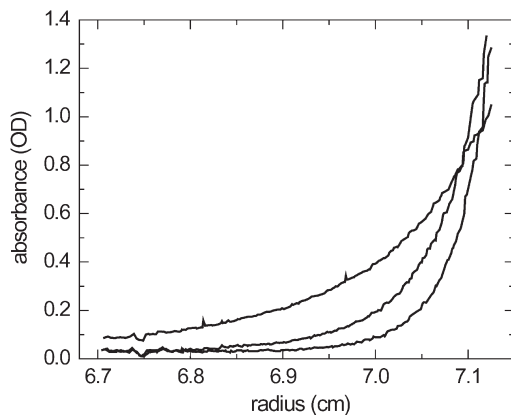


Figure 9 Typical shapes of absorbance distributions in sedimentation equilibrium at different rotor speeds

with short-column techniques.^{18,71} It should be noted, however, that the information content of the concentration profiles significantly increases with larger column height.¹⁶ Longer times are required for proteins with higher molar mass; as well as for reversible proteins complexes where the dissociation rate constant is very small, and for very elongated molecules which have unusually low diffusion coefficients. Additionally when working with buffers of high viscosity or density, such as 5–10% glycerol, additional time to reach equilibrium is required. Typically, in order to minimize protein degradation during the extended time of the experiments, in the absence of other considerations, SE is frequently conducted at low temperature.

It is very important to check that equilibrium has truly been attained. This cannot be assessed from the shape of the concentration profiles, but is done, typically, by subtracting successive SE traces taken in intervals of at least 4 h apart. When no systematic deviation is seen across the cell, then the sample is regarded to have attained equilibrium. The software WINMATCH⁷² greatly facilitates this procedure, and permits to visualize the asymptotic approach to equilibrium by displaying the r.m.s. difference of all previous scans relative to the last scan.

SE experiments are *always* conducted over a range of initial loading concentrations. This should ideally be combined with data obtained at multiple wavelengths of the ABS detection and ideally in combination with interference optics in order to increase the dynamic range of the concentration gradient detected in the SE experiment. For the latter, however, a mechanical stabilization process for the centrifuge cell assemblies has to be applied,^{22–24,73} and the systematic radial-dependent signal offsets of the IF detection have to be removed by subtraction of a water blank or by computational analysis.²² The use of multiple rotor speeds and loading concentrations is particularly important for the analysis of protein–protein and protein–DNA interactions. The data acquired over a range of signals, concentrations, and rotor speeds can then be analysed by global analysis, and from these appropriate errors in the parameters can be defined. More details on the planning SE experiments can be found in the protocol in ref. 33.

As mentioned above, it is highly recommended to perform a characterization of the proteins by SV prior to the SE experiment, in order to establish the interaction model and the purity of the samples. As in SV, the study of heterogeneous protein–protein interactions requires the individual components to be characterized separately.

For the data analysis, beyond the global analysis of data from different rotor speeds and loading concentration, strategies to improve the modelling and the resulting parameter estimates include the application of a constraint for conservation of mass, which relates the total observed mass in equilibrium with the loading concentration¹⁷ or among the data from the different rotor speeds,^{22,74} as well as the algebraic decomposition of systematic signal offsets.²² An analysis of optimal statistical weights for the data points is provided in Chapter 8. Several software packages are available for the analysis of SE, with varying degree of flexibility and capabilities for the analysis of self-association and hetero-association.

5.2 Self-Association

Sedimentation equilibrium data can be fitted with specific models for self-association and from this, binding constants can be determined. For any interacting system, SE is established if the sample is both in ‘mechanical’ sedimentation–diffusion equilibrium and in chemical equilibrium, and therefore, all interacting species distribute so that they follow mass action law at all places in the centrifuge cell. Taking the simplest example, a monomer–dimer self-association, mass action means that the amount of dimer can be expressed as $c_2 = K_{12}c_1^2$ (with concentrations and binding constants expressed in molar units), and for the SE profile then follows the relationship

$$a(r) = c_1(r_0)\varepsilon_1d \exp\left[M_b \frac{\omega^2}{2RT}(r^2 - r_0^2)\right] + K_{12}c_1(r_0)^2\varepsilon_1d \exp\left[2M_b \frac{\omega^2}{2RT}(r^2 - r_0^2)\right] \quad (15)$$

where M_b denotes the monomer buoyant molar mass $M = (1 - \bar{v}\rho)$. This assumes an ideally sedimenting system in the absence of repulsive interactions from charge or steric repulsion, which is fulfilled in most cases for dilute protein solutions ($<1 \text{ mg mL}^{-1}$) in buffers containing sufficient ionic strength. Wills and Winzor discuss non-ideality at some length in SE in Chapter 4; Minton and co-workers⁷⁵ also assess the effect of non-ideality for highly concentrated solutions rigorously.

As the sedimentation creates a concentration gradient, a SE experiment over several speeds and loading concentrations has the potential to contain information over a large portion of the binding isotherm. Several other biophysical techniques require fitting of exponentials to experimental data. However, in contrast to most techniques, the exponents in SE are usually known (*i.e.*, M_b should be known), which makes the problem more tractable. As outlined above, the effective partial-specific volume, and hence M_b , for most proteins under common experimental conditions can be predicted with sufficient accuracy from amino acid composition or mass spectrometry^{19,34} (and for glycoproteins the carbohydrate composition).^{76,77} For proteins under conditions of high co-solvent concentration leading to preferential solvation,^{23,34,36}

or proteins containing other non-amino acid components, such as protein–detergent or protein–lipid complexes, the density increment (or effective partial-specific volume) can be determined, for example, by conducting SE⁷⁸ or SV⁷⁹ experiments in H₂O and D₂O, or by densitometry.⁸⁰ This topic is addressed in Chapters 7 and 19. Fortunately, SE itself can often be used to determine directly experimentally M_b , if conditions can be established where the proteins are monomeric. If M_b is included as a fitting parameter, correlations of M_b and K_{12} can occur that increase the uncertainty of both, in particular for experiments where concentrations are achieved that only partially span the binding isotherm.

5.3 Heterogeneous Interactions

The simplest heterogeneous interaction is for two components A and B to form a 1:1 complex. Here, we can specify the mass action law $c_{AB} = K_a c_A c_B$. The total signal seen at sedimentation equilibrium then becomes:

$$a(r) = c_A(r_0)\varepsilon_A d \exp\left[M_{b,A} \frac{\omega^2}{2RT}(r^2 - r_0^2)\right] + c_B(r_0)\varepsilon_B d \exp\left[M_{b,B} \frac{\omega^2}{2RT}(r^2 - r_0^2)\right] + K_a c_A(r_0)c_B(r_0)(\varepsilon_A + \varepsilon_B)d \exp\left[(M_{b,A} + M_{b,B}) \frac{\omega^2}{2RT}(r^2 - r_0^2)\right] \quad (16)$$

It is possible to combine Equations (15) and (16) for more extended association equilibria, and to include mixed self-association and further hetero-associations. More in-depth thermodynamic theory of SE of multicomponent mixtures can be found in the chapter by Willis and Winzor (Chapter 4), and in the work by Minton and co-workers.⁷⁵ In the analysis of heterogeneous associations, the buoyant molar mass values $M_{b,A}$ and $M_{b,B}$ can be determined directly by studying the individual protein components separately.

One difficulty in the analysis of hetero-associations arises if the molar mass values of the components differ by less than 20% or more than a factor 5, because in this case the unbound or the bound components cannot be directly distinguished from the shape of the sedimentation profiles due to mathematical correlation of their exponential signal contributions. This makes a direct analysis of the binding very difficult. This problem can be overcome in two ways. First, multiwavelength analysis techniques can be applied if the protein components exhibit significantly different absorption properties (for example, extinction coefficients at 280 and 250 nm), or if one of the components has extrinsic chromophoric labels.^{22,76,81,82} This is particularly useful for protein–DNA interactions (see Chapter 10). Similar, the global multisignal analysis with ABS and IF detection system is possible, exploiting differences in the weight-based extinction coefficients of proteins.²² It is also possible to intrinsically label proteins by incorporating fluorescent analogues of tryptophan, such as 7-azatryptophan or 5-hydroxytryptophan.^{83,84,85} Second, such interactions can also be studied by imposing a mathematical constraint whereby the total mass of each component in the solution column is related either to the loading concentration^{10,17,86} or conserved between the equilibria at different rotor speeds.^{22,74} Further improvement in the data analysis can be possible if the experiment is conducted in titration series

with serial dilutions or if one component is kept constant – designs that enable to establish mathematical relationships between the total mass of material implied in the mathematical models for the different cells.^{22,23}

6 Conclusions

This review is by no means exhaustive. Further important topics covered in this book include methods to determine large-scale properties of genomes (Chapter 5), synthetic boundary formation (Chapter 6), and polysaccharide and polymer characterization (Chapters 12, 147 and 23). One particular fertile set of methodologies is relating sedimentation studies to those structures derived from X-ray crystallography (Chapters 20 and 21) and small-angle scattering (Chapters 13, 15, 16 and 18). It is hoped that this review has given some help to the novice in understanding the principles behind an analytical ultracentrifugation experiment and that it facilitated access to the in-depth discussion of the techniques found in subsequent chapters in this book, and the references therein.

References

1. H. K. Schachman, *Ultracentrifugation in Biochemistry*, Academic Press, New York, 1959.
2. C. Tanford, *Physical Chemistry of Macromolecules*, Wiley, New York, 1961.
3. H. Fujita, *Foundations of Ultracentrifugal Analysis*, Wiley, New York, 1975.
4. T. Svedberg and K. O. Pedersen, *The Ultracentrifuge*, Oxford University Press, London, 1940.
5. B. Elzen, Scientists and rotors. The development of biochemical ultracentrifuges, Dissertation, University Twente, Enschede, 1988.
6. H. K. Schachman, in *Analytical Ultracentrifugation in Biochemistry and Polymer Science*, S. E. Harding, A. J. Rowe and J. C. Horton (eds), Royal Society of Chemistry, Cambridge, 1992, 3–15.
7. D. J. Cox and R. S. Dale, in *Protein-Protein Interactions*, C. Frieden and L. W. Nichol (eds), Wiley, New York, 1981.
8. B. Demeler, H. Saber and J. C. Hansen, *Biophys. J.*, 1997, **72**, 397–407.
9. J. C. Hansen, J. Lebowitz and B. Demeler, *Biochemistry*, 1994, **33**, 13155–13163.
10. J. S. Philo, *Proceedings of Ultrasensitive Biochemical Diagnostics II*, SPIE, San Jose, CA, 1997, 170–177.
11. J. Lebowitz, M. Teale and P. Schuck, *Biochem. Soc. Transact.*, 1998, **26**, 745–749.
12. G. Rivas, W. Stafford and A. P. Minton, *Methods: A Companion to Methods in Enzymology*, 1999, **19**, 194–212.
13. F. Arisaka, *Tanpakushitsu Kakusan Koso*, 1999, **44**, 82–91.
14. T. M. Laue, *Current Protocols Protein Sci.*, 1999, 20.3.1–20.3.13.
15. T. M. Laue and W. F. I. Stafford, *Annu. Rev. Biophys. Biomol. Struct.*, 1999, **28**, 75–100.
16. P. Schuck and E. H. Braswell, in *Current Protocols in Immunology*, J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach and W. Strober (eds), John Wiley & Sons, New York, 2000, 18.8.1–18.8.22.
17. J. S. Philo, *Method. Enzymol.*, 2000, **321**, 100–120.
18. J. J. Correia and D. A. Yphantis, in *Analytical Ultracentrifugation in Biochemistry and Polymer Science*, S. E. Harding, A. J. Rowe and J. C. Horton (eds), The Royal Society of Chemistry, Cambridge, U.K., 1992, 231–252.

19. J. Lebowitz, M. S. Lewis and P. Schuck, *Protein Sci.*, 2002, **11**, 2067–2079.
20. P. Schuck, *Anal. Biochem.*, 2003, **320**, 104–124.
21. J. L. Cole, *Method. Enzymol.*, 2004, **384**, 212–232.
22. J. Vistica, J. Dam, A. Balbo, E. Yikilmaz, R. A. Mariuzza, T. A. Rouault and P. Schuck, *Anal. Biochem.*, 2004, **326**, 234–256.
23. A. Balbo and P. Schuck, in *Protein–Protein Interactions*, E. Golemis and P. D. Adams (eds), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, in print.
24. www.nih.gov/od/ors/dbeps/PBR/AUC.htm
25. www.analyticalultracentrifugation.com
26. M. S. Springer and H. K. Schachman, *Biochemistry*, 1974, **13**, 3726–3733.
27. D. J. Cox, *Science*, 1966, **152**, 359–361.
28. G. J. Howlett and P. D. Jeffrey, *J. Phys. Chem.*, 1973, **77**, 1250–1258.
29. J. B. Ifft, *Biophys. Chem.*, 1976, **5**, 137–157.
30. M. A. Bothwell, G. J. Howlett and H. K. Schachman, *J. Biol. Chem.*, 1978, **253**, 2073–2077.
31. S. Darawshe, G. Rivas and A.P. Minton, *Anal. Biochem.*, 1993, **209**, 130–135.
32. D. A. Yphantis and T. Arakawa, *Biochemistry*, 1987, **26**, 5422–5427.
33. <http://www.nih.gov/od/ors/dbeps/PBR/AUCProtocols.htm>
34. J. Lebowitz, M. S. Lewis and P. Schuck, *Protein Sci.*, 2003, **12**, 2649–2650.
35. T. M. Laue, B.D. Shah, T. M. Ridgeway and S. L. Pelletier, in *Analytical Ultracentrifugation in Biochemistry and Polymer Science*, S. E. Harding, A. J. Rowe and J. C. Horton (eds), The Royal Society of Chemistry, Cambridge, 1992, 90–125.
36. H. Eisenberg, *Biological Macromolecules and Polyelectrolytes in Solution*, Clarendon Press, Oxford, 1976.
37. J. Garcia De La Torre, M. L. Huertas and B. Carrasco, *Biophys. J.*, 2000, **78**, 719–730.
38. O. Lamm, *Ark. Mat. Astr. Fys.*, 1929, **21B**(2), 1–4.
39. J.-M. Claverie, H. Dreux and R. Cohen, *Biopolymers*, 1975, **14**, 1685–1700.
40. L. A. Holladay, *Biophys. Chem.*, 1979, **10**, 187–190.
41. J. S. Philo, *Biophys. J.*, 1997, **72**, 435–444.
42. P. Schuck, *Biophys. J.*, 1998, **75**, 1503–1512.
43. P. Schuck, C. E. MacPhee and G. J. Howlett, *Biophys. J.*, 1998, **74**, 466–474.
44. J. Behlke and O. Ristau, *Biophys. Chem.*, 2002, **95**, 59–68.
45. P. Schuck, *Biophys. Chem.*, 2004, **108**, 187–200.
46. P. Schuck, *Biophys. Chem.*, 2004, **187**, 201–214.
47. A. Solovyova, P. Schuck, L. Costenaro and C. Ebel, *Biophys. J.*, 2001, **81**, 1868–1880.
48. J. Dam and P. Schuck, *Method. Enzymol.*, 2004, **384**, 185–212.
49. K. E. van Holde and W. O. Weischet, *Biopolymers*, 1978, **17**, 1387–1403.
50. P. Schuck, M. A. Perugini, N. R. Gonzales, G. J. Howlett and D. Schubert, *Biophys. J.*, 2002, **82**, 1096–1111.
51. B. Demeler and van K. E. Holde, *Anal. Biochem.*, 2004, **335**, 279–288.
52. W. F. Stafford, *Anal. Biochem.*, 1992, **203**, 295–301.
53. J. S. Philo, *Anal. Biochem.*, 2000, **279**, 151–163.
54. P. Schuck and P. Rossmanith, *Biopolymers*, 2000, **54**, 328–341.
55. W. F. Stafford, *Method. Enzymol.*, 1994, **240**, 478–501.
56. P. Schuck and B. Demeler, *Biophys. J.*, 1999, **76**, 2288–2296.
57. P. Schuck, *Biophys. J.*, 2000, **78**, 1606–1619.
58. A. Balbo, K. H. Minor, C. A. Velikovskiy, R. Mariuzza, C. B. Peterson and P. Schuck, *Proc. Natl. Acad. Sci. USA*, 2005, **102**, 81–86.
59. R. F. Steiner, *Arch. Biochem. Biophys.*, 1954, **49**, 400–416.
60. J. J. Correia, *Method. Enzymol.*, 2000, **321**, 81–100.

61. G. A. Gilbert and R. C. Jenkins, *Nature*, 1956, **177**, 853–854.
62. G. A. Gilbert, *Proc. Royal Soc. London A*, 1959, **250**, 377–388.
63. J. Dam and P. Schuck, *Biophys. J.*, 2005, **89**, 651–666.
64. J. Dam, C. A. Velikovskiy, R. Mariuzza, C. Urbanke and P. Schuck, *Biophys. J.*, 2005, **89**, 619–634.
65. W. F. Stafford and P. J. Sherwood, *Biophys. Chem.*, 2004, **108**, 231–243.
66. www.analyticalultracentrifugation.com/sedphat/sedphat.htm
67. S. Tarabykina, D. J. Scott, P. Heizyle, T. J. Hill, J. R. Tame, M. Kriajevska, I. Lafitte, P. J. Denick, G. G. Dodson, N. J. Maitland, E. R. Lukanidia, I. B. Bronstein, *J. Biol. Chem.*, 2001, **276**, 24212–24222.
68. J. Dam, R. Guan, K. Natarajan, N. Dimasi, L.K. Chlewicki, D.M. Kranz, P. Schuck, D.H. Margulies and R. A. Mariuzza, *Nature Immunol.*, 2003, **4**, 1213–1222.
69. E. G. Richards, D. C. Teller and H.K. Schachman, *Biochemistry*, 1968, **7**, 1054–1076.
70. D. A. Yphantis, *Biochemistry*, 1964, **3**, 297–317.
71. D. L. Sackett and R. E. Lippoldt, *Biochemistry*, 1991, **30**, 3511–3517.
72. www.biotech.uconn.edu/auf
73. A. T. Ansevin, D. E. Roark and D. A. Yphantis, *Anal. Biochem.*, 1970, **34**, 237–261.
74. D. E. Roark, *Biophys. Chem.*, 1976, **5**, 185–196.
75. S. Zorrilla, M. Jimenez, P. Lillo, G. Rivas and A. P. Minton, *Biophys. Chem.*, 2004, **108**, 89–100.
76. M. S. Lewis, R. I. Shragar and S.-J. Kim, in *Modern Analytical Ultracentrifugation*, T.M. Schuster and T. M. Laue (eds), Birkhäuser, Boston, 1993, 94–115.
77. S. Shire, *Determination of Molecular Weight of Glycoproteins by Analytical Ultracentrifugation*, Beckman Instruments, Palo Alto, CA, 1992.
78. J. Edelstein and H. Schachman, *J. Biol. Chem.*, 1967, **242**, 306–311.
79. Y. Gohon, G. Pavlov, P. Timmins, C. Tribet, J. L. Popot, and C. Ebel, *Anal. Biochem.*, 2004, **334**, 318–334.
80. C. Ebel, H. Eisenberg and R. Ghirlando, *Biophys. J.*, 2000, **78**, 385–393.
81. P. Schuck, *Progr. Colloid. Polym. Sci.*, 1994, **94**, 1–13.
82. F. Dölle and D. Schubert, *Progr. Colloid Polym. Sci.*, 1997, **107**, 77–81.
83. T. M. Laue, D. F. Senear, S. Eaton and J. B. Ross, *Biochemistry*, 1993, **32**, 2469–2472.
84. D. J. Scott, S. Leejeerajumnean, J. A. Brannigan, R. J. Lewis, A. J. Wilkinson and J. G. Hoggett, *J. Mol. Biol.*, 1999, **293**, 997–1004.
85. D. J. Scott, A. L. Ferguson, M. T. Gallegos, M. Pitt, M. Buck and J. G. Hoggett, *Biochem. J.*, 2000, **352**, 539–547.
86. M. S. Lewis and R. J. Youle, *J. Biol. Chem.*, 1986, **261**, 11572–11577.