Published in final edited form as:
Curr Protoc Protein Sci. 2013 February ; 0 20: . doi:10.1002/0471140864.ps2012s71.

Current Methods in Sedimentation Velocity and Sedimentation Equilibrium Analytical Ultracentrifugation

Huaying Zhao[*], **Chad A. Brautigam**[#], **Rodolfo Ghirlando**[$], and **Peter Schuck**[*]

[*]Dynamics of Macromolecular Assembly Section, Laboratory of Cellular Imaging and Macromolecular Biophysics, National Institute of Biomedical Imaging and Bioengineering, National Institutes of Health, Bethesda, Maryland 20892

[#]Department of Biochemistry, The University of Texas Southwestern Medical Center, Dallas, Texas 75390

[$]Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892

Abstract

Significant progress in the interpretation of analytical ultracentrifugation (AUC) data in the last decade has led to profound changes in the practice of AUC, both for sedimentation velocity (SV) and sedimentation equilibrium (SE). Modern computational strategies have allowed for the direct modeling of the sedimentation process of heterogeneous mixtures, resulting in SV size-distribution analyses with significantly improved detection limits and strongly enhanced resolution. These advances have transformed the practice of SV, rendering it the primary method of choice for most existing applications of AUC, such as the study of protein self- and hetero-association, the study of membrane proteins, and applications in biotechnology. New global multi-signal modeling and mass conservation approaches in SV and SE, in conjunction with the effective-particle framework for interpreting the sedimentation boundary structure of interacting systems, as well as tools for explicit modeling of the reaction/diffusion/sedimentation equations to experimental data, have led to more robust and more powerful strategies for the study of reversible

Internet Resources

There are abundant resources available on the internet covering software, detailed experimental step-by-step protocols, video instructions, and discussion forums. The following is only a small subset facilitating the application of the methods described in the present Unit.

General practical tools to help designing AUC experiments can be found at https://sedfitsedphat.nibib.nih.gov/tools/Protocols/Forms/AllItems.aspx, including detailed step-by-step protocols for conducting SE and SV experiments, a table with criteria for buffer selection, a grid for SV run conditions with different optical systems, instructions for radial calibration of the interference system, and a tutorial for the test for equilibrium. A video showing the assembly of cells can be found at http://www.jove.com/video/1530/assembly-loading-and-alignment-of-an-analytical-ultracentrifuge-sample-cell (Balbo et al., 2009).

All methods discussed above are implemented in SEDFIT and SEDPHAT, unless otherwise stated. They can be freely downloaded from https://sedfitsedphat.nibib.nih.gov/software/default.aspx, and an extensive online help system and documentation can be found at http://www.analyticalultracentrifugation.com/default.htm, including a tutorial for ‘getting started’ in SEDFIT, and step-by-step examples for the data analysis with *c(s)* http://www.analyticalultracentrifugation.com/examples.htm. ‘Getting started’ tools for SEDPHAT are available at http://www.analyticalultracentrifugation.com/sedphat/concepts_for_getting_started.htm.

Tutorials on special topics of AUC and/or SEDFIT or SEDPHAT analyses can be accessed at http://www.analyticalultracentrifugation.com/tutorials.htm. Movies illustrating the mechanism of migration o f there action boundary can be found at https://sedfitsedphat.nibib.nih.gov/tools/Reaction%20Boundary%20Movies/Forms/AllItems.aspx.

GUSSI is are available for download at http://biophysics.swmed.edu/MBR/software.

SEDNTERP for the calculation of buffer densities and viscosities, as well as protein extinction coefficients and partial specific volumes can be found at John Philo’s website http://www.jphilo.mailway.com/download.htm

Active discussion forums are SEDFIT-L (https://list.nih.gov/cgi-bin/wa.exe?SUBED1=SEDFIT-L&A=1), SEDPHAT-L (https://list.nih.gov/cgi-bin/wa.exe?A0=SEDPHAT-L), and RASMB (http://rasmb.bbri.org/cgi-bin/mailman/listinfo/rasmb).

The websites https://sedfitsedphat.nibib.nih.gov/workshop/default.aspx and http://www.analyticalultracentrifugation.com/default.htm will have current information on upcoming workshops.

protein interactions and multi-protein complexes. Furthermore, modern mathematical modeling capabilities have allowed for a detailed description of many experimental aspects of the acquired data, thus enabling novel experimental opportunities, with important implications for both sample preparation and data acquisition. The goal of the current commentary is to supplement previous AUC protocols, *Current Protocols in Protein Science* 20.3 (1999) and 20.7 (2003), and 7.12 (2008), and provide an update describing the current tools for the study of soluble proteins, detergent-solubilized membrane proteins and their interactions by SV and SE.

Keywords

sedimentation equilibrium; sedimentation velocity; chemical equilibria; reversible interaction; size-distribution; multi-protein complex; analytical ultracentrifugation; protein hydrodynamics

INTRODUCTION

Analytical ultracentrifugation (AUC) is a classical first-principle method of physical biochemistry, firmly rooted in solution thermodynamics. It allows for the real-time observation of the redistribution of proteins in dilute solution following the application of a strong centrifugal field. In sedimentation velocity (SV), the evolution of the sedimentation process is studied, whereas in sedimentation equilibrium (SE), the final equilibrium distribution is examined. SV and SE are complementary and together provide an information-rich characterization of many aspects of protein behavior in solution, including the protein mass and size, density, hydrodynamic shape, size-distribution and purity, weak protein interactions, specific reversible interactions, and the formation of multi-protein complexes. Among the virtues of AUC are the absence of surfaces, the absence of label requirements, and the powerful size resolution for protein-sized macromolecules and larger. By adjusting the rotor speed, peptides and proteins with molar masses ranging from 100 g/mol to 10^8 g/mol can be studied; furthermore the broad dynamic range of the detection systems in combination with current data analysis strategies allows for the characterization of interacting systems having equilibrium dissociation (K_d) between 10 nM and 10 mM (Zhao et al., 2012; Rowe, 2011).

Current major applications in protein science include the study of peptide and protein self-association and the characterization of oligomeric state, the characterization of heterogeneous protein-protein interactions and multi-protein complexes with regard to the number and size of co-existing complexes, as well as their affinity, the study of membrane proteins in detergent solutions, and the study of protein solution conformation and ligand-induced conformational changes.

Compared to the classical approaches, a profound change in the application of AUC has taken place approximately a decade ago, mainly due to the confluence of three new developments for SV: (1) Sophisticated and efficient algorithms for the precise numerical solution of the differential equation of the sedimentation/diffusion process, which can be applied routinely for the non-linear regression of experimental data using modern personal computer hardware. Historically, this has been a major limitation in SV methods for most of the previous century. (2) The introduction of mathematical tools, borrowed from modern image analysis, for direct data fitting based on integral equations for determining an unknown size distribution. In this manner, the level of detail of the mathematical model is adjusted to the experimental sensitivity of the method for polydispersity and trace impurities. (3) The adaptation of the model to accommodate the specific noise structure displayed by SV data, which is a combination of time-invariant and radius-invariant baseline offsets. These can now be directly determined from the least-squares fit of the data as part of

the model. This advance avoids 'manual' alignment, ambiguous data subset selection, and noise-amplifying data differencing. Consequently, the experimenter can now take full advantage of the large number of experimental data points, leading to lower detection limits, higher dynamic range, and an improved precision of the parameter estimates. These, together with other developments, have spawned many new approaches for studying mixtures of non-interacting and interacting proteins. The generalization of these concepts to multi-signal and density-contrast global analysis has led to new tools particularly suitable for heterogeneous protein interactions, multi-protein complexes, and membrane proteins in detergent solution. Complementarily, a new, more physically intuitive theoretical framework for the interpretation of the boundary structure of rapidly interacting systems was developed, which can be used for data analysis as well as experimental planning. Furthermore, more precise predictions of hydrodynamic friction from structure-based models are available. New software capabilities embed all these tools.

As a consequence of these developments, the majority of AUC applications, including the study of detergent-solubilized membrane proteins, have shifted towards the use of the SV mode. However, fresh developments have also taken place in SE. In particular, global-modeling capabilities have been extended, permitting a multi-signal, multi-speed global analysis and a global density-contrast analysis. In combination with new implicit mass-conservation constraints, these extend both the range of interacting systems and affinities that can be measured by SE, and the size range of binding partners in heterogeneous protein interactions.

Because the basic principles of both the instrumentation and the technique have not changed, the present commentary is aimed at providing additional and updated information to the previous more detailed Units for SE and SV of soluble proteins (*Current Protocols in Protein Science Units* 20.3 (1999), 20.7 (2003), 7.13 (2010)), as well as SE for detergent-solubilized membrane proteins (*Current Protocols in Protein Science* 7.12 (2008)). Elements of the newer techniques were presented previously in *Current Protocols in Immunology* 18.8 (2007) and 18.15 (2008) and will be referenced. A more recent detailed introduction to SV can be found in (Schuck et al., 2010), and a review for SE can be found in (Ghirlando, 2011). For the data analysis, we recommend our software SEDFIT and SEDPHAT, in which are implemented all tools discussed below. It can be downloaded from https://sedfitsedphat.nibib.nih.gov and an extensive online help system can be found at http://www.analyticalultracentrifugation.com. Workshops on current AUC methodology are held regularly in our laboratory at the National Institutes of Health, Bethesda, Maryland.

BASIC PRINCIPLES

For a description of the basic setup of AUC, we refer the reader to the previous Units cited above, as well as recent reviews (Rowe, 2010; Schuck, 2012; Zhao et al., 2012).

Sedimentation Velocity Analysis of Non-Interacting Systems

Basic Theory—The basic theory is also described in Unit 20.7, but briefly recapitulated here to provide the context and to introduce the symbols. We can define the sedimentation coefficient *s* as the linear velocity *u* of sedimentation a protein exhibits per unit gravitational field $\omega^2 r$ (with rotor angular velocity ω and distance from the center of rotation *r*):

$$s = \frac{u}{\omega^2 r} \quad \text{Equation 1}$$

It is measured in units of Svedberg, with 1S = 10^{-13} sec. Because it is dependent on buffer viscosity and density, this experimental *s*-value (s_{exp}) is often normalized to standard solution conditions of water at 20° C ($s_{20,w}$) according to

$$s_{20,w}=s_{\exp}\left(\frac{\eta_{\exp}}{\eta_{20,w}}\right)\left(\frac{1-\overline{v}\rho_{20,w}}{1-\overline{v}\rho_{\exp}}\right) \quad \text{Equation 2}$$

with ρ and η denoting the buffer density and viscosity, respectively, and $\bar{v}$ the partial-specific volume of the protein. The sedimentation coefficient depends on the translational friction coefficient *f* and molar mass *M*

$$f=\frac{M(1-\overline{v}\rho)}{sN_A}=6\pi\eta R_S \quad \text{Equation 3}$$

(with Avogadro's number N_A, and can be related to the Stokes radius R_S). It is useful to relate the frictional coefficient to that of an ideal, smooth, compact sphere of the same density and mass, which leads to the frictional ratio f/f_0 and to the ratio of *s*-values of the observed particle and the corresponding ideal, smooth, compact sphere:

$$f/f_0=s_{sphere,20w}/s_{20w} \quad \text{Equation 4}$$

We note that sedimentation velocities in excess of any particle's equivalent sphere, $s_{sphere,\ 20w}$, are impossible. Typically, due to frictional contribution from bound hydration water, f/f_0 -values between 1.1 – 1.2 are obtained for highly globular particles, 1.3 – 1.5 for moderately asymmetric particles, and in excess of 1.5 for highly asymmetric particles. From Eq. 3 we can also derive the Svedberg equation

$$M=\frac{sRT}{D(1-\overline{v}r)} \quad \text{Equation 5}$$

, which is of key importance because it relates the sedimentation coefficient, *s*, and diffusion coefficient, *D*, with the protein molar mass, *M*. It is based on the insight that, for dilute solutions, the frictional coefficient for sedimentation is the same as that governing diffusion (Svedberg and Pedersen, 1940).

Direct Boundary Modeling with Lamm Equation Solutions—In SV, we are not observing single particles, but rather an ensemble of particles with a radial- and time-dependent concentration $\chi(r,t)$. Generally, sedimentation starts with a well-mixed sample at a uniform loading concentration $\chi(r,t=0)=c_0$. For a single class of proteins, the evolution of the concentration for a sector-shaped cell in the centrifugal field is governed by the Lamm equation (Lamm, 1929)

$$\frac{\partial\chi}{\partial t}=-\frac{1}{r}\frac{\partial}{\partial r}\left(\chi s\omega^2r^2-Dr\frac{\partial\chi}{\partial r}\right) \quad \text{Equation 6}$$

The lack of analytical solutions to this equation has presented a significant obstacle for fully exploiting SV during most of the 20th century, even though numerical solutions have been sparsely used to simulate and analyze data (Claverie et al., 1975; Dishon et al., 1967; Cox, 1965). Abundant computational resources and the introduction of highly efficient numerical

solutions have now enabled a transformation of SV: rather than measuring the velocity of the boundary midpoint and calculating *s* via Eq. 1, we can fit directly solutions of Eq. 6 to experimental data $a(r,t)$ in a direct, least-squares boundary model (see below). Different strategies for solving Eq. 6 are implemented in different software packages: SEDFIT and SEDPHAT are based on the solutions described in (Brown et al., 2008) with adaptive adjustment to pre-determined precision. A different approach has been proposed by (Cao and Demeler, 2005) and implemented in ULTRASCAN, but if applied as suggested by Cao & Demeler it lacks the precision required for modeling SV boundaries of medium and large proteins (Schuck, 2009). Examples for the evolution of concentration profiles of different size particles are shown in Figure 1.

When experimental data are acquired using the absorbance optical (ABS) system, the radial profiles are not always well described in terms of a temporal snapshot due to the finite velocity of the scanner reaching higher radii with a time delay. This can produce significant errors in the sedimentation coefficients of large proteins and protein complexes (Brown et al., 2009). However, by evaluating the sedimentation profiles that would be measured in scan *i* as $\chi'(r,t_i) = \chi(r,t_i+(r-r_1)/v_{scan})$ and assigning later times to larger radii than the initial scan time stamp t_i, SEDFIT and SEDPHAT are able to account for the finite scanning velocity v_{scan} (typically 40 μm/sec).

The use of Eq. 6 in modeling data requires a definition of the initial conditions, including the geometric limits of the solution column, specifically the radial positions for the meniscus, *m*, and the cell bottom, *b*. Traditionally, the assignment of *m* was separate from, and had to precede, the evaluation of the scans. This was true in the pre-computer approach of plotting the movement of the boundary midpoint $\log(r/m)$ as a function of time, which has a slope of $\omega^2 s$ based on an integral form of Eq. 1. Similarly, this was true for the $g(s^*)$ transformation (Stafford, 1992), as implemented in SEDANAL and DCDT+. However, in the modern least-squares modeling of SV data, the meniscus is a fitted parameter and thus can be subject to the same optimization based on the quality of fit, much like other fitting parameters. Different groups have established that such a treatment of the meniscus parameter leads to a well-defined estimate, which is more precise than that discerned from visual inspection of the experimental scans (Brown et al., 2009; Besong et al., 2012; Gabrielson et al., 2007), mainly due to the simple facts that the location of the true meniscus location is obscured by optical imaging artifacts, and that the radial resolution of experimental scans is not sufficiently high. Correspondingly, refining the meniscus parameter usually improves the fit significantly and also leads to well-determined macromolecular sedimentation parameters. Correlation of the meniscus parameter with protein sedimentation parameters of interest can occur for very broad boundaries either of small peptide sedimentation, of data at low rotor speeds, of SV configurations with short columns, and/or data with highly polydisperse material or extended association schemes. For this reason, it is recommended to use graphically determined bounds as upper and lower limits for the non-linear regression of the meniscus parameter.

If normal high-speed SV experiments with well-formed sedimentation boundaries between clear solvent and solution plateaus lead to optimized meniscus values at or outside these bounds, this is usually a strong indication for the presence of convection. In this case, the best practice is to repeat the experiment, aiming for a more thorough temperature equilibration. However, some information can be rescued from such data by excluding some initial scans prior to temperature stabilization and either (1) letting the meniscus position refine to its best-fit position as determined by the progression of the boundaries after convection has ceased (Brown et al., 2009); or (2) using a model where sedimentation is initialized at some point after start of the sedimentation with the concentration distribution extracted from an experimental scan at that time (Cox, 1966; Schuck et al., 1998).

In analogy, the bottom position of the solution column can also be refined from the non-linear regression of the data. However, this step is only necessary when the experimental data exhibit back-diffusion in the radial range to be analyzed. Unless there are small solute components in the sample, this is usually not the case, since the back-diffusion of proteins in typical high-speed SV is too steep to optically image reliably, and produces concentrations that are too high for Eq. 6 to be sensible. Thus, the steep back-diffusion sections of the data are customarily discarded, and the bottom position may become irrelevant for the model.

Another important ingredient for modeling experimental SV data is the ability to describe the experimental imperfections that produce signals offsetting the macromolecular sedimentation signal of interest. For example, Figure 2 shows raw SV data acquired at a series of time points with the interference optical (IF) detection system during the sedimentation of a protein. Two types of offsets can clearly be discerned. (1) Time-dependent vertical offsets of the whole signal pattern caused partly by inconsistent offsets in the zero fringe shift assignment (which is intrinsically arbitrary due to the periodicity of the fringe pattern), and partly by vibrations or other adventitious time-dependent changes in the optical pathlengths on the nm scale. These offsets are termed RI noise, for ‘radius-invariant’ noise. (2) A constant radial signal pattern that is superimposed equally on all scans. This phenomenon is caused by radius-dependent imperfections in the optical system, and can have both elements of short and long spatial correlations. It is referred to as TI noise, for ‘time-invariant’ noise. These noise components are also present when other SV data acquisition systems are used, even though their magnitude is smaller. For example, data from the ABS system often have negligible RI noise, but invariably show significant TI noise components that consist of a background profile with short spatial correlation.

Previous approaches to address these noise components have required *ad hoc* alignment of scans to eliminate RI noise, e.g., from operator-selected data subsets in regions that are thought not to change with time. To eliminate the TI noise, schemes for pair-wise subtraction of scans were devised to eliminate the TI noise from difference data. This approach will usually introduce some bias and invariably lead to noise amplification. This strategy is implemented, for example, in SEDANAL and DCDT+. Recently, a more straightforward method was introduced that does not pose any of these drawbacks, by simply incorporating these TI and RI signal terms directly into the model as terms $b(r)$ and $\beta(t)$, respectively (Schuck and Demeler, 1999). The latter approach rigorously honors all degrees of freedom presented by the analysis problem, and was shown to be statistically optimal (Schuck, 2010c). The additional linear baseline parameters can be easily solved for when using appropriate modern optimization techniques.

The practical significance of modern noise decomposition is an increased precision, resulting, for example, in the ability to extract reliable information from SV data at very low signal/noise ratio and to use lower protein concentrations (such as 10 nM of a 50 kg/mol protein (Zhao et al., 2012)). After the fit, the noise estimates can be subtracted from the raw data to allow the visualization of the remaining macromolecular sedimentation data in the original data space for critical visual inspection of the model and of the information content of the data (Dam and Schuck, 2004).

To extend the description of the baselines, we can add to the SV model expressions for the signal arising from a mismatch in the geometry of the sample and reference solution column (solution column height) and/or of the buffer composition in the sample and reference sector (Zhao et al., 2010). Buffer salts exhibit significant sedimentation, which can be described extremely well with Eq. 6 (typically, with $D \sim 1.4 \times 10^{-5}$ cm^2/s and $s \sim 0.14$ S for NaCl in water at 20° C). If the resulting signal contributions from sedimentation in both sectors do not cancel, a contribution $b_{buff}(r,t)$ is superimposed on the data, usually taking the form of a

slowly varying characteristic tilt of the baseline. The ability to include this term into the sedimentation model, where necessary, relieves experimental constraints in the perfectness of sample preparation and filling of the centrifugal cells (see below). When such buffer signal offsets are present, their inclusion in the model prevents errors in the protein sedimentation parameters, and often results in better fits with tighter confidence limits, despite the introduction of additional sedimentation parameters for the co-solvents. Despite this newfound ability to computationally compensate for meniscus and buffer mismatches, the best experimental practice is still to match them as rigorously as possible.

At very high co-solvent concentrations, whether or not they contribute to the signal directly, the sedimentation of the co-solvents will generate significant local changes in solvent density and viscosity that will affect, in turn, the sedimentation process of the protein of interest. For example, these concentration gradients can result in characteristic retardation of macromolecular sedimentation process with time, and even produce isopycnic bands of the protein in the density gradients after long times. It is possible to extend Eq. 6 to conditions of locally varying and time-dependent density and viscosity, and we have described a method how the linked sedimentation of co-solvent and macromolecule can be numerically simulated and fitted to experimental data (Schuck, 2004). Applications of this approach have been described in the characterization of protein pharmaceuticals (Gabrielson et al., 2009).

In cases where the protein itself is at a concentration exceeding the limit where hydrodynamic non-ideality is small (e.g., 1–2 mg/ml for globular proteins, lower limits for more elongated structures), terms can be added to the Lamm equation (Eq. 6) that describe the concentration-dependence of sedimentation, *s(w)*, and diffusion coefficients, *D(w)*, arising from the long-range hydrodynamic interactions (Dishon et al., 1967; Solovyova et al., 2001). SEDFIT, SEDPHAT, and SEDANAL provide for such models, which use linear approximation of the sedimentation coefficient with a non-ideality coefficient k_S. An example for the application of direct boundary modeling with a non-ideality model can be found in the study of malate dehydrogenase in complex solvents (Solovyova et al., 2001). These models can be extended to multiple species, where Johnston-Ogston effects (i.e. the expulsion of slower sedimenting species from the region of the faster-sedimenting boundary at higher total concentration) are displayed. However, even though modeling of such systems is possible, SV data cannot be expected to provide sufficient information for the reliable determination of different non-ideality coefficients and mutual interaction coefficients necessary in systems of multi-species and multi-component sedimentation with hydrodynamic interactions. In this regard, direct boundary modeling of non-ideal solutions is currently still limited. Furthermore, the linear approximation of the concentration-dependent sedimentation coefficient breaks down at higher than 5% total macromolecular volume exclusion (e.g. ~ 70 mg/ml protein). This fact renders impossible the full, quantitative interpretation of the sedimentation boundaries measured under crowded conditions.

Whichever model is used for the macromolecular sedimentation concentration profiles *c(r,t)* of the proteins of interest, in the direct-boundary model, the experimental data are fitted by least-squares with terms of the form

$$a(r,t) \cong c(r,t,\{\ldots\}_{protein}) + b(r) + \beta(t) + b_{buff}\ \left(r,t,\{\ldots\}_{buff}\right) \quad \text{Equation 7}$$

, where the brackets emphasize the non-linear parameters that depend on the precise model of the sedimenting species, as well as non-linear parameters that depend on the solution geometry. In these optimizations it is of critical importance to monitor the residuals of the fit. As illustrated in (Dam and Schuck, 2004), fits with incorrect models may sometimes

roughly describe the boundary data, yet provide qualitatively inaccurate answers if the boundary broadening from unrecognized heterogeneity is modeled as though originating from diffusion. To inspect the residuals, in addition to an overlay of the residuals to all scans, we have introduced a residuals bitmap that displays the radius- and time-dependent residuals in the form of a greyscale picture (Dam and Schuck, 2004). An example for an unacceptable fit to the data of Figure 4A using a single-species model and an acceptable fit with a distribution model is shown in Figure 4B and 4C, respectively.

Usually, fits within the noise of data acquisition and without systematicity should be expected from SV analyses. It has proven useful to compare models using criteria from F-statistics (Johnson and Straume, 1994; Johnson, 1992).

Direct Boundary Modeling With Size Distributions—A key development for the direct-boundary modeling of SV was the explicit, constructive description of polydisperse systems as a size distribution. SV data are highly sensitive to polydispersity and even trace impurities due to the strongly size-dependent hydrodynamic separation. Without equipping the model with appropriate terms to describe the effect of polydispersity, in most cases no satisfactory fit will be achieved. The basic approach comprises the description of the ensemble of sedimenting molecules as a differential distribution *c(s)* that describes the concentration of species with sedimentation coefficients between *s* and *s+ds* (Schuck, 2000)

$$a(r,t) \cong \int_{s_{\min}}^{s_{\max}} c(s)\chi_1(s, D(s), r, t)ds \quad \text{Equation 8}$$

; for greater clarity, from here onwards we omit the baseline, noise, and buffer terms of Eq. 7.

Diffusion takes place with a $\sqrt{t}$ dependence, but separation from differential sedimentation occurs with a e^{w^2st} time dependence. Thus, the description of the diffusion process and the sedimentation coefficient distribution usually do not strongly correlate. The solution to *D(s)* exploits some knowledge of the sedimenting macromolecules, often in the form of a hydrodynamic scaling law for certain classes of particles. Sometimes, different rules for calculating *D*(*s*) are applied in different segments of the distribution, dependent on *s* (see below). This generality and adaptability is possible due to the numerical evaluation of the Fredholm integral equation Eq. 8, which exploits techniques borrowed from modern image analysis, with details outlined elsewhere (Schuck, 2000, 2009).

Data fitting to a size-distribution of diffusionally broadened sedimentation profiles deconvolutes diffusion effects from the sedimentation boundary, much like removing blurring from point-spread functions in image processing. An example of the gain in resolution of *c(s)* when compared to a *ls-g*(s) 'data transformation'* that reflects largely the boundary shape is shown in Figure 3. In addition to the gain in resolution, because the *c(s)* model can usually describe the complete set of experimental data, it can report with exquisite sensitivity the presence of trace aggregates. With detection limits of 1% or better for oligomers of antibodies, this approach was found to be highly useful in biotechnology applications (Berkowitz, 2006; Gabrielson and Arthur, 2011). Another feature of the fit of Eq. 8 is the ability to include scans from the entire time-range of sedimentation from all species, without requirement that sedimentation boundaries of all species present be visible at the same time. This advantage results in a very wide (often 100–1000-fold) range of *s*-values that can be covered in the *c(s)* distribution from a single experiment.

For folded proteins, typically the most productive scaling law is that for compact particles, where *D(s)* is governed by an average frictional ratio $f/f_{0,w}$ for all particles (in a given segment) (Schuck, 2000; Schuck et al., 2002). The value of $f/f_{0,w}$ is usually refined as part of the non-linear regression. This model exploits the fact that actual shapes must vary quite significantly in order to produce a different frictional ratio, and many different shapes can have the same frictional ratio. Usually considered a downside of hydrodynamics for the assessment of protein shapes, the low shape resolution turns out to be advantageous in the present case. Furthermore, as shown in (Schuck et al., 2002), deviations of the best-fit frictional ratio from the actual frictional ratio will mainly result in broadening of the *c(s)* peaks. Due to the high resolution, robustness, and sensitivity of this approach, the vast majority of the > 1000 applications of *c(s)* reported in the literature for folded proteins have used this *c(s)* method with a refined average frictional ratio as scaling law *D(s)*. In this manner, this analysis has become the first tool used for the understanding of experimental SV data of unknown samples, providing a stepping stone towards the appropriate data analysis and improved experimental design.

Once sedimentation and best-fit frictional ratios have been assigned, each pair of *s*, and *D* can be inserted in the Svedberg equation Eq. 5 to produce a molar mass distribution *c(M)*. Provided that correct parameter values for solvent density and viscosity and for the protein partial specific volume are available, the precision of molar masses reported from single *c(M)* peaks is ~ 5–10%.

If there are multiple boundaries formed, then multiple *c(s)* segments can be constructed with a different frictional ratio attributed to each boundary. Usually, not more than one piece of diffusion information can be extracted in a well-conditioned manner from a single sedimentation boundary (see below). For example, for data showing two sedimentation boundaries, a model with a bimodal frictional ratio $f/f_{0,w}$ can be used, and molar masses of isolated peaks corresponding to each sedimentation boundary may be obtained.

Other scaling-law *c(s)* variants available in SEDFIT or SEDPHAT are for worm-like chains (applicable, for example, to amyloid fibrils (MacRaild et al., 2003)), arbitrarily user-defined scaling laws (applicable to some unfolded biological macromolecules (Harding et al., 2011), and proteins with ligand-induced conformational changes. The latter is a variation of the segmented distribution: for a certain *s*-range, the diffusion coefficient *D(s)* is calculated by the Svedberg equation (Eq. 5) based on a constant molar mass. Thus, this strategy reflects different possible hydrodynamic friction coefficients exhibited by the same protein (Schuck et al., 2000). For protein complexes that are known to have a certain frictional ratio, for example, after imaging of globular or spherical particles by electron microscopy, it is also possible to phrase the unknown extent of diffusion *D(s)* in terms of a known frictional ratio but unknown partial-specific volume, which can be refined in the SV boundary analysis. In this way, partial-specific volumes can be evaluated, for example, for samples of detergent-solubilized membrane proteins (Ebel, 2011), or for nano-particles.

Another noteworthy special case is the limit of $f/f_{0,w}$ being infinite, appropriate for non-diffusing species (Schuck and Rossmanith, 2000). In this case, the resulting *c(s)* distribution is referred to as *ls-g*(s)* due to its relationship with the previously derived apparent sedimentation coefficient distribution *g(s*)* from the time-derivative approach (Stafford, 1992), which was popular in the 1990s. In principle, the latter is theoretically also based on – though not practically applicable to – non-diffusing particles. Many nano-particles, large proteins, or other macromolecular complexes can be considered ‘non-diffusing’ species for the purpose of SV. This approximation is valid because, during the short experimental time provided by their rapid sedimentation, the extent of diffusional boundary broadening is negligible (which may be the case even though the boundary shows broadening from

differential sedimentation of polydisperse mixtures). The application of *ls-g*(s)* allows the sedimentation coefficient distribution of such "non-diffusing" species to be derived. When applied to diffusing species, a misfit of *ls-g*(s)* arises, which can be removed by restricting the data to a subset of the scans, similar to the *g(s*)* method. In this case, due to the lack of diffusional deconvolution, a diffusion-broadened sedimentation coefficient distribution is produced, reflecting the raw boundary shapes in the dimension of apparent *s*-values (hence the attribute '*g*'). Like the previous *g(s*)* method, one could model the spread of apparent sedimentation coefficients to estimate the molar mass of a discrete species. However, much more accurate and robust estimates can be achieved using the *c(s)/c(M)* method to evaluate diffusion information directly, accounting at the same time for polydispersity (Brown et al., 2009). In contrast to the previous *g(s*)* from the time-derivative method (Stafford, 1992), however, *ls-g*(s)* does not suffer from the same artificial broadening and peak shifts as reported for *g(s*)* caused by the problematic approximation of the time-derivative by finite time differences (Schuck and Rossmanith, 2000). Thus, *ls-g*(s)* has been used often when diffusional deconvolution was not desired. For example, *ls-g**(*s*) has been applied to large, essentially 'non-diffusing' particles and/or to distributions of particles with a very wide range of sedimentation coefficients.

Importantly, like *c(s)*, the *ls-g*(s)* method has generally no limitation in the shape, time-interval or total time-range of scans and can incorporate boundaries that are not simultaneously visible in the observed radial range. In comparison with the time-derivative method common in the 1990s, the *ls-g**(*s*) approach eliminates practical constraints that impose the selection of relatively low rotor speeds in SV, while maintaining a 1000-fold dynamic range of sedimentation coefficients in a single run. Regarding experimental design, this possibility enables the use of the highest possible rotor speed (50,000 or 60,000 rpm as limited by the rotor) to achieve the best possible hydrodynamic separation in SV.

Information Content and Regularization—One practically important aspect of the direct boundary modeling with size distributions, well-known in all fields employing Fredholm integral equations of the type Eq. 8, is that the solution is ill-conditioned and requires regularization. When Eq. 8 is solved strictly as stated, typically the absolute best-fit distributions consist of series of sharp, baseline-separated spikes, the number and location of which is dependent on very small features in the data, such as details of the noise. Although it may seem tempting to take this best-fit distribution as if reflecting a true size distribution, these spikes are highly unreliable and do not reflect the true information content of the data. In fact, a broad set of such spiky distributions usually fit the data almost equally well. From these properties of the inverse problem arises the requirement for regularization. Regularization is the selection, from all the distributions that fit the data statistically indistinguishably, of one particular solution that has the least information content. As described in detail by Provencher (Provencher, 1982), by minimizing the information content of the distributions and adjusting the detail of the distribution to what is reliably extracted from the experimental data, misleading details of the peak structure are avoided and the risk for over-interpretation is minimized. This is typically accomplished by a secondary optimization that quantifies the level of detail in the distribution. The exact form can be chosen dependent on the system under study. To achieve that, the error surface of the fit is skewed by a regularization penalty term and minimized, while the magnitude of the penalty term is adjusted such that the difference in the quality of fit to the raw data of the penalized solution does not diminish relative to the un-penalized overall best-fit by more than a permissible factor governed by F-statistics on a given, user-determined confidence level (e.g., P = 0.683). This produces a statistically acceptable fit that conforms as much as possible to the parsimony expressed in the regularization terms. (Conversely, it will be appreciated that, by this design of regularization, the original spiky best-fit solution cannot

provide a statistically better fit, and that therefore any information from a naive, unregularized *c(s)* solution will present details that are misleading.)

There is a large body of literature on regularization, as it is an important and well-studied problem in data analysis in many fields of physical science. Regularization methods often employed in biophysics are Tikhonov regularization and maximum-entropy regularization. As implemented in SEDFIT and SEDPHAT, Tikhonov regularization uses a measure of the total curvature $-\int(c''(s))^2\,ds$ to favor distribution curves that have as few peaks as possible, and as broad peaks as consistent with the experimental data. This is appropriate, for example, for macromolecules with a broad intrinsic size distribution, for example, polymeric products, or heavily glycosylated proteins with an extensive and polydisperse degree of glycosylation. Maximum-entropy regularization, also implemented in SEDFIT, minimizes the information content by maximizing the Shannon-Jaynes entropy $-\int c(s)\ln(c(s)/p(s))ds$, where *p(s)* is a default distribution usually taken to be constant (see below). This model is particularly attractive for systems that likely have a few intrinsically discrete species, such as purified protein samples. This is the default choice in SEDFIT. Because maximum entropy regularization has a tendency to produce oscillations for broader distributions (Schuck et al., 2002), in such cases the user should switch to the Tikhonov method. No regularization option is currently available for the *c(s)* implementation in ULTRASCAN.

In practice, for typical SV data with a high signal/noise ratio, both Tikhonov and maximum entropy regularization will result in very similar distributions. For data with a low signal/noise ratio, some distinct differences in the peak shapes and widths might be observed. The type and extent of regularization can be a dominating factor for the peak widths in the *c(s)* method. When considering fine details of the distribution, it is generally useful to probe the family of distributions that fit the data equally well and ask whether or not a certain feature is imposed by the data or by the regularization. This can be achieved by alternating the regularization method and the regularization level (P-value).

An additional, highly useful tool is the customization of the regularization to what is already known or to what can be hypothesized about the sample. This modification of the regularization can be achieved by imposing Bayesian prior distributions. For example, a non-constant prior distribution (*p(s)*, above) can be employed (Brown et al., 2007b). SEDFIT and SEDPHAT allow these priors to be constructed in variety of ways. It is often useful to use a *c(s)* distribution obtained from one sample as a prior hypothesis for the expected distribution of another sample. For example, this can be applied to a dilution series to determine whether the sedimentation coefficient distribution of a more dilute sample is consistent with that of a higher concentration. In this manner, one can address the problem observed at the lower concentrations, namely that the decreased signal/noise level results in broader peaks with the default regularization. For instance, this approach can better address the question of whether an oligomer dissociates upon dilution. Also, one can hypothesize that a certain peak may be described by a single species by using the numerical equivalent of a Dirac δ-function as a prior hypothesis (Figure 4F). If microheterogeneity contributes to a *c(s)* peak, then the resulting distribution, termed $c^{(p\delta)}(s)$, will not conform to the sharpness of the δ-function. (Due to its wide utility when studying purified proteins, this analysis can be conducted by pressing the keyboard shortcut control-X after obtaining the *c*(*s*) distributions.) The δ-function approach can help to assess the purity of a preparation, and/or to adjust the regularization of the *c(s)* distribution to utilize prior knowledge, as may be obtained from mass spectrometry. Alternatively, when obtaining slightly bimodal, not fully separated *c(s)* distributions in studies of species with similar *s*-values, we may use Gaussians placed at the centroids of each partial peak to try to baseline-separate the signals from both species. This can help in the quantification of subpopulations. *Vice versa*, we may test

whether a single, broader peak would also be consistent with the data, or if the data have statistically unambiguous information that the peaks can be sub-divided.

The way in which the Bayesian approach utilizes this additional knowledge is fundamentally different from hard-wired constraints. Due to the imperfections of the experimental data alone (originating from both data acquisition and sample), a hard-wired constraint that strictly imposes a certain model to the analysis will in most cases lead to a statistically significantly worse fits, which often then requires further judgment regarding whether it should be rejected or is still acceptable. In contrast, the Bayesian approach always fully honors the complete information of the data and, by design, ensures the same quality of fit, by allowing the result to refute the constraint if it is wrong, or by subtly adding on missing pieces in the interpretation. Notably, the Bayesian approach described above does not provide a quantitative measure of the agreement with the prior; rather, that is assessed graphically by the user.

In summary, SV analysis has progressed from a simple 'data transform' to a size-distribution analysis that fully utilizes all the data and invites the utilization of our certain knowledge of the system, such as the validity of the laws of diffusion and hydrodynamic scaling laws, to arrive at higher-resolution results. The Bayesian analysis represents a highly useful further step that now lets us actively probe the consistency of different interpretations with the experimental data.

Multi-dimensional generalizations of c(s)—Several useful generalizations of the *c(s)* are available that make use of additional data or distribution dimensions.

First, it is possible to generalize the sedimentation coefficient distribution to a size-and-shape distribution $c(s,f/f_0)$ (Brown and Schuck, 2006) that allows for a distribution of hydrodynamic friction values for each single *s*-value, thereby abandoning the need for a hydrodynamic scaling law:

$$a(r,t) \cong \int\int c(s,f/f_0)\chi_1(s,D(s,f/f_0),r,t)\,ds\,df\!f_0 \quad \text{Equation 9}$$

An example is shown in Figure 4. The size-and-shape distribution may equivalently be presented in *s*-*D* or *s*-R_S coordinates, although it is easier to create an efficiently discretized mesh in *s*-f/f_0 coordinates. Computationally, this problem can be solved exactly, in the same rigorous way as the standard *c(s)* distribution available in SEDFIT. A fundamentally different approach, called '2DSA', with a rather unique computational strategy was proposed by Demeler & colleagues (Brookes et al., 2009) to achieve a size-and-shape distribution in ULTRASCAN, as described in *CPPS Unit* 7.13 (2010). However, we are unaware of attempts to prove the correctness of the algorithm, and 2DSA remains empirical. In fact, upon detailed examination (Schuck, 2009), several key points of this method were found to be in conflict with mathematics and incompatible with general scientific principles of data analysis, and the method is therefore unlikely to give correct results.

Ordinarily, SV data do not contain sufficient information to create well-defined peaks in the diffusion dimension except for the major peaks (Brown and Schuck, 2006), thus providing very similar information as *c(s)*. In fact, by integration of the diffusion dimension, $c(s,f/f_0)$ can be collapsed back to a sedimentation coefficient distribution, called '*general c(s,*)*', that is independent of scaling laws but typically is virtually identical to the standard *c(s)*. However, in applications where scaling laws cannot be easily phrased or where species with very different molar masses co-sediment at the same *s*-value, the $c(s,f/f_0)$ and its related

distributions may be very useful. An example of the former case occurs in ensembles of particles with density distributions (Carney et al., 2011).

Another extension of *c(s)* is for the global modeling of SV data sets acquired in experiments at different solvent densities (Brown et al., 2011)

$$a_r(r,t) \cong \alpha_\rho \int c(s_{20,\omega})\chi_1(s_{20,\omega}, D_{20,\omega}(s_{20,\omega}), r, t)ds_{20,\omega} \quad for\ all\ \rho \quad \text{Equation 10}$$

In units of sedimentation and diffusion parameters at standard conditions (*20,w*), all experiments share the same sedimentation coefficient distribution, which will be mapped to the measured experimental conditions based on the experimental solvent density and viscosity. A scaling factor a_ρ accounts for the experimental fact that not all samples may be precisely at the same concentration, for example, due to slight dilution errors. This allows us to treat the protein partial-specific volume as an unknown parameter to be refined globally in the non-linear regression, and thereby provides a method for the determination of the partial-specific volumes, for example, of proteins or protein-detergent complexes.

This density-contrast SV approach improves on the precision of the classical Edelstein-Schachman SE approach due to the significantly higher precision of sedimentation coefficients, and due to the allowance for protein heterogeneity and impurities (Brown et al., 2011).

Undoubtedly the most powerful extension of the *c(s)* method for multi-component mixtures is the global analysis of SV data acquired in the same experiment simultaneously with different optical signals λ (Balbo et al., 2005):

$$a_\lambda(r,t) \cong \sum_{k=1}^{k\leq\Lambda} \varepsilon_{\lambda k} d \int c_k(s)\chi_1(s, D(s), r, t)ds \quad for\ all\ \lambda{=}1 \ldots \Lambda \quad \text{Equation 11}$$

The signals can be absorbance data at different wavelengths and/or a combination of interference and absorbance optical signals. If a total number of Λ different signals are available, then up to the same number of different macromolecular components can be spectrally discriminated based on their molar extinction (or signal) coefficient $\varepsilon_{\lambda k}$, and separate sedimentation coefficient distributions $c_k(s)$ can be calculated for each component *k* (Figure 5). (The remaining required quantity in Eq. 11 is the optical pathlength, *d*, which will be 1.2 cm for standard double sector centerpieces.) As shown in (Balbo et al., 2005), the spectral and hydrodynamic resolution can exhibit synergy, and the information content of the set of component $c_k(s)$ distributions can be far beyond that of separate standard *c(s)* analyses of the individual signals.

This multi-signal SV (MSSV) technique is naturally of great utility in the study of protein interactions, where the stoichiometry of protein complexes can be deduced from the area ratio of co-localized $c_k(s)$ peaks. MSSV has found many applications, as recently reviewed in (Padrick and Brautigam, 2011; Padrick et al., 2010). In the current version of SEDPHAT, up to three different signals may be included in the analysis, thus enabling the study of co-existing binary and ternary complexes formed in mixtures of three different proteins. Another important application is the analysis of protein-detergent mixtures, where this approach can reveal the protein/detergent ratio of the detergent-solubilized protein (Salvay et al., 2007).

For this multi-signal SV approach to work, sufficiently different spectral signatures of the interacting components are required. Padrick & Brautigam (Padrick and Brautigam, 2011) have determined a predictor, D_{norm}, that is calculated based on the known extinction coefficients and allows the prediction of whether different components are likely to be distinguishable in a two- or three-component MSSV analysis (see below). Often, interacting proteins have sufficiently different fractions of aromatic amino acids for the combination of interference signals with absorbance at 280 and/or 250 nm to generate a sufficiently high D_{norm}. The same is true for many protein interactions with nucleic acids (Berke and Modis, 2012), lipids or detergents (Ebel, 2011), or strongly glycosylated proteins. Thus, MSSV can often be conducted label-free (Balbo et al., 2005). Examples for the analysis of MSSV to mixtures of three protein components can be found in (Houtman et al., 2006; Barda-Saad et al., 2010). Detailed instructions for the application of MSSV can be found in the web-based SEDPHAT help system and in (Padrick et al., 2010). Although D_{norm} is based on simple matrix manipulations, some may find it non-trivial to calculate. As a courtesy, SEDPHAT can calculate and present this quantity to allow for experimental planning.

A current limitation is that the absence of hyper- or hypochromicity is assumed. Furthermore, it should be noted that the integrals of $c_k(s)$ peaks of different components foremost reflects the composition of the sedimentation boundary. For these quantities to reflect the complex stoichiometry, the complex must be *quasi* stable, which is the case close to saturation or for slow dissociation kinetics, both of which can be tested via concentration-independence of the *c(s)* peaks (see below). An exception can be the reaction boundaries of rapidly interacting molecules with very similar sedimentation coefficient in their free form (see below).

Sedimentation Velocity Analysis of Interacting Systems

Direct boundary modeling with a set of coupled Lamm equations for reacting systems—A key application of SV in protein science is the study of interacting systems. In recent years, it has become possible to use solutions of coupled sets of Lamm equations of chemically reacting systems directly for the non-linear regression of experimental SV data. This takes the form (Fujita, 1975)

$$\frac{\partial \chi_i}{\partial t} + \frac{1}{r}\frac{\partial}{\partial r}\left[r\left(s_i \omega^2 r \chi_i - D_i \frac{\partial \chi_i}{\partial r} \right) \right] = q_i \quad \text{Equation 12}$$

where i enumerates the species participating in the interaction with their signal contribution $\varepsilon_i dc_i(r, t)$, and q_i is the local chemical reaction flux between the species. For example, for a simple bimolecular reaction forming a 1:1 complex, q_i becomes $q_1 = q_2 = -q_3 = -q$ with $q = k_{on}\chi_1\chi_2 - k_{off}\chi_3$, and the equilibrium dissociation constant $K_d = k_{off}/k_{on}$. Because the sedimentation behavior and fractional saturation strongly depend on the molar loading concentrations, it is generally highly advisable to run multiple experiments at concentrations that bracket K_d by a factor 10, if possible, on either side. After a reaction scheme has been determined, both SEDPHAT and SEDANAL are capable of globally modeling all experimental data with Eq. 12, SEDANAL in the data differencing mode (Correia and Stafford, 2009) and SEDPHAT in the direct boundary modeling mode (Dam et al., 2005). Applications and strategies have been reviewed recently by Brautigam (Brautigam, 2011).

In principle, the Lamm-equation modeling for an interacting system allows kinetic rate constants to be extracted, as this parameter modulates the shape of the boundaries and their separation. However, as shown in (Dam et al., 2005), and not surprisingly, the sensitivity of the data for different kinetic rate constants is limited roughly to the time-scale on which the

SV experiment takes place, which is dependent on protein size, but typically restricts the range of k_{off} to around 10^{-4}/sec – 10^{-3}/sec.

Furthermore, Eq. 12 is intrinsically a discrete model, and may not always yield a good fit and/or accurate results if sample imperfections such as microheterogeneity and trace impurities are present that will affect the boundary broadening. These deviations from single-species behavior are most critical for the kinetic rate constants, as the information on their parameter values also resides in the boundary broadening. For heterogeneous associations between proteins that do not self-associate, sufficient purity can be tested by the applicability of single discrete-species models to the individual components. The impact of aggregate impurities may be reduced by restricting the radial range of the analysis, such as in partial boundary modeling (Brown et al., 2009). Unfortunately, the influence of breakdown products or impurities sedimenting slower than the species of interest cannot be excluded in the same way, as their diffusionally broadened boundary often makes significant contributions to the sedimentation data across the whole available radial range and typically overlaps the boundary of interest, especially at early times.

Although very powerful, the direct-boundary modeling with Lamm equations of reacting systems is usually not the first method of choice, and more robust methods that are more consistent with the sensitivity of SV for sample imperfections will be highly desirable, especially if kinetic rate constants are not of primary importance. One could consider it a virtue of the Lamm-equation modeling of reacting systems that all peculiarities of the coupled reaction/diffusion/sedimentation process are delegated to the Lamm equation solver and modeling. However, for successfully designing and interpreting SV studies of heterogeneous interacting systems, it is essential to understand the salient features of reaction boundaries.

Effective particle model for the sedimentation of heterogeneous interacting systems—For rapidly interacting systems (on the time-scale of sedimentation, i.e. usually $k_{off} > 10^{-2}$/sec), not all sedimentation boundaries reflect real sedimenting species. To examine this phenomenon, let us consider at first a simple 1:1 interaction of the type A + B forming an AB complex in a rapid reaction, with the nomenclature such that $s_A < s_B < s_{AB}$. It has been long known that sedimentation boundaries of rapidly interacting proteins exhibit several seemingly non-intuitive properties (Fujita, 1975): (1) The data show not three but at most two boundaries, one always sedimenting at the *s*-value of either free A or free B, termed undisturbed boundary, and the other at a composition-dependent *s*-value between s_B and s_{AB}, termed reaction boundary (2) In a titration series, the *s*-value of the undisturbed boundary could switch from s_A to s_B, but at a transition point where the solution composition does *not* correspond to the complex stoichiometry. (3) At a certain solution composition, also *not* equal to the complex stoichiometry, there can be only a single boundary. It is crucially important not to mistake the reaction boundary for that of a stable, independently sedimenting species.

These properties can be intuitively better understood in the framework of the recently developed effective particle theory (EPT) for the coupled sedimentation/reaction process (Schuck, 2010b), which was derived directly from Eq. 8 in the approximation of non-diffusing particles and rectangular geometry. Due to the instability of the complex in rapid equilibrium, obviously no separate boundary can stably exist for the complex. Instead, the whole system with components of free A, free B, and complex AB, as determined by mass action law, sediments jointly. Thus, in the reaction boundary, there is always a mixture of free A, free B, and complex. There can be only one undisturbed boundary comprised of the excess material of one of the binding partners (since the presence of the other would lead to complex formation, hence a reaction boundary). From the ergodicity of the sedimentation

process we can conclude that, in the reaction boundary, the relative populations of free and complexed A and B, respectively, determine their constituent time-average sedimentation coefficients s_A^- and s_B^-, respectively. A fundamental, intuitive insight is that for the material in the reaction boundary these time-average sedimentation coefficient experienced by molecules A and molecules B must match, $s_A^- = s_B^-$. From this directly follows the s-value of the reaction boundary, $s_{A\cdots B}$, its composition, $R_{A\cdots B}$, as well as the nature and magnitude of the undisturbed boundary (Schuck, 2010b):

$$s_{A\cdots B} = \begin{cases} (c_A s_A + c_A c_B K s_{AB})/(c_A + c_A c_B K) & \text{for} \quad c_{Btot} > c^*_{Btot}(c_{Atot}) \\ (c_B s_B + c_A c_B K s_{AB})/(c_B + c_A c_B K) & \text{else} \end{cases}$$

$$R_{A\cdots B} = \begin{cases} 1-(s_B - s_A)/K c_B (s_{AB} - s_B) & \text{for} \quad c_{Btot} > c^*_{Btot}(c_{Atot}) \\ 1-(1+K c_A (s_{AB} - s_A)/(s_B - s_A))^{-1} & \text{else} \end{cases} \qquad \text{Equation 13}$$

$$c^*_{Btot}(c_{Atot}) = c_{Atot} + \frac{(s_B - s_A)}{2K(s_{AB} - s_B)} \left(1 + \sqrt{1 + \frac{4 c_{Atot} K (s_{AB} - s_B)}{(s_{AB} - s_A)}}\right)$$

These relationships are illustrated in Figure 6. For example, in our 1:1 binding scheme, it follows that there always must be less free A in the reaction boundary than free B, due to the lower *s*-value of free A lowering the time-average of all A more than free B lowers the time-average of all B. Therefore, at equimolar total concentrations, the excess material constituting the undisturbed boundary must always be A. It will always require a certain excess of total B over total A in solution, termed $c^*_{Btot}(c_{Atot})$, to allow s_A^- to exceed that of the free B species s_B. This condition is necessary for B to produce the undisturbed boundary, because the reaction boundary must always be faster than free B. This $c^*_{Btot}(c_{Atot})$ is the phase transition point where there is no undisturbed boundary, and at $c_{Btot} > c^*_{Btot}(c_{Atot})$ we have excess B comprising the undisturbed boundary, and at $c_{Btot} < c^*_{Btot}(c_{Atot})$ A must be in the undisturbed boundary. Because the free species' sedimentation coefficients are different, the phase transition of the coupled system is asymmetric.

There are several practical implications of these considerations.

1. To conceptually understand the mechanism of reaction/sedimentation processes: When applying SV to interacting systems that form complexes that can dissociate and re-associate during the sedimentation process, the time-average sedimentation coefficient experienced by all components is crucial for determining the sedimentation properties. This mode of sedimentation is fundamentally different from the sedimentation of mixtures of stable, discrete species. Movies displaying schematically the mechanism of propagation of the reaction boundaries at different concentrations and sedimentation coefficients can be created for user-defined parameters in SEDPHAT. Careful consideration of EPT may prevent naïve misinterpretations, for example, the false assumption that the molar ratio at the transition between undisturbed boundaries will simply reflect the complex stoichiometry.
2. For experimental design: Because the salient features of the boundary structure are expressed in simple terms in Eq. 13, it is possible for any given system with expected K_d and estimated *s*-values to predict the *s*-value and amplitudes of sedimentation boundaries for the entire range of possible total loading concentrations of all components. For several interaction schemes, an experimental design tool is available in SEDPHAT that surveys the boundary properties similar to Figure 6. In addition, it allows one to consider practical constraints in the available stock sample concentrations, enter minimum detection limits, and decide on a general strategy for an experimental concentration series that brackets the

most significant and informative changes in the sedimentation behavior. SEDPHAT will translate the graphical user input and propose a detailed sample mixture and dilution plan. If possible, it will approximate the user input to create a titration series, dilution series, or Job series, generating additional constraints that may be useful in the data analysis (see below).

3. For quantitative analysis: Experimental data can be analyzed and average *s*-values and amplitudes of the undisturbed and the reaction boundaries can be easily determined, for example, using *c(s)* followed by peak integration. These data can be assembled into isotherms as a function of total loading concentrations and subjected to quantitative analysis with the equations of Eq. 13. This approach exploits the clearly visible boundary structure without the need to resort to a detailed interpretation of the boundary shapes. Thus, the EPT-based isotherm analysis is a robust approach that can contribute significant information. For example, the isotherm of the weighted-average *s*-value s_w of the interacting system remains far below the complex *s*-value s_{AB} for all concentrations except when the molar loading concentration ratio matches and saturates the complex stoichiometry. In contrast, the *s*-value of the reaction boundary will always be close to s_{AB} when the molar ratio of any component is in excess of the complex stoichiometry. Furthermore, the boundary pattern reveals very clear information on the location of the phase transition, which is closely related to s_{AB} and the equilibrium constant (Eq. 13). These features allow the binding parameters as well as the binding stoichiometry to be better determined. The accuracy of this approach has been demonstrated and it has already been successfully applied to several experimental systems (Zhi et al., 2010; Schuck, 2010b; Zhao et al., 2011a; Berke and Modis, 2012).

In practice, such an isotherm analysis can be accomplished by creating an ASCII text file with a table that contains two columns for micromolar concentrations of A and B, followed in a third column, for example, by the signal-weighted *s*-value of the reaction boundary. Optionally, this format can be extended by incorporation upper and lower confidence limits for each *s*-value, which represent individual error bars to be considered in the analysis. This file can then be loaded into SEDPHAT as an SV isotherm data type. After specification of ancillary information such as signal coefficients, optical pathlengths, isotherm type, these data can be included in the global analysis. When fitting the user-selected global model for the interaction, the appropriate model function for each specific data type is applied, and the model is mapped accordingly into the data space of each isotherm (or other) data set.

Application of c(s) differential sedimentation coefficient distributions—Even though the sedimentation process of rapidly reacting systems is very different from that of noninteracting mixtures, it is still possible to apply the same differential sedimentation coefficient distributions to their SV data. In particular, it has been widely observed that the diffusional deconvolution in *c(s)* works equally well as in the application to non-interacting mixtures. In fact, it was shown recently, with the help of EPT for the approximation of gradients and diffusion fluxes, that despite the different mechanism of propagation of reaction boundaries, their diffusion proceeds normally in a very good approximation (Schuck, 2010a). However, like the *s*-values, the apparent diffusion coefficients reflect the interacting system and must be interpreted in this way.

The theoretical expressions for the isotherms of *D* and *M* (via insertion of *s* and *D* into the equation Svedberg) of the ‘effective particle’ as a function of total loading concentration, in analogy of Eq. 13, are presented in (Schuck, 2010a). The *M*-value associated with the reaction boundary is intermediary to that of A, B, and AB, but *not* equal to the population-average or weight-average molar mass of the system. Because it is a particular strength of

the EPT analysis not to require interpretation of the diffusional spread and precise boundary shape, as that feature is most sensitive to sample imperfections, the quantitative analysis of the isotherm of *D* or *M* has not been suggested.

The application of *c(s)* to reaction boundaries is highly useful for the qualitative discrimination whether or not there is an interaction. For this end, experiments should be conducted over as wide a range of concentrations as possible, and it is useful (though not mandatory) to adopt a dilution or titration series. For a dilution series, if the *c(s)* curves obtained at different protein concentrations exhibit invariant peak positions and invariant peak ratios, then there is no interaction. If the peak positions are invariant but their ratio changes, an interaction with slow dissociation on the time-scale of sedimentation is indicated. If the peak positions depend on concentration, the presence of a reaction boundary from a rapidly equilibrating system is demonstrated. Similar conclusions can be drawn from a titration series, obviously allowing for the titrant peak to grow with concentration. These considerations should also account for the fact that a standard *c(s)* analysis will exhibit lower resolution at lower signal/noise ratios, which can be counter-acted by using Bayesian prior distributions, as indicated above.

A second, very important application of *c(s)* to interacting systems, fast or slow, is the ability to determine signal-weighted-average sedimentation coefficients of the system. Based on the second moment method (Schachman, 1959), the weighted-average sedimentation coefficient can be rigorously defined based on the total macromolecular flux through an imaginary plane in the solution plateau at a radius r_p *rp*

$$s_w = -\frac{1}{\omega^2 r_p^2 c_p}\frac{d}{dt}\int_{r_m}^{r_p} c(r,t)\,r\,dr \quad \text{Equation 14}$$

This measure is independent of reaction kinetics and diffusion. It is solely a function of solution composition and the individual species' *s*-values. The right-hand side (rhs) of Eq. 14 prescribes a method for determining s_w from the change in the integral over the observed concentration distribution. Any direct-boundary model that is faithful to this integral will produce the same s_w, independent of the physical motivation or model parameters. For this property, it is sufficient to ensure that the model produces a good fit to the data. In particular, this applies to models from *c(s)*, which typically lead to excellent fits of the raw data. Therefore, we can extend Eq. 14 as

$$s_w = -\frac{1}{\omega^2 r_p^2 c_p}\frac{d}{dt}\int_{r_m}^{r_p} c(r,t)\,r\,dr = \int_{s_{\min}}^{s_{\max}} c(s)\,s\,ds \Big/ \int_{s_{\min}}^{s_{\max}} c(s)\,ds \quad \text{Equation 15}$$

There are some fine points regarding the time-dependence of the integrals in Eqs. 14 and 15, its relation to reaction kinetics, and the weighting of scans from different time-points that enter the calculation of *c(s)*, but these factors have a negligible effect on the analysis of binding isotherms (Schuck, 2003). Although Eq. 14 is strictly applicable only to conditions producing true solution plateaus, the generalization on the rhs of Eq. 15 is straightforward (Schuck, 2003). For accurate determination of s_w, it is therefore essential that the quality of fit be monitored, verifying that the model is faithful to the integrals under the concentration profile. It should be noted that the van Holde-Weischet analysis, which was a useful qualitative tool at the time of pre-computer SV analysis (Demeler et al., 1997; Schuck et al., 2002) (but is now obsolete), is unsuitable for this purpose because it abandons the information on absolute boundary height (Schuck, 2003). When calculating *g(s*)* based on the time-derivative method, even though the integrals on the rhs of Eq. 15 can easily be

formed and their identity with s_w is sometimes taken for granted, the relationship of *g(s*)* with the integral on the lhs of Eq. 15 is more tenuous. This uncertainty is due to the presence of numerical distortions (see above) and the fact that the quality of fit to the raw data is not tested. However, a tool in SEDFIT allows to back-transform *g(s*)* into the original data space as a direct boundary model, allowing one to assess the faithfulness to the integral of the concentration profiles (Schuck and Rossmanith, 2000).

The assembly of s_w-values as a function of solution composition allows an isotherm analysis for the interacting system, independent of the reaction kinetics. Similar to the EPT analysis described above, the s_w isotherm can be loaded as an ASCII file into SEDPHAT, where an analysis with many different binding models can be carried out, either in single-isotherm analysis mode or in a global analysis with other isotherms. The latter is accomplished simply by loading multiple isotherm files side-by-side into SEDPHAT. When multiple signals are available, the separate determination of signal weighted-average s_w-values and EPT-based boundary structure isotherms from the different signals and their global analysis is extremely useful.

Overall strategy for the SV analysis of interacting systems—In summary, the analysis of interacting systems by SV usually consists of several main steps. For previously well-characterized samples, the experiments may be performed side-by-side.

1. **Assuming a heterogeneous association, characterize A and B with regard to purity and possible self-association:** For this purpose, run as wide a concentration series as possible, and inspect the resulting *c(s)* distributions. The concentrations used need to bracket, or at least include, those used for the study of the interaction. Acquire data at all signals that can be expected to allow spectral discrimination of A and B. A concentration-independent peak position (except where hydrodynamic interactions would occur, e.g. > 1–2 mg/ml) belies the presence of self-association. Determine the average properties of the material in the main peak (control-M in SEDFIT), and double-check the molar mass estimate to be within 5–10% of the known value. Compare the signal in the peak with the total sedimenting signal to determine purity.
2. **Assemble a hypothesis for the interaction with regard to the stoichiometry and binding affinity and plan the binding experiment.** Either from structural and functional prior knowledge about the system under investigation or from initial exploratory SV-AUC experiments of mixtures at different molar ratios, formulate a hypothesis of the binding model. In SEDPHAT, select the hypothesized binding model with known and hypothesized parameters, and open the effective particle explorer tool. Familiarize yourself with the expected boundary structure and the overall s_w as a function of loading concentrations, similar to the plots in Figure 6. Enter the available stock concentrations, signal coefficients, and sketch a trajectory in the feasible part of the parameter space that encompasses significant changes in the expected observables. SEDPHAT will propose a protocol with log-spaced concentrations along this trajectory (or optionally along a nearby trajectory that generates additional constraints), with detailed dilution instructions for sample preparation. For centrifugal run conditions, one typically will pick the highest possible rotor speed for optimal hydrodynamic separation (50,000 rpm in an 8 hole rotor). Set up data acquisition at all ABS wavelengths and/or IF, dependent on the potential spectral discrimination: Calculate D_{norm} by entering the extinction coefficient matrix in the corresponding calculator tool of SEDPHAT.
3. **Experimentally determine c(s) on the mixtures and qualitatively determine reaction kinetics and reaction scheme.** Create a normalized superposition of c(s)

traces, for example in GUSSI, which is the plotting companion program to SEDFIT. Discern whether peaks exhibit a characteristic concentration-dependent shift in relative populations and in peak *s*-value. This will allow for a classification of the reaction kinetics. For slow reactions, information on the stoichiometry can be assessed from the *s*-value (and attendant *M*-value) of the complex peak. If A and B are spectrally distinguishable, determine the complex stoichiometry by MSSV analysis of the data under saturating or kinetically stable conditions.

4. **Determine the binding affinity and complex s-value(s).** In the *c(s)* analyses of all mixtures, determine the undisturbed and reaction boundary, or all species boundaries, dependent on the reaction kinetics. Perform an overall integration of *c(s)* in SEDFIT to determine the overall s_w, for all signals. Enter these in a table in an ASCII file, e.g. created using Microsoft NOTEPAD or GUSSI version 1.0.2 or higher, and load these s_w isotherms in SEDPHAT. Similarly, determine the amplitudes and *s*-values of all the boundaries of the sedimenting interacting system. Depending on the kinetics of the interaction, add to the global SEDPHAT window isotherms from species signal amplitudes, or the isotherms described by EPT, respectively. (For statistical refinement, determine error bars for all isotherm values.) If a reaction scheme is already identified, select the corresponding model and fit globally all isotherms. Perform a statistical error analysis on the refined parameter values. If multiple models seem conceivably consistent with the data, sequentially pick the different models and compare the overall quality of fit. Utilize constraints derived from hydrodynamic scaling laws for the species *s*-values, considering lowest and highest possible shape asymmetry.

5. **If possible, conduct a direct-boundary model with Lamm equation solutions of the reacting system.** Provided the individual species can be reasonably well described as discrete species with the correct molecular weight, then a global direct-boundary model with coupled Lamm equations may be conducted. Load all raw data into SEDPHAT, representing the entire sedimentation process, from all the cells and all the signals. In the local parameters of each cell, allow for the meniscus to float, and enter appropriate signal coefficients. If possible, add constraints linking some of the loading concentrations of different experiments (see below). Conduct a global fit, at first using parameter values fixed to those obtained in the isotherm model of Step 3, at a later stage after partial convergence relaxing these constraints. For unknown kinetics, it can be useful to initialize $log_{10}(k_{off})$ at an intermediate value of −3.5. If a good global boundary model is achieved, calculate the error projection maps of all parameters of interest. If a satisfactory global boundary model cannot be found, resort to the isotherm model in Step 3, and calculate error estimates using the projection method.

Obviously, this description cannot be more than a general guide. The individual steps may need to be supplemented and/or adapted to address specific properties of the system at hand. What is described is the common denominator of the workflow that we have developed and applied successfully to many interacting systems in our laboratories through the last several years.

Sedimentation Equilibrium Analysis of Interacting Systems

As in the comments on SV above, we will not re-introduce exhaustively the principles and practice of SE. Instead, we restrict the presentation here to new developments in the decade since the last Unit on SE (CPPS Unit 20.3 (1999)), and just recapitulate the basic theoretical framework for nomenclature.

While SV studies the evolution of the sedimentation process with time, SE examines the equilibrium state. In the limit of infinite time, a single ideal species at an initial loading concentration $c = \bar{c}(r,t = 0)$ in a sector-shaped solution column extending from the meniscus *m* to the bottom *b* will slowly redistribute to assume a Boltzmann distribution

$$\begin{aligned} c(r) &= c(b)\exp\ (M_b\xi_\omega(r)) \\ \xi_\omega(r) &= (r^2 - b^2)\omega^2/2RT \\ c(b) &= \bar{c}\frac{M_b\xi_\omega(m)}{\exp\ (M_b\xi_\omega(m)) - 1} \end{aligned} \quad \text{Equation 16}$$

, with a curvature determined by the rotor speed and the buoyant molar mass M_b, which is a function of its molar mass *M* and density increment $(d\rho/dw)_{T,\mu}$, $M_b = M\,(d\rho/dw)_{T,\mu}$ (with *w* referring to the weight concentration of the protein). In dilute solution with simple buffers, the latter can be approximated very well with the buoyancy factor given by the partial-specific volume, such that $M_b = M\,(1 - v\bar{r})$. Often Eq. 16 is presented without the reference to the initial loading concentration, in which case the concentration *c(b)* at the bottom of the solution column (or any other arbitrarily selected reference radius) will serve as a pre-exponential factor. It should be noted that the relationship between the loading concentration and the amplitudes of the pre-exponential term will be significantly different from the form of Eq. 16 if the column is not sector-shaped. For example, if standard six-channel centerpieces or for external loading six-channel centerpieces are used instead of the sector-shaped double-sector centerpieces, the correct pre-exponential amplitudes are larger because less material is in the highest concentration regions close to the bottom of the cell, and they have to be calculated by numerical integration considering the precise geometry. For this reason, the exact type of centerpieces used in the SE experiment should be recorded and entered in SEDPHAT.

For interacting systems, mass action law will hold between the local species concentrations throughout the solution column. For example, in a system where A binds B to form an AB complex with an association equilibrium constant K_a, the concentration of complex will be $c_{AB}(r) = c_A(r) \times c_B(r) \times K_a$, which directly leads to the expected behavior that the radial distribution of the complex is likewise a Boltzmann exponential, which will be governed by a buoyant molar mass that is the sum of the individual species' buoyant molar masses:

$$\begin{aligned} c_{AB}(r) &= c_A(r)c_B(r)K_a \\ M_{b,AB} &= M_{b,A} + M_{b,B} \end{aligned} \quad \text{Equation 17}$$

Some deviations from Eq. 17 are possible in the presence of repulsive interactions, which usually are not encountered below concentrations of a few mg/ml, except for highly charged proteins at low ionic strength.

Also, corrections are required if binding or conformational changes in the complex causes significant changes in the hydration properties of the proteins affecting the buoyancy. This would lead to pressure-dependent, and thereby rotor-speed-dependent, behavior. This phenomenon though possible, is rarely observed (Harrington and Kegeles, 1973), and therefore not discussed further.

For beginners, it may be confusing that the buoyant molar masses are products of the buoyancy factor and true molar mass, such that a given buoyant molar mass can arbitrarily be expressed as apparent molar mass M^* on any given v^* scale, as long as the product produces the correct buoyant molar mass, $M_b = M^*(1 - v^*\bar{\rho})$. For studying heterogeneous interactions, it is highly convenient to express both components' molar masses on the same

v^* scale. If this v^* is arbitrarily taken as that of one of the components, then the apparent molar mass of the other component will be different from the true molar mass, but all components' molar masses will remain additive towards the complex molar mass. This technicality is inconsequential, and only concerns the units of measuring buoyant molar mass, which is often not the desired refined parameter of such an analysis. (Otherwise, for each different complex a different complex partial specific volume, determined by the weight-average of its components would have to be considered. This seems simpler, at first, but turns more complicated in other models when the complex molar mass is an adjustable parameter.) This exercise is required, as SEDPHAT separates the true molar mass and macromolecular partial specific volume; an alternative rigorous workaround is to set v^* to zero and let the buoyant molar mass represent the parameter for the molar mass.

A key difficulty in SE analyses, especially of interacting systems, is the well-known ill-posed nature of the decomposition of noisy exponentials. Different from most other fields of biophysics where this problem occurs, the exponents (molar masses) are pre-determined in SE, which allows us to distinguish several components, provided we have sufficient solution column length (> 3 mm) and data with sufficient curvature. For arriving at statistically well-defined binding constants, many data sets are obtained at multiple loading concentrations and rotor speeds, and – if possible – multiple signals. These sets must be included in a global analysis. The ability to perform this decomposition routinely, with the help of various software packages, has developed over the last two decades. Unfortunately, the global modeling of SE with N = 10 – 20 different profiles introduces 2N unknown concentration parameters as pre-exponential factors for two-component mixtures. This difficulty creates a complex error surface notorious for exhibiting local minima and that is very difficult to optimize.

Two strategies have been introduced to address this central problem. They both revolve around the idea of mass conservation (MC), i.e. that the total amount of material for each component that was initially inserted in the experiment should also be found in the equilibrium profile. Because this has become a key tool to modern SE analysis, we have emphasized MC in Eq. 16 by introducing the notion of an effective loading concentration $\bar{c} = V^{-1}\int c(r)dV$, which corresponds to the average concentration we would achieve by shaking up the contents of the material in the equilibrium gradient and can be compared with the true known loading concentration c_0. Importantly, when using MC for interacting systems this relationship becomes slightly more complicated with the new requirement $\bar{c_A} + \bar{c_{AB}} = c_{Atot,0}$ and $\bar{c_B} + \bar{c_{AB}} = c_{Btot,0}$, which can be numerically easily solved. (It should be noted that mass action law will *not* hold between the volume-averaged concentrations, i.e. $\bar{c_{AB}} \neq \bar{c_A}\bar{c_B}K_a$.)

The first strategy to use these MC relationships merely incorporated them as numerical tools to aid in the optimization of the error surface to find the best fit, but not necessarily to enforce these constraints once the putative best-fit has been found (Philo, 2000). This has been termed 'soft MC'. This is the form implemented, for example, in the program HETEROANALYSIS, with the caveat that mass is only calculated up to the fitting limits and the amount of material between the highest observed radius and the bottom of the solution column is ignored (HeteroAnalysis Contents, 2012). Unfortunately, due to the strong concentration gradient and increasing concentrations towards the bottom of the cell, a very significant amount of material may remain unaccounted for. Thus, although it may be useful as flexible bounds to intermittently guide the optimization algorithm, a true account of the mass of material for each component loaded into the centrifugal cell is not accomplished.

The second strategy is to use MC, accounting correctly for all material, as a true constraint. Historically, what has prevented this approach from entering routine analysis was the

difficulty in assigning the bottom position of the solution column, which is obscured by optical artifacts. (This is in contrast to SV, where the meniscus position is the key geometric parameter for the analysis of the sedimentation process; in SE the bottom position is more important as the highest solution concentrations are close to the bottom.) In 1976, Roark (Roark, 1976) proposed a method based on the inclusion of SE profiles at multiple rotor speeds, which was applied recently by Vistica et al. (Vistica et al., 2004) as a method to solve the problem of the unknown bottom position of the solution column. It is based on the assumption that the total amount of material in solution is the same at all rotor speeds. Because significant redistribution of material occurs at the different rotor speeds, MC can be used to calculate the bottom position of the solution column, which is shared by all SE profiles from the same cell (accounting for trivial corrections due to rotor stretching). Accordingly, the bottom position of the solution column does not need to be assigned *a priori*, but can remain an adjustable fitting parameter. Due to this flexibility, this method was also characterized ‘soft’, but in a different sense, and we propose it should better be referred to as the ‘implicit MC’ method.

In practice, SE analysis with implicit MC works best with a series of three rotor speeds that span the conditions of typical low-speed SE (Richards et al., 1968) and high-speed meniscus depletion SE (Yphantis, 1964). This has the additional benefit of providing the potential to determine the baseline parameters, including radius-dependent features (Vistica et al., 2004). Dependent on the molar masses of interest in the experiment, an algorithm for the prediction of optimal rotor speeds is implemented as a tool in SEDFIT (see below).

It is also important to recognize that our absolute knowledge of protein concentrations prior to the AUC experiment is invariably not precise enough to be used as hard constraint. Therefore, the total concentrations should be allowed to refine in the analysis within bounds. However, it is possible to create the samples in a way that produces safe constraints. For example, in a titration series with careful pipetting, the concentration of one component can be held sufficiently constant in different samples, justifying the addition of a constraint forcing them to be identical. Similarly, dilution series of stock mixtures are certain to have the same molar ratio of total material of both components. Using these approaches the total number of concentration unknowns can be reduced to just a few. Further constraints between data sets can often be established, for example, between the solution geometry of data sets from the same physical cell but different detection wavelength.

The implicit MC method not only stabilizes the optimization algorithm, but considerably tightens the error intervals of the unknown parameters. It also extends the range of binding constants that can be examined, as well as the general applicability of SE analysis. For example, just based on the curvature of the SE profiles, co-existing species are impossible to reliably discriminate if their molar masses differ by less than ~20–30%. This leads to requirements on the mass ratio of proteins whose interactions can be studied to be less than a factor 3 – 5, unless the components have significant spectral differences that can be exploited in a multi-signal analysis. The implicit MC approach overcomes these limitations, and can be applied even for binding partners that cannot be detected.

The SE analysis with this method is carried out in SEDPHAT using data sets of the type ‘multi-speed SE’, which are groups of SE profiles from the same cell, recorded with the same signal, at different rotor speeds. This implicit MC method has proven to be extremely useful in practice, as judged by many studies published in the literature using this approach (Ghirlando, 2011).

Characterizing Detergent-Solubilized Membrane Proteins

At the time of the pioneering work of Tanford & Reynolds on the use of AUC for the study of protein/detergent complexes, advanced SV methods were unavailable. SV has led to new approaches that have proven extremely useful and surpassed the resolution and reliability of SE in the study of protein-detergent complexes. In recent years this potential was systematically explored and applied in many systems by the laboratory of Christine Ebel (Ebel, 2011). Because the basic relationships have already been described by Fleming (CPPS Unit 7.12 (2008)), we will just briefly highlight the new tools and refer to the specialized literature: (1) The buoyant molar mass can now be reliably extracted from SV experiments, which is especially profitable in the context of distribution models that allow one to observe and exclude the effects from sample heterogeneity. When interpreting the *s*-value in terms of a Stokes radius, all contributions from bound detergent and lipid must be accounted for. (2) The extension of SV to density-contrast experiments (Ebel, 2011; Brown et al., 2011) can allow one to determine the average partial-specific volume of the protein/detergent/lipid complex more reliably than SE in the case of preparations that do not lead to single species. In either case, H-D exchange must be accounted for when using deuterated water. This allows the study of detergent systems that cannot be readily density matched. Although combinations of different oligomeric states, complex frictional ratios, and binding coefficients can sometimes describe SV data equally well for SV experiments at a single density, this ambiguity is resolved in experiments at multiple densities. An ‘Ebel-plot’ of binding coefficients *vs* partial specific volume can allow the identification and elimination of hypothetical oligomeric protein states that imply un-realistic solvation (Ebel, 2011; Nury et al., 2008). (3) The composition of the protein/detergent/lipid complex can be estimated in an orthogonal approach using multi-signal SV, exploiting the fact that UV absorbance of most detergents is smaller than that of the proteins under study (Ebel, 2011; Salvay et al., 2007). While these methods can improve technical hurdles in the determination of the oligomeric state of membrane proteins, of course they do not address more fundamental questions of detergents affecting the oligomeric state (Dorwart et al., 2010). Nanodiscs are a promising new tool in the study of integral membrane proteins in a lipid membrane environment, and their use in analytical ultracentrifugation studies as recently been described in detail (Inagaki et al., 2012).

EXPERIMENTAL DESIGN AND PROTOCOLS

In the following, we comment mostly on the experimental design features and protocols that have changed as a result of the new theoretical and data analysis approaches. For a more complete and recent systematic discussion of experimental details, the reader is referred to *Current Protocols in Immunology* 18.8 (2007) and 18.15 (2008), as well as the internet resources listed below.

SV or SE?—The developments described have now led to the general acceptance of SV as the primary method of choice for the AUC characterization of non-interacting single and polydisperse systems, as well as interacting systems. The simple, yet exquisitely profound *c(s)* analysis can be diagnostic in characterizing the solution behavior of the system of interest, with the added advantage that the SV experimental setup, data collection, and analysis can be successfully completed within a day or two. The use of an eight-hole rotor allows the simultaneous study of seven samples – in the case of an interacting system or self-association, these seven samples can easily cover nearly a 100-fold concentration range, allowing the construction of a binding isotherm, as recently demonstrated for the dimerization of enzyme I from *E. coli* (Schwieters et al., 2010). Another advantage in the use of SV is that the initial *c(s)* analysis only describes the diffusion-deconvoluted behavior of the sedimenting boundary; no specific model is required for its interpretation, thus

making it an unbiased reporter of the solute's behavior. The one disadvantage of SV over SE is that it requires much more material, especially when characterizing an interacting system. Even though this is usually not a problem when AUC is used to study systems being analyzed by either NMR or X-ray crystallography, the issue may prove to be more challenging for biologically interesting systems not amenable to overexpression using recombinant technologies. Despite this, given the sensitivity of SE to impurities, it is highly recommended to characterize samples destined for SE studies initially by SV in order to confirm purity. In this respect, SV is well suited for determining the presence of impurities.

In the case of interacting systems, both SV and SE have their significant advantages – whereas SV can report the stoichiometry or lifetime of the complexes, SE provides no kinetic information which can be particularly advantageous when the *c(s)* distributions describing the free and reaction boundary are unintuitive. Unlike certain instances of SV, however, the model best describing an interacting system may not be immediately obvious from SE data. In terms of a detailed comparison of SV and SE methods for a self-associating system with known binding mode, it may be relevant that the s_w analysis in SV may be conducted independent of the molecular buoyancy. In an extensive series of AUC experiments of glutamate receptor variants covering ~5 orders of magnitude of K_d conducted side-by-side in SE and SV mode, binding constants were found to be in agreement within a factor of two, with the exception of high-affinity binding (Zhao et al., 2012).

For very high-affinity and very low-affinity binding, degradation during the long experimental time in SE can become a significant problem, in which case SV will be more reliable (Zhao et al., 2012). For high-affinity binding, if the total loading concentration of the material is significantly above the true K_d, even trace amounts of degradation products with smaller molar mass can lead to artificially elevated best-fit concentrations of free monomer, thus artificially increasing the best-fit estimate of K_d, as examined in detail in (Zhao et al., 2012). Additionally the formation of irreversible aggregates may pose a similar problem exacerbated in systems at low-affinity binding. The global analysis of SE and SV in interacting systems is provided for in SEDPHAT.

Concentrations—For SV analysis, the lower limit of useful concentrations provides signals slightly below the noise of the data acquisition. Dependent on the detection used and the signal increment, this may be in the range of 1 μg/ml. For example, isotherm analyses of a 50 kg/mol protein with high-affinity dimerization have taken advantage of ABS detection at concentrations as low as 10 nM (Zhao et al., 2012). However, in the absence of carrier protein it is not always predictable if surface adsorption to the windows or centerpiece depletes the concentration of protein remaining in solution. For SE, due to the formation of a concentration gradient that in meniscus-depletion configuration approaches zero concentration independent of loading concentration, it is generally not necessary to load concentrations at the detection limit. The highest concentration is usually constrained by the available concentration, or by the hydrodynamic or thermodynamic nonideality interactions that often start to set in at concentrations in excess of a few mg/ml.

In order to facilitate the experimental design of rapidly interacting system, we implemented several practical tools in SEDPHAT. Designing series concentrations of samples for SV is often time consuming due to the required comprehensive consideration of multiple factors such as extinction coefficient, dynamic range of signals from optical detection systems, binding constant, etc. The Effective Particle Explorer function in SEDPHAT takes all the constraints into account and can generate a complete and practical series of concentrations for the user (Zhao et al., 2011a).

With regard to the relationship between signal and concentration, it has long been assumed that IF data can be analyzed on the basis of a constant, protein-independent refractive index increment. However, this is not always very accurate, since the protein's refractive index increment does significantly depend on its amino acid composition. The deviations from a 'consensus' value decrease with increasing protein size, but especially peptides < 10 kg/mol may exhibit significant deviations (Zhao et al., 2011b). For this reason, a calculator function has been implemented in SEDFIT that, side-by-side with the extinction coefficient and partial-specific volume, calculates the refractive index increments and a molar signal coefficient.

In choosing concentrations to be used in an MSSV experiment, additional concentration restraints are present: the interacting heterocomplex must be well populated and kinetically stable. These requirements are usually met by exploring high-concentration domains of concentration space. Of course, one must balance the need to populate the complex against the linearity of the signal from the mixture. These considerations are discussed in detail in (Padrick and Brautigam, 2011), and SEDPHAT's Effective Particle Explorer can be utilized to find optimal concentrations with these caveats in mind.

Buffers—Buffer matching of the reference and solvent chambers of the AUC cell is critical for the success of both SV and SE, particularly when using the IF optical detection system, as this reports on all of the components sedimenting in solution, including the buffer salts. Buffer matching is best achieved by equilibrium dialysis and use of the dialysis buffer as a reference. As the final step in protein purification usually involves size-exclusion chromatography, we have found that use of the actual buffer used for SEC also works well. If added to the purification protocol as a substitute for dialysis, it has the added benefit of further purification.

Meniscus matching is also critical for both SE and SV when using IF detection, less so for SV as meniscus and/or buffer mismatches can be modeled computationally in SEDFIT for a *c(s)* analysis (Zhao et al., 2010). The buffer/meniscus mismatch model can be accessed in SEDFIT by pressing Ctrl-B. In addition to buffer mismatches arising from different concentrations of the same buffer components in the sample or reference sectors (i.e. having the same sedimentation coefficient) the model can also account for situations in which a different buffer is present in the reference sector. Meniscus mismatches, which provide different starting points for buffer sedimentation, can also be treated in this model. We find that such buffer corrections become more important when using buffers having a high refractive index (e.g. 2M NaCl). In cases when sufficiently precise matching becomes infeasible, it may be prudent to eliminate buffer components from the reference cell altogether and fill it with water, and simply add to the SV sedimentation model an explicit term for the sedimentation of small co-solutes.

Partial Specfic Volume—The protein partial specific volume ($\bar{v}$) can be calculated from its amino acid composition. This prediction is based on the tabulated values of the partial specific volumes of amino acid residues, which change slightly with solution conditions. A SEDFIT calculator function has implemented the relevant calculation algorithm for predicting the $\bar{v}$ of proteins (Zhao et al., 2011b). The calculated values of proteins are extensively used in practice and the accuracy is generally satisfactory. However, experimental determination of $\bar{v}$ might be necessary in some circumstances such as proteins with post-translational modification and hydration measurements. Conventionally, the measurement of $\bar{v}$ is performed using H-D density contrast SE experiments (Edelstein and Schachman 1967). However, use of H_2O and D_2O introduces the issue of H-D exchange, which results in a variation of the $\bar{v}$ value. In order to precisely determine $\bar{v}$, H-D exchange

must be taken into account in analyzing density contrast SE data, for example, using direct global non-linear regression in SEDPHAT.

Further, in order to address the intrinsic high susceptibility to sample heterogeneity of SE density contrast measurements, we developed a strategy using SV with buffers at different densities by varying the concentration of H_2O and $H_2{}^{18}O$ (Brown et al., 2011). By globally modeling the sedimentation boundaries of the data with different buffer densities, the protein partial-specific volume and sedimentation coefficient distribution are optimized simultaneously. This approach has been demonstrated to allow precise experimental determination of ^{v} even with levels of sample heterogeneity that would prohibit the use of the classical Edelstein-Schachman approach (Edelstein et al., 1967).

Selecting Optical System and Setting Scan Parameters

Absorbance or interference data acquisition?

In general, the choice of optical system (and if absorbance optics are used, which wavelength to use) is sample-dependent. Because of its speed and high data-point density, the interference (IF) optical system is very desirable, but it requires scrupulous compositional balancing of the reference and sample buffers, or additional parameters for describing the sedimentation of the buffer components need to be introduced (see above). One experimental setting where IF optics are necessary is for a protein that requires a high concentration of a UV-absorbing chemical to retain stability or activity; as long as the UV-absorbing material can be balanced in the reference and sample sectors, IF could be used to follow the sedimentation of the protein whereas the absorbance (ABS) optics would not. ABS optics should be used if a non-absorbing buffer component cannot be balanced and the user does not wish to use the computational compensations described above. Because the ABS and IF optical systems offer complementary information, we find it useful to use both to collect AUC data if possible.

Since ABS data are usually acquired with quartz windows in the cell assembly, whereas IF data are collected with sapphire windows, the question arises regarding which windows to use for the joint acquisition of IF and ABS data. In our experience, both kinds of windows can work well. Quartz windows are less mechanically stable and can result in higher systematic TI noise, that sometimes is not completely stationary by undergoes small shifts. On the other hand, most sapphire windows have a slight absorption of light in the far UV causing higher noise in ABS detection. (Dependent on the supplier, the absorbance may vary considerably.) Balancing these two factors, since the elevated noise in the ABS system is uniform whereas shifting TI patterns cannot be modeled, the default choice is usually sapphire windows.

Furthermore, using the IF optics with SE experiments carries with it special requirements: the centrifugation cells must be mechanically “aged”, and water blanks should be collected (Ansevin et al., 1970). The former step is necessary to ensure that components of the cell do not shift when different centrifugal fields are applied to them, as is routinely practiced in SE experiments. The aging process is described elsewhere (Balbo et al., 2007). The water blanks allow the subtraction of the noise structure of the IF data from the acquired concentration profile. A water- or buffer-containing cell should be scanned at the experimental rotor speeds, and a tool is available in SEDFIT for subtracting the “water blank” when the data are being pre-processed. In practice, a similar correction can be effected by analyzing the data in multi-speed mode and subtracting TI noise features (and possibly accounting for rotor stretch) (Vistica et al., 2004).

For the absorbance optics, the choice of wavelength depends on the chemical nature of the sample and the sample’s concentration. Convenient wavelengths to study for proteins of

low, moderate, and high concentrations are 230 nm, 280 nm, and 250 nm, respectively. The first wavelength is measuring absorbance from the peptide bonds in the protein (this wavelength corresponds to an intensity maximum from the xenon flash lamp used in the centrifuge), the second is measuring absorbance from aromatic side chains, and the third is at a spectral minimum where the proximal tails of the first two absorbance features overlap. Importantly, many small molecules absorb light in the far-UV range, restricting what buffer components may be used (phosphate is a good choice for far-UV work, but TRIS and HEPES have significant absorbance there). For SE experiments, it is usually advantageous to collect data using at least two wavelengths; 280 nm and 250 nm are often the most useful. For most SV experiments, usually a single wavelength is chosen based on the expected signal. If DNA or RNA is under study, then the absorbance maximum at ca. 260 nm is a natural choice for data collection.

Data acquisition for multi-signal SV

Multi-signal SV (MSSV) experiments require careful considerations of the signal and wavelength choices. Up to three components can be analyzed in such experiments, but it should be noted that at least as many signals must be collected as components present. In what follows, we assume the simplest case: a two-component interacting system (A + B $\leftrightarrow$ AB). But the principles discussed are extensible to three-component systems. Before initiating an MSSV experiment, it is advisable to assess whether the proteins are spectrally distinguishable. In previous work (Padrick and Brautigam, 2011), we have established a numerical criterion (D_{norm}) that indicates whether the experimenter can reasonably expect a successful outcome from an MSSV experiment. For a two-component system in which the amino-acid compositions of both components are known and assuming the use of both ABS and IF,

$$D_{norm}=\frac{\left\|\left(\varepsilon_{IF}^{A}\varepsilon_{ABS}^{B}-\varepsilon_{IF}^{B}\varepsilon_{ABS}^{A}\right)\right\|}{\sqrt{\varepsilon_{IF}^{A\ 2}+\varepsilon_{ABS}^{A\ 2}}\sqrt{\varepsilon_{IF}^{B\ 2}+\varepsilon_{ABS}^{B\ 2}}}, \quad \text{Equation 18}$$

where the ε's are molar signal increments, their subscripts indicate the optical detection system, and the superscripts denote the component's identity. A D_{norm} 0.04 or higher is desirable for two-component MSSV experiments. Although we find the combination of IF and A_{280} to be particularly powerful for such experiments, other combinations of detection signals are available, and A_{280}/A_{250} can be employed successfully (Padrick et al., 2010). We recommend that the user determine D_{norm} of all possible dual combinations of A_{280}, A_{250}, and IF, and choose the combination that has the highest D_{norm}. The wavelengths 280 nm and 250 nm are significant because they usually represent a maximum and a minimum, respectively, in typical protein absorbance spectra. It is important to choose such features in the UV spectrum for MSSV because the precision of the Beckman absorbance optics is not sufficient to guarantee that all scans will be performed at exactly the same wavelength. Choosing maxima and minima therefore minimizes scan-to-scan variation in the detected absorbances, and makes the analysis tolerant with regard to 1 or 2 nm shifts of the data acquisition wavelength. Thus, one should be aware of the proteins' spectral properties, and if they deviate from the expectations described above, the detection wavelengths should be adjusted accordingly. If one of the proteins has a prosthetic chromophore with a sufficiently high molar extinction coefficient, its absorbance maximum (usually in the visual part of the spectrum) can be chosen. Likewise, if one of the proteins is labeled with an absorbing moiety (e.g. a fluorophore, see Hellman et al. 2011), its absorbance maximum can be utilized. The latter two cases should afford excellent spectral discrimination if only one of the proteins is thus modified (Padrick et al., 2011).

Absorbance data acquisition in intensity mode

We have shown that negative logarithm of raw intensity data acquired with the ABS system for each a single sector of a double sector cell also has the same signal structure that can be modeled with a combination of TI and RI noise, thus allowing the use of intensity data for SV experiments (Kar et al., 2000). In this manner, SV intensity data can be collected on a maximum of 14 rather than 7 samples. There are, however, a number of technical limitations which should be borne in mind by the experienced user. SV absorbance data are usually collected in a 'continuous' mode – prior to data collection, the scanner records the intensity of the reference sector at 6.5 cm and adjusts the PMT voltage for this single scan. For this reason, it is usually advisable to avoid menisci or other optical artifacts at this radial position. Furthermore, it is essential to only have low absorbance (< 0.5 AU) solutions in the reference sector when collecting data in intensity mode. The use of highly absorbing solutions in the reference sector will result in high PMT voltages that may bring about PMT saturation at solute-free regions in the sample sector. As this mode of detection is not supported by Beckman Coulter, we advise that the reader approach such apparent time-saving measures with due caution. We further note that SE absorbance data are usually collected in a 'step' mode – in this mode the PMT voltage for the reference and sample sectors is set for each radial measurement based on the reference intensity.

Post-centrifugal fractionation and tracer detection

Much greater freedom in the detection method is achieved if the sedimentation process takes place in preparative centrifuges and is recorded after post-centrifugal fractionation (Bothwell et al., 1978; Darawshe et al., 1993). Proteins can be tagged using radioactive or chromophorically labels, or may have its concentration determined using an immunoassay. In the case of nucleic acids, quantitative polymerase chain reaction (qPCR) can be utilized. Interestingly, qPCR was implemented successfully to study the physical properties of native chromatin fragments through a combination of 'tracer' band sedimentation and isopycnic ultracentrifugation on preparative ultracentrifuges (Ghirlando et al., 2004; Ghirlando and Felsenfeld, 2008). In recent years, tracer SE experiments after post-centrifugal fractionation has been used, in particular, to study the behavior of labeled components under non-ideal conditions and crowded conditions, as pioneered by Allen Minton and colleagues (Darawshe et al., 1993; Rivas et al., 1999). For a recent review of theory and applications, see (Rivas and Minton, 2011).

Centrifuge Run and Scanning Parameters—For SV data collection, once the rotor is placed into the centrifuge, the chamber is evacuated, and the temperature of the rotor arrives at the experimental temperature after some time. For 20° C experiments, we routinely observe a 1-hour equilibration period after the rotor has reached this temperature before commencing centrifugation. For experiments at 4° C, this waiting period is longer, i.e. between 8–16 hours. The temperature equilibration should be conducted while the rotor is at rest. Previous recommendations to equilibrate temperature at 3,000 rpm were based on less-detailed data analyses than currently in use. The rotor is then accelerated from 0 rpm to the experimental rotor speed. In general, for the maximal hydrodynamic resolution, we recommend the highest speed possible. The Beckman An60-Ti and An50-Ti rotors are rated for 60,000 and 50,000 rpm, respectively. For most proteins, 50,000 rpm is suitable, but for large proteins or complexes (> ca. 200–300 kg/mol), a slower rotor speed is necessary to ensure time to collect adequate data before the sample has completely sedimented. In such a case, we recommend that the user experiment with various rotor speeds prior to starting the run by simulating the centrifugation using the "Generate" option in SEDFIT. This command allows the user to input sample parameters and rotor speeds, and it simulates the experiment, allowing the user to visually assess whether the inputted parameters will lead to a fruitful

centrifugation run. It should be kept in mind that as few as 20 scans reporting boundary movements throughout the cell are sufficient for the analysis.

For SE experiments, no temperature equilibration prior to centrifugation is required, as this end will be accomplished simultaneously with the establishment of the mechanical and thermodynamic equilibrium. However, the question remains of which rotor speeds to use. As described above, for a well-conditioned analysis utilizing implicit mass constraints (Vistica et al., 2004), we have found it useful to consider at least three rotor speeds: one with a shallow concentration gradient with information on high-mass species, one with a steeper gradient with more information on smaller species, and one with significant meniscus depletion that assists in the determination of baseline offsets and has information on the smallest species. Calculator functions exist in SEDFIT to estimate these speeds ("Calculator → estimate equilibrium rotor speeds"). This calculator is appropriate for interacting systems because it queries the user for the highest and lowest expected mass values. The calculator also gives a rough estimate for the time of the system to reach equilibrium given the inputted parameters and assuming a temperature of 4° C. A separate calculator function ("Calculator → minimum time to equilibrium") is also available and can consider different experimental conditions to estimate the time to reach equilibrium (van Holde and Baldwin, 1958).

For SV experiments, the data should be collected quickly. This does not represent a difficulty for the IF optical system, because it only takes a few seconds to acquire a scan from a given cell. If IF data alone are being collected, it is appropriate to program a small (1–5 min.) wait time between scans so that the maximum scan number (999) is not exceeded before centrifugation is complete. However, the ABS system, which monitors light intensity as a function of radius by using a stepper motor to move a slit over a photomultiplier tube, is significantly slower. For this reason, we set the scanning parameters for ABS optics to collect the data as quickly as practicable without compromising the data quality. We find using the "continuous" collection mode with a step size of 0.003 cm and 1 replicate strikes a good balance between speed and data quality. Also, the data-collection method is modified such that there is no wait time between scans. For SE data taken with the ABS system, there is ample time to collect the scans, so we use the stepping mode with 0.001 cm steps and 10–20 replicates. We generally wait 4–6 hours between scans.

DATA ANALYSIS

Considerations for the Analysis of SV data

For data analysis of SV data, all scans that have information on the sedimentation process should be loaded into SEDFIT or SEDPHAT. Inexperienced users often load too many scans; late scans in which no sedimentation is evidenced should be excluded. However, it is notable that the analytical methods used in these programs do not require one to restrict the data to a particular time domain. In practice, usually about 50–200 scans are input. Depending on the scan parameters used, the user may need to down-sample the scans, for example selecting only every 2nd available scan. The user graphically chooses initial estimates for the meniscus, bottom, and fitting-limit positions.

The most commonly used analysis performed on SV data in SEDFIT and SEDPHAT is the $c(s)$ analysis. Useful parameters to employ for this analysis of most protein samples include a resolution of 50–100, an s-range spanning the expected sedimentation coefficients, a starting frictional ratio of 1.2–1.4 (this parameter is set to be refined), and a regularization P-value of 0.68. This latter setting controls the regularization level used in the analysis (Schuck, 2000; Schuck et al., 2002), and a value of 0.68 strikes an excellent balance between the resolution of the distribution and the accuracy of its information content. For IF data, both TI and RI noise must be calculated to achieve accurate modeling of the SV data.

For ABS data, the RI noise decomposition is rarely appropriate, but the TI noise calculation is helpful to remove data features due to window imperfections. The meniscus position should always be refined, as it is probably not known by the user with sufficient accuracy. The bottom position is only refined when small molecules with significant back-diffusion are analyzed. The analysis is initiated by clicking “Fit” in the main menu of the program, during which an iterative optimization of the linear and nonlinear parameters of the system is performed. Once the fitting session has converged, the $c(s)$ distribution is displayed along with the corresponding data, fit lines, and residuals. The fit should be examined for defects. The rmsd should be low; less than 0.01 signal units under most experimental settings is adequate. The residuals should be nonsystematic. This desired feature can be assessed by examination of the residuals bitmap (introduced above), which should be free of strong horizontal, vertical, or diagonal features. Systematic residuals usually indicate that a parameter that should have been refined was actually fixed at a non-optimal value; two leading candidates are the frictional ratio and the sample meniscus position. If this is the case, the user should correct the problem and reinitiate the fitting session. Another possible source of systematicity is that the underlying assumptions of the $c(s)$ model are violated. In particular, the user should consider whether an interaction may be occurring (the $c(s)$ strictly distribution assumes noninteracting species and the presence of thermodynamic or hydrodynamic interactions can lead to increased residuals) or species with varying frictional ratios are present (the simple $c(s)$ distribution assumes a single frictional ratio for all species).

An examination of the final refined distribution may indicate a buffer mismatch. This defect, which is especially common in interference data, usually manifests as a sharp peak or upward feature at the low-s extreme of the $c(s)$ distribution. As mentioned above, recent advances (Zhao et al., 2010) have been introduced that allow the user to compensate for the mismatch computationally.

Also described above is the fact that the sedimentation coefficients of rapidly sedimenting species may be inaccurately refined in ABS data analyzed by these programs because the sedimenting boundary moves significantly during the finite scanning time (i.e. after the time stamp of the scan has been recorded). This discrepancy results in an overestimate of the s-value. Compensatory terms to the Lamm-equation calculation (accessed by “Options → Fitting Options → Lamm Equation Parameters” in SEDFIT and “Options → Lamm Equation Options” in SEDPHAT) can be added to ameliorate this error (Brown et al., 2009).

Many variations on the $c(s)$ distribution are available. For example, it may be known that two or more hydrodynamically resolved species have divergent frictional ratios; to obtain accurate estimates of the respective frictional ratios, it is possible to divide s-space into several segments, each having an independent frictional ratio. Carrying this idea further, it is feasible to calculate the frictional ratios for species that are not hydrodynamically resolved using the $c(s, f/f_0)$ distribution discussed above (Brown and Schuck, 2006). Further, it may be advantageous to model one region of s-space with discrete species and another with a continuous distribution (an example might be a sample with highly populated monomers (the discrete species) with aggregates (the continuous distribution)). These $c(s)$ offshoots are implemented in SEDFIT and SEDPHAT.

Following a $c(s)$ analysis of an interacting system, it is advisable to construct isotherms based on the results. Such isotherms can be analyzed to arrive at an equilibrium binding constant (or constants) for the system. For more details regarding isotherm construction, see “Considerations for the Analysis of Isotherms,” below.

For systems with no contaminants and only a few species evident in the *c*(*s*) distribution, it is possible that a more accurate description of the data will be obtained with a discrete-species analysis. As mentioned above, for interacting systems, it can be desirable to directly model the boundary with Lamm-equation solutions to the reacting system. Because this method of SV data analysis may take some time to converge, we recommend that an isotherm analysis on the system (see below) be performed first to establish excellent parametric estimates that can be entered into the Lamm-equation analysis. While the goal of perfect purity can be difficult to achieve for biological samples, SEDPHAT has features that can minimize the influence of these sample contaminants. For example a “non-participating” species can be introduced to compensate for a single contaminant. If aggregates are encountered, a “partial boundary model” can be employed, which restricts the analysis to only the relevant radial portion of the scans (Brown et al., 2009), eliminating contributions from aggregates. Many of the strategies employed for SE data analysis (see below) are also applicable for the Lamm-equation approach to interacting systems. The analysis should be as constrained as possible. For example, the individual components of a two-component system should be studied so that their sedimentation coefficients and molar masses may be treated as constants in the analysis of mixtures. If data from multiple signals were collected, this step also allows the refinement of extinction coefficients, in a manner equivalent to that of the MSSV experiment detailed below. These refined values would be entered and fixed in the global analysis of the mixtures. For these mixtures, concentrations should be studied that result in the measurement of adequate populations of the complex and individual components. For example, a titration series can be examined in which one component can be held at a constant concentration while those of the other are varied. In the global analysis (and by analogy to the SE data analysis presented below), the concentrations of the first component can be linked across all data sets, removing several parameters from the analysis. In general, component concentrations should be refined (i.e. not fixed). If the proper domains of concentration space are surveyed, then refined values for K_d and $s_{complex}$ can be reliably and precisely obtained. The value for k_{off} may also be refined, but, as mentioned above, it is only accurately estimated when it is in the 10^{-3}–10^{-4} s^{-1} range. After the parameters of interest have been refined, several criteria should be checked to establish the goodness of the fit. These include non-systematic residuals, low rmsd’s, modest drifts from input concentrations, and physically intuitive values of the parameters (see “Considerations for Isotherm Analysis”, below). The error intervals of the parameters of interest should be obtained using the methods described below (“General Statistical Methods”). Further details on this type of analysis are available in (Brautigam, 2011; Dam et al., 2005).

Another *c*(*s*)-based analysis enabled by SEDPHAT is the MSSV experiment. A detailed examination of this method is available elsewhere (Padrick et al., 2010), so we restrict ourselves here to a brief summary. Below, we will assume that the IF molar signal increments are known from the amino-acid compositions of the proteins (i.e. via SEDFIT’s calculator function). Knowing one of the signal increments well for each protein is a prerequisite of the experiment, because this increment will be used as a standard against which the signal increments for the other detection signal will be refined. In the experiment, one centrifugal cell will contain only A, one only B, and a third will contain both proteins, with one (usually the smaller) at a significant molar excess. This expedient helps to ensure adequate occupancy and kinetic stability of the expected AB complex. There are three steps taken to accomplish this analysis: (1) refinement of $\varepsilon_{\mathrm{ABS280}}$ for A, (2) refinement of $\varepsilon_{\mathrm{ABS280}}$ for B, and (3) the MSSV analysis of the mixture.

For step (1), the SV data sets for protein A (IF and A_{280}) are loaded into SEDPHAT, and a multi-signal analysis is carried out, holding ε^{A}_{IF} constant and allowing ε^{A}_{ABS280} to refine. In this analysis, the data are globally modeled by a single *c*(*s*) distribution. Of course,

variations on this strategy are available, including segmented *c*(*s*) distributions and hybrid discrete/continuous distributions. Once a satisfactory fit to the data is obtained, the configuration is saved and the refined ε^{A}_{ABS280} noted. In step (2), the above is repeated for protein B, resulting in a refined value for ε^{B}_{ABS280}. In step (3), the SV data sets for the mixture are loaded into SEDPHAT, and the known and refined signal increments are input. A global analysis is performed using two overlapping distributions—one accounting for species A ($c_A(s)$) and the other for B ($c_B(s)$). For co-sedimenting complexes, these distributions will have peaks at the same sedimentation coefficient (and presumably a sedimentation coefficient that is greater than that of either of the two species). The molar ratio of the species present in the co-sedimenting *s*-range can be obtained by integration of the distributions. The stoichiometry of the complex may be derived from this information and an estimate of the mass of the complex.

We suggest that four criteria be used at this point to assess the reliability of the molar-ratio estimate. They are:

1. **Rationality of the calculated molar ratio.** A 20:1 ratio of an AB complex probably is not "rational", whereas a 1:0.9 ratio is likely, within error, to indicate a 1:1 complex.
2. **Rationality of the distributions.** Detection of a small component alone at very high *s*-values or *vice versa* is not likely and probably indicates poor spectral resolution.
3. **Mass conservation.** The concentrations of the proteins in the mixed sample should be known, and facilitating this calculation are the concentrations available from the SV experiments performed with the samples alone. Integration over the entire *s*-range in the analysis of the mixture should comport with this expectation.
4. **Statistical tests of alternative explanations given the data.** For example, if a 1:1 complex is observed, can a 2:1 or 1:2 complex be ruled out? Our ability to use the statistical tools and segmented distributions in SEDPHAT can provide answers to this question. For example, a segment of *s*-space can be defined where the complex is observed, and a fit performed as above. The χ^2 of the best fit (called χ^2_b), is noted. Then, a critical $\chi^2(\chi^2_c)$ can be defined by using the "Statistics" menu of SEDPHAT ("Statistics → Critical chi-square for error surface projections"), choosing a confidence level of 0.95 (i.e. 2 σ), and confirming the defaults for degrees of freedom. The program will report χ^2_c. Then the *c*(*s*) segment with the complex can be constrained to have either 2:1 or 1:2 complexes. After repeating the fit with the constraint in place, the χ^2 is examined. If this value is higher than χ^2_c, we may safely reject the alternative hypothesis. If not, the data do not have enough discriminatory power to confidently reject likely alternative hypotheses.

Considerations for the Analysis of SE Data

A successful sedimentation equilibrium experiment, be it for the analysis of a single, non-interacting solute or the characterization of an interacting system, requires data collection at a minimum of three rotor speeds that report on all of the species present in solution. Furthermore, it is crucial to demonstrate that sedimentation equilibrium has been in fact attained. Typically, sedimentation data are collected at 6-hour intervals and compared using either WinMATCH or SEDFIT. Equilibrium is reached when differences between successive scans can be described as non-systematic random noise typical for the method of detection being used. The time to reach equilibrium depends on many factors, including the solution column height, the experimental temperature (which directly influences the solution

viscosity), and the asymmetry of the macromolecule. A typical solution column of 5 mm containing globular proteins at 20.0°C usually requires at least 48 hours to reach equilibrium; as a guide, SEDFIT also provides a calculator for estimating the minimum time to equilibrium. As the sample concentrates at the bottom of the AUC cell, multi-speed SE data are best collected starting at the lowest speed and incrementally raising the rotor speed once equilibrium has been reached. This approach has the added advantage that it typically takes less time to re-establish equilibrium when raising the rotor speed.

Having established equilibrium at all rotor speeds data are preprocessed in SEDFIT for export into SEDPHAT. Multi-speed data for each detection system are loaded into SEDFIT and sorted into equilibrium scans; in this manner, the attainment of equilibrium for each rotor speed is confirmed, and data are exported as experimental ('xp') files into SEDPHAT. Each experimental file contains the parameters used for data collection, specifically the solution density, viscosity and temperature as well as the effective partial specific volume and extinction coefficients of the macromolecules of interest. As each experimental file contains multi-speed data, it is possible to apply TI noise corrections to the data (see below).

The analysis of hetero-interacting systems requires characterization of the individual components by SE using the same conditions used to study the interaction. An analysis of the multi-wavelength SE data collected for each of the components can be conducted with a model termed 'A (single species of interacting system)'; this does not only provide a method for experimentally the determining partial specific volume (assuming that the macromolecule does not self-associate), but also a method for providing the effective extinction coefficients at the absorbance wavelengths of interest. Such a determination is carried out by fixing one of the values for the extinction coefficient (this process is analogous to that described above for MSSV). Typically, when IF data are collected, the molar signal increment for the protein calculated in SEDFIT is fixed and absorbance extinction coefficients at 280, 250 and/or 230 nm are experimentally determined through the implementation of implicit MC constraints. In cases when interference data are not available, then the extinction coefficient at 280 nm is fixed. This value, along with the interference molar signal increment can be calculated based on its amino-acid composition in SEDFIT. As noted above, the collection of SE IF data requires the use of mechanically aged cells and the subtraction of previously measured water blanks. The use of internally loaded 6-channel centerpieces is thus precluded. Also note that absorbance data for protein and protein-protein interactions are typically collected at 280, 250 and 230 nm. By choosing these absorbance wavelengths, one not only minimizes possible issues with the AUC monochromator but also extends the range of accessible protein concentrations. Data collection at these wavelengths in combination with the use of 12 or 3 mm pathlength cells allows for the characterization of high and low-affinity interactions.

The parameters obtained from the characterization of the individual components are used to initialize the analysis of the interacting system. SE data collected at multiple wavelengths and detection systems, and concentrations spanning the expected K_d, in the form of 'xp files', are assembled into SEDPHAT. Currently, SEDPHAT supports a maximum of 30 xp files, which is usually far more than sufficient. To implement the use of implicit MC, the loading concentrations of each of the components are entered, cells for which data are collected at multiple wavelengths are linked in terms of a single concentration(s). Furthermore, in the case of an experiment designed using a dilution series, the ratio of component B to A can be fixed or floated as a single global parameter, providing an example in which the experimental design can be used for the simplification and optimization of the data analysis. The implementation of MC also requires that the value of the cell bottom be refined during data fitting; in this case data collected for the same cell using the same detection system (e.g. absorbance data at multiple wavelengths) can be

linked in terms of a common meniscus and bottom. We do not recommend linking bottom radii for interference and absorbance data, as they utilize independent radial calibration procedures.

Having assembled all of the xp files and initial constraints, the data are modeled in terms of the appropriate interaction model. It is important at this stage not to implement the use of TI noise, as this may become correlated with the interaction model used. Fitting is carried out implementing the simplex, Marquardt-Levenberg and simulated-annealing fitting routines to establish a global minimum on the SE data error surface. The r.m.s.d for each xp file and the systematicity in the residuals are used as criteria to establish the goodness of fit and satisfaction with the model used to fit the data. In the case of an unsatisfactory fit, the model can be improved or further refined (e.g. is this an A + B ↔ AB or A + B + B ↔ ABB interaction, does one need to include positive or negative cooperativity in the case of the latter?). The optimized concentrations can also be used as a criterion for the goodness of fit and propriety of the model as these are usually close (within 10–15%, but dependent on the technique used to determine protein concentrations prior to AUC) to those expected. It is also essential to check that the optimized values for the cell bottom are not constrained by the limits placed on the parameter and/or totally unreasonable. Following the analysis and optimization, the best model that describes the experimental data is further refined by inclusion of TI noise for the absorbance as well as the interference data, assuming that the latter had the corresponding water blanks subtracted. The implementation of TI noise will result in a decrease of the r.m.s.d. and improve the quality of the residuals. Care should be taken to note that the TI noise should only account for time invariant systematic noise in the data it should be reasonably flat and show no curvature.

The optimized fit should be further analyzed to obtain confidence limits in the measured equilibrium dissociation constant(s) and demonstrate that all of the species involved in the chemical equilibrium are actually populated. Both of these features are implemented in SEDPHAT – the former, based on the projection method, is now automated whereas the latter can be accessed through the appropriate sub-menu within the ‘display’ function. These steps are important for the evaluation of the information content of the data and possible design of further experiments. It should be noted that a simple evaluation of whether an interaction is actually occurring or not can be obtained when the single xp files are being constructed and assembled – an analysis in terms of a single non-interacting solute provides a signal-weighted-average molar mass; this parameter should increase with increasing concentrations or under conditions that favor complex formation.

A special case is presented in the study of self-associating systems, in that it is not possible to experimentally determine partial specific volume of the protein of interest by SE. Either SEDFIT or SEDNTERP can be used to provide the partial specific volume of proteins, and these values are generally deemed as reliable. In the case of non-protein self-associating systems, the use of SE presents a particular challenge, as the experimentally determined affinity is correlated to the value of the partial specific volume. In a recent study on the self-association of peptide-nucleic acids (PNA), we confirmed that the partial specific volume was correctly determined (Durchschlag and Zipper, 1997, 1994) by comparing the affinity of interaction obtained by SE with that obtained from a *c(s)* SV analysis (Englund et al., 2012). This result highlights the importance of combining SE and SV experiments.

These considerations can also be applied to protein nucleic-acid interactions. One key feature to note is that both the effective partial specific volume and refractive index increment for nucleic acids depend on the buffer composition and in particular the ionic strength. For this reason, it is crucial that the effective partial specific volume be determined experimentally by SE. Also, when IF data are collected, the molar signal increment needs be

determined experimentally using the defined extinction coefficient at 260 nm as a reference (e.g. see idtdna.com for single-stranded nucleic acids, or biophysics/idtdna.com for double-stranded nucleic acids) as the refractive index increment dn/dc depends on the ionic strength (Kam et al., 1981).

Considerations for the Analysis of Isotherms

As mentioned above, there are several SV isotherm-based analyses available in SEDPHAT. The first step in constructing an isotherm is to analyze all relevant SV data sets in SEDFIT. Regardless of the kinetics of the interaction, there are always at least two isotherms available for the analysis, and they should be globally analyzed. For bimolecular interactions that display slow kinetics (i.e. at least three peaks in a $c(s)$ distribution), two isotherms should be globally analyzed: the s_w isotherm and a population isotherm. The s_w isotherm is derived by integrating in SEDFIT the entirety of the distributions and tabulating the reported signal-average s-values along with the respective total component concentrations in an ASCII file. The population isotherm is constructed similarly, except the signal populations, not the s-values, of the *individual peaks* are tabulated. For rapidly interacting systems, three isotherms can be input: the s_w isotherm, the s_{fast} isotherm, and the "Effective Particle" population isotherm. The s_{fast} isotherm is constructed exactly as the s_w isotherm, except only the reaction boundary is integrated. The Effective Particle population isotherm is constructed like the population isotherm described above, but, because there are only two peaks present (the stationary and reaction boundaries), only these peaks are integrated and their signal populations tabulated. SEDFIT has a tool available that facilitates these analyses and integrations ("Options → Fitting Options → Serialize Fit"). These isotherms are loaded into SEDPHAT by executing the "Data → Load New AUC Isotherm Data" command from the main menu. The extinction information for the components must be input so that the signal information from the isotherms can be converted to concentration space. At this point, it is necessary to impose a model (e.g. A + B ↔ AB); information on the model may come from orthogonal methods or from complementary AUC experiments, such as an MSSV analysis. The s-values of the components must be known and input; presumably these are known from separate SV analyses. The s-value of the complex(es) must be estimated and input. For a slowly exchanging system, $s_{complex}$ is known, but for systems with a fast k_{off}, this s-value may be estimated from the mass of the complex and some hydrodynamic assumptions using calculators available in SEDFIT. If a crystal structure of the complex is known, the pdb file may be submitted to a hydrodynamic modeling program like BEST (Aragon, 2011; Aragon and Hahn, 2006) or HYDROPRO (Garcia de la Torre et al., 2000) for estimating $s_{complex}$. For systems with multivalent components, educated guesses at complex s-values may be the only available option. $s_{complex}$ may be refined, but this optimization should only be attempted if the isotherms reveal significant information on its value, as in the s_{fast} isotherm or certain concentration domains (high concentrations, equimolar) of the s_w isotherm. The total concentrations of the components are taken from the isotherm files. Refinement can also be constrained to certain s-value ranges, based, for example, on hydrodynamic considerations. However, if one suspects that the active concentrations of a component systematically deviate from the input values, a global correction factor can be applied and refined.

After inputting a reasonable initial guess for K_d, a global refinement is initiated that will usually converge rapidly. The fit lines should be scrutinized closely for correspondence to the data points and for systematic deviations therefrom. Poor, very systematic fits may indicate an incorrect model, and the user may experiment with other likely models if appropriate. If the fit is deemed good, the fitted parameters should be examined and their rationality assessed. For example, an attomolar equilibrium dissociation constant and an $s_{complex}$ greater than physically possible would indicate an unrealistic result, and the cause of

these wayward parameters should be investigated. The model may be incorrect, an errant parameter may have been input, the parameter may not be well restrained by the data, or the data may simply not support the simultaneous refinement of all of the parameters. In the absence of an error, one strategy that can be explored is to fix or constrain (within user-defined limits) one or more of the previously fitted parameters at or near to a reasonable value and repeat the fitting session. If more meaningful parameters result, the constraint can be provisionally released to examine its potential to refine rationally under this new, hopefully more stable set of related parameters. A global multi-method analysis, based on the availability of data from multiple biophysical methods, can also be applied. This methodology, implemented in SEDPHAT, uses the orthogonal information content from different data sets as mutual restraints (Zhao and Schuck, 2012). After a reliable fit has been achieved, confidence intervals for the parameters of interest can be obtained using the statistical approaches described in the next section.

General Statistical Methods

With any model, once a best-fit has been found, it is crucial to examine the statistical significance of the result. For this purpose, three fundamentally different approaches have been used in AUC, all implemented in SEDPHAT.

For simple parameter error surfaces that can be optimized well with automated optimization algorithms, Monte-Carlo methods are possible. It requires the generation of synthetic data sets that follow the best-fit model but have different random noise, with similar noise properties as the original data. In each iteration, the model is re-fitted to the synthetic data set, and a distribution of best-fit parameter values is built up. From this, for example, the 95% confidence interval is determined by parameter values encompassing the 2.5% and 97.5% quantile. In order to create meaningful precision of the confidence interval, typically 1000 – 10,000 iterations are required. A significant drawback of this method is presented by the problem that automated optimization does not work well for complex error surfaces. If at some iterations the ‘best-fit’ parameters are caught in local minima, the resulting parameter value distribution will be skewed. This is likely to occur with SE analyses, which have the most complex error surfaces.

A better approach is to trace the projections of the error surface on a parameter of interest (Bevington and Robinson, 1992). This approach fixes the parameter of interest at a sub-optimal value close to the best-fit value, and tests to what extent the fit quality degrades while allowing all other parameters to compensate for the constraint. This test is repeated multiple times moving the parameter of interest further from the best-fit value. Once the increase in the rmsd of the fit exceeds a statistically pre-set level based on F-statistics, one of the sides of the error interval is reached. The same approach is used for the other side of the confidence interval. In SEDPHAT, this procedure has been automated (which is simpler than in the Monte-Carlo case since one parameter is always fixed), providing on user request just the confidence interval or complete traces of the error surface projections. The latter can be highly instructive, in particular since ubiquitous systematic errors in the data make the choice of the confidence level ambiguous.

A third approach, long recognized to be highly unreliable in actually estimating parameter uncertainties, is the co-variance matrix. In this method, essentially a parabolic approximation of the error surface is made at the global minimum, and its curvature is used to predict parameter errors. In contrast to the F-statistics method above, it relies on the absolute errors of the data acquisition to be known. Clearly, the parabolic approximation of the error surface may not be realistic over the range of parameter values in question. However, this method can display the cross-correlation coefficients of the parameter values, which is useful for quickly examining the rigidity of the model.

Generally, it cannot be over-emphasized that the parameter error estimates can be greatly influenced by the application of constraints, for example, hydrodynamic constraints on *s*-values, constraints on the loading concentrations by suitable experimental design in both SE and SV, or mutual constraints arising in the global analysis of multiple data sets or multiple techniques (Zhao and Schuck, 2012).

Using GUSSI to produce publication-quality graphs of SV and SE data and analyses

GUSSI is a stand-alone, companion program for SEDFIT and SEDPHAT. GUSSI can be called from the Plot menus of these analytical programs to automatically parse and present the results of the analysis. It gives the user considerable control over the appearance of the output, and can write high-resolution images that are suitable for presentations and publication. The GUSSI interface features a large view of the graph along with a GUI-based "Control Panel" that eases access to the most-used graph parameter values. Changes in the parameters are updated in the graph window in real time. Once the user has settled on an appearance for the graph, a "state" file may be saved that contains all information necessary to reproduce the plot. Also, plot can be saved in one of four file formats: EPS, PNG, PDF, and SVG. The following protocols are meant as brief introductions; further information is available in the GUSSI documentation, which is available from that program's Help menu.

For the c(s) distribution

1. Load SV data into SEDFIT or SEDPHAT and analyze using a continuous distribution, a segmented distribution, or a hybrid discrete/continuous distribution.
2. In the analysis program, choose "Plot → GUSSI c(s) plot". If you are working in SEDPHAT, you may be asked some questions about the appearance of the distribution(s).
3. GUSSI will open. A splash screen will appear, followed by the initial GUSSI $c(s)$ plot with defaults applied.
4. Add distributions as needed. This task can be accomplished by:
    a. Adding saved distributions.
    b. Copying and pasting. One can choose "Copy → Copy Distribution Table" in a different SEDFIT or SEDPHAT session and then paste these data into GUSSI (Ctrl-V or "Distributions → Paste a Distribution").
5. Adjust individual distribution properties as desired. Available options for each distribution are:
    a. Line color, type, and thickness.
    b. Marker existence, type, color, and size.
    c. Area fills below the distribution
6. Adjust global plot properties as desired, including:
    a. Normalization.
    b. Axis properties (ticks, text, and limits).
    c. Existence and properties of a legend.
    d. Display of integral plots in addition to the differential $c(s)$ distributions.
    e. Conversion of the *s*-values of the distribution to standard conditions ($s_{20,w}$).

f. Figure size and dpi.

7. Save the GUSSI state ("File → Save GUSSI State") if recall of the graph for further manipulation is desired. Also, the data only can be saved and input later into the program.

8. Save the figure by choosing "Save Figure Only" from GUSSI's File Menu.

For a data/fit/residual plot of SV data

1. Load SV data into SEDFIT or SEDPHAT and analyze using any available model that will apply the model directly to the SV data.

2. In the analysis program, choose "Plot → GUSSI data-fit-residuals plot". If multiple data sets are loaded into SEDPHAT as "Experiments", the SEDPHAT will prompt you to select one of them for plotting (only one at a time may currently be plotted in this GUSSI module).

3. GUSSI will open after displaying its splash screen. The data will be displayed with default parameters applied. As in SEDFIT and SEDPHAT, the data and fits are presented in an upper panel, whereas the residual plot is presented below.

4. Adjust parameters as desired, including:

   a. Color mode. By default, early scans are colored violet, and late scans are red; scans between these extremes are drawn progressing through rainbow colors. Other modes available are a SEDFIT-like color gradient, a grayscale gradient, black-and-white, a uniform color, or a user-defined gradient.

   b. The existence, color, type, and thickness of the fit lines.

   c. The existence, color, and size of the data markers.

   d. The existence of the residual plot and appearance of its lines.

   e. Data sampling; SV data are usually too dense to show all of them on a single presentation-quality graph, so GUSSI by default down-samples the data it receives to every 3rd scan and every 3rd data point. These sampling rates are user-adjustable.

   f. Which systematic noise features are subtracted from the data and fit lines.

   g. Axis ticks, text, and limits.

   h. Figure size and dpi.

5. Save the GUSSI state ("File → Save GUSSI State") so that the adjusted graph state can be recalled if necessary. As with the $c(s)$ distribution, the data only may also be saved ("File → Save Data Only") and loaded later into the program.

6. Save the figure by choosing "Save Figure Only" from GUSSI's File Menu.

For a data/fit/residuals plot of SE data

1. Load and analyze SE data in SEDPHAT (this step is facilitated by utility functions in SEDFIT).

2. Once the analysis is final, choose "Plot → GUSSI data-fit-residuals plot".

3. If multiple "Experiments" are being analyzed, SEDPHAT will prompt the user to select one for plotting. GUSSI is optimized to display multi-speed SE data sets that

are recommended for use in SEDPHAT (see “Considerations for the Analysis of SE Data,” above).

4. After GUSSI’s splash screen disappears, it displays the data, fit lines, and residuals as in SEDPHAT, with the data and fit lines in an upper panel and the residuals in a lower panel.
5. Adjust the appearance of the figure as desired. Adjustable parameters for individual concentration distributions include:
    a. Fit line color, style, and width.
    b. Marker existence, color, style, and size.
    c. Existence of the residual plot and the appearance of its lines.
    d. Data-point sampling. Often, especially for long-column data, the data are too dense to be shown on this style of plot, and so the data-points can easily be down-sampled.
6. Adjust the overall parameters of the plot, including:
    a. Axis ticks, text, and limits.
    b. Subtraction of noise from the data and fit lines.
    c. Existence and properties of a legend.
    d. Figure size and dpi.
7. Save the GUSSI state (“File → Save GUSSI State”) if it is desirable to alter the graph later. The as with the other two protocols, the data alone may be saved (“File → Save Data Only”) and later input back into GUSSI.
8. Save the figure by choosing “Save Figure Only” from GUSSI’s File Menu.

GUSSI may also be started independently of SEDFIT or SEDPHAT to facilitate the updating of saved data or “ states. ” The program is a vailable free-of-charge from http://biophysics.swmed.edu/MBR/software.

Acknowledgments

This work was supported by the Intramural Research Program of the National Institutes of Health, National Institute of Biomedical Imaging and Bioengineering (H.Z. and P.S) and National Institute of Diabetes and Digestive and Kidney Diseases (R.G.). C.A.B. acknowledges support from grants AI056305, GM056322, and DK026758 to Michael Norgard, Michael Rosen, and David Chuang, respectively.

Literature Cited

Ansevin AT, Roark DE, Yphantis DA. Improved ultracentrifuge cells for high-speed sedimentation equilibrium studies with interference optics. Anal Biochem. 1970; 34:237–261. [PubMed: 4314972]

Aragon S, Hahn DK. Precise boundary element computation of protein transport properties: Diffusion tensors, specific volume, and hydration. Biophys J. 2006; 91:1591–1603. [PubMed: 16714342]

Aragon SR. Recent advances in macromolecular hydrodynamic modeling. Methods. 2011; 54:101–114. [PubMed: 21073955]

Balbo, A.; Brown, PH.; Schuck, P. Experimental protocol for sedimentation equilibrium analytical ultracentrifugation. 2007.

Balbo A, Minor KH, Velikovsky CA, Mariuzza R, Peterson CB, Schuck P. Studying multi-protein complexes by multi-signal sedimentation velocity analytical ultracentrifugation. Proc Nat Acad Sci USA. 2005; 102:81–86. [PubMed: 15613487]

Balbo A, Zhao H, Brown PH, Schuck P. Assembly, loading, and alignment of an analytical ultracentrifuge sample cell. J Vis Exp. 2009; 5:1530. [PubMed: 19893484]

Berke IC, Modis Y. MDA5 cooperatively forms dimers and ATP-sensitive filaments upon binding double-stranded RNA. EMBO J. 2012; 31:1714–1726. [PubMed: 22314235]

Berkowitz SA. Role of analytical ultracentrifugation in assessing the aggregation of protein biopharmaceuticals. AAPS J. 2006; 8:E590–605. [PubMed: 17025277]

Besong TMD, Harding SE, Winzor DJ. The effective time of centrifugation for the analysis of boundary spreading in sedimentation velocity experiments. Anal Biochem. 2012; 421:755–758. [PubMed: 22197415]

Bevington, PR.; Robinson, DK. Data Reduction and Error Analysis for the Physical Sciences. 1992.

Bothwell MA, Howlett GJ, Schachman HK. A sedimentation equilibrium method for determining molecular weights of proteins with a tabletop high speed air turbine centrifuge. J Biol Chem. 1978; 253:2073–2077. [PubMed: 632258]

Brautigam CA. Using Lamm-Equation modeling of sedimentation velocity data to determine the kinetic and thermodynamic properties of macromolecular interactions. Methods. 2011; 54:4–15. [PubMed: 21187153]

Brookes E, Cao W, Demeler B. A two-dimensional spectrum analysis for sedimentation velocity experiments of mixtures with heterogeneity in molecular weight and shape. Eur Biophys J. 2009; 39:405–414. [PubMed: 19247646]

Brown PH, Balbo A, Schuck P. Using prior knowledge in the determination of macromolecular size-distributions by analytical ultracentrifugation. Biomacromolecules. 2007; 8:2011–2024. [PubMed: 17521163]

Brown PH, Balbo A, Schuck P. On the analysis of sedimentation velocity in the study of protein complexes. Eur Biophys J. 2009; 38:1079–1099. [PubMed: 19644686]

Brown PH, Balbo A, Zhao H, Ebel C, Schuck P. Density contrast sedimentation velocity for the determination of protein partial-specific volumes. PLoS One. 2011; 6:e26221. [PubMed: 22028836]

Brown PH, Schuck P. Macromolecular Size-And-Shape Distributions by Sedimentation Velocity Analytical Ultracentrifugation. Biophys J. 2006; 90:4651–4661. [PubMed: 16565040]

Brown PH, Schuck P. A new adaptive grid-size algorithm for the simulation of sedimentation velocity profiles in analytical ultracentrifugation. Comput Phys Commun. 2008; 178:105–120. [PubMed: 18196178]

Cao W, Demeler B. Modeling analytical ultracentrifugation experiments with an adaptive space-time finite element solution for multicomponent reacting systems. Biophys J. 2008; 95:54–65. [PubMed: 18390609]

Carney RP, Kim JY, Qian H, Jin R, Mehenni H, Stellacci F, Bakr OM. Determination of nanoparticle size distribution together with density or molecular weight by 2D analytical ultracentrifugation. Nat Commun. 2011; 2:335. [PubMed: 21654635]

Claverie JM, Dreux H, Cohen R. Sedimentation of generalized systems of interacting particles. I Solution of systems of complete Lamm equations. Biopolymers. 1975; 14:1685–1700. [PubMed: 1156660]

Correia JJ, Stafford WF. Extracting equilibrium constants from kinetically limited reacting systems. Methods Enzymol. 2009; 455:419–446. [PubMed: 19289215]

Cox DJ. Computer simulation of sedimentation in the ultracentrifuge. II Concentration-independent sedimentation. Arch Biochem Biophys. 1965; 112:259–266.

Cox DJ. Sedimentation of an initially skewed boundary. Science. 1966; 152:359–361. [PubMed: 17775166]

Dam J, Schuck P. Calculating sedimentation coefficient distributions by direct modeling of sedimentation velocity profiles. Methods Enzymol. 2004; 384:185–212. [PubMed: 15081688]

Dam J, Velikovsky CA, Mariuzza R, Urbanke C, Schuck P. Sedimentation velocity analysis of protein-protein interactions: Lamm equation modeling and sedimentation coefficient distributions c(s). Biophys J. 2005; 89:619–634. [PubMed: 15863475]

Darawshe S, Rivas G, Minton AP. Rapid and accurate microfractionation of the contents of small centrifuge tubes: application in the measurement of molecular weights of proteins via sedimentation equilibrium. Anal Biochem. 1993; 209:130–135. [PubMed: 8465945]

Demeler B, Saber H, Hansen JC. Identification and interpretation of complexity in sedimentation velocity boundaries. Biophys J. 1997; 72:397–407. [PubMed: 8994626]

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

Dorwart MR, Wray R, Brautigam CA, Jiang Y, Blount P. S. aureus MscL is a pentamer in vivo but of variable stoichiometries in vitro: implications for detergent-solubilized membrane proteins. PLoS biology. 2010; 8:e1000555. [PubMed: 21151884]

Durchschlag H, Zipper P. Calculation of the partial specific volume of organic compounds and polymers. Progr Colloid Polym Sci. 1994; 94:20–39.

Durchschlag H, Zipper P. Calculation of partial specific volumes and other volumetic properties of small molecules and polymers. J Appl Cryst. 1997; 30:803–807.

Ebel C. Sedimentation velocity to characterize surfactants and solubilized membrane proteins. Methods. 2011; 54:56–66. [PubMed: 21112401]

Englund, Ea; Wang, D.; Fujigaki, H.; Sakai, H.; Micklitsch, CM.; Ghirlando, R.; Martin-Manso, G.; Pendrak, ML.; Roberts, DD.; Durell, SR.; Appella, DH. Programmable multivalent display of receptor ligands using peptide nucleic acid nanoscaffolds. Nat Commun. 2012; 3:614. [PubMed: 22233624]

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

Gabrielson JP, Arthur KK. Measuring low levels of protein aggregation by sedimentation velocity. Methods. 2011; 54:83–91. [PubMed: 21187149]

Gabrielson JP, Arthur KK, Kendrick BS, Randolph TW, Stoner MR. Common excipients impair detection of protein aggregates during sedimentation velocity analytical ultracentrifugation. J Pharm Sci. 2009; 98:50–62. [PubMed: 18425806]

Gabrielson JP, Brader ML, Pekar AH, Mathis KB, Winter G, Carpenter JF, Randolph TW. Quantitation of aggregate levels in a recombinant humanized monoclonal antibody formulation by size-exclusion chromatography, asymmetrical flow field flow fractionation, and sedimentation velocity. J Pharm Sci. 2007; 96:268–279. [PubMed: 17080424]

Garcia de la Torre J, Huertas ML, Carrasco B. Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. Biophys J. 2000; 78:719–730. [PubMed: 10653785]

Ghirlando R. The analysis of macromolecular interactions by sedimentation equilibrium. Methods. 2011; 54:145–156. [PubMed: 21167941]

Ghirlando R, Felsenfeld G. Hydrodynamic studies on defined heterochromatin fragments support a 30-nm fiber having six nucleosomes per turn. J Mol Biol. 2008; 376:1417–1425. [PubMed: 18234217]

Ghirlando R, Litt MD, Prioleau MN, Recillas-Targa F, Felsenfeld G. Physical properties of a genomic condensed chromatin fragment. J Mol Biol. 2004; 336:597–605. [PubMed: 15095975]

Harding SE, Schuck P, Abdelhameed AS, Adams G, Kök MS, Morris GA. Extended Fujita approach to the molecular weight distribution of polysaccharides and other polymeric systems. Methods. 2011; 54:136–144. [PubMed: 21276851]

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

Hellman LM, Zhao C, Melikishvili M, Tao X, Hopper JE, Whiteheart SW, Fried MG. Histidine-tag-directed chromophores for tracer analyses in the analytical ultracentrifuge. Methods. 2011; 54:31–38. [PubMed: 21187151]

Houtman JC, Yamaguchi H, Barda-Saad M, Braiman A, Bowden B, Appella E, Schuck P, Samelson LE. Oligomerization of signaling complexes by the multipoint binding of GRB2 to both LAT and SOS1. Nat Struct Mol Biol. 2006; 13:798–805. [PubMed: 16906159]

HeteroAnalysis Contents. 2012. http://www.biotech.uconn.edu/auf/ha-help/hs2024.htm

Inagaki S, Ghirlando R, Grisshammer R. Biophysical characterization of membrane proteins in nanodiscs. Methods. 2012 S1046-2023(12)00301-5.

Johnson ML. Why, when, and how biochemists should use least squares. Anal Biochem. 1992; 225:215–225. [PubMed: 1443589]

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

Kam Z, Borochov N, Eisenberg H. Dependence of laser light scattering of DNA on NaCl concentration. Biopolymers. 1981; 20:2671–2690. [PubMed: 7034800]

Kar SR, Kinsbury JS, Lewis MS, Laue TM, Schuck P. Analysis of transport experiment using pseudo-absorbance data. Anal Biochem. 2000; 285:135–142. [PubMed: 10998273]

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

MacRaild CA, Hatters DM, Lawrence LJ, Howlett GJ. Sedimentation velocity analysis of flexible macromolecules: self-association and tangling of amyloid fibrils. Biophys J. 2003; 84:2562–2569. [PubMed: 12668464]

Nury H, Manon F, Arnou B, le Maire M, Pebay-Peyroula E, Ebel C. Mitochondrial bovine ADP/ATP carrier in detergent is predominantly monomeric but also forms multimeric species. Biochemistry. 2008; 47:12319–12331. [PubMed: 18980386]

Padrick SB, Brautigam CA. Evaluating the stoichiometry of macromolecular complexes using multisignal sedimentation velocity. Methods. 2011; 54:39–55. [PubMed: 21256217]

Padrick SB, Deka RK, Chuang JL, Wynn RM, Chuang DT, Norgard MV, Rosen MK, Brautigam CA. Determination of protein complex stoichiometry through multisignal sedimentation velocity experiments. Anal Biochem. 2010; 407:89–103. [PubMed: 20667444]

Padrick SB, Doolittle LK, Brautigam CA, King DS, Rosen MK. Arp2/3 complex is bound and activated by two WASP proteins. Proc Nat Acad Sci USA. 2011; 108:E472–479. [PubMed: 21676863]

Philo JS. Sedimentation equilibrium analysis of mixed associations using numerical constraints to impose mass or signal conservation. Methods Enzymol. 2000; 321:100–120. [PubMed: 10909053]

Provencher SW. A constrained regularization method for inverting data represented by linear algebraic or integral equations. Comput Phys Commun. 1982; 27:213–227.

Richards EG, Teller DC, Schachman HK. Ultracentrifuge studies with Rayleigh interference optics. II Low-speed sedimentation equilibrium of homogeneous systems. Biochemistry. 1968; 7:1054–1076. [PubMed: 5657846]

Rivas G, Fernandez JA, Minton AP. 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. 1999; 38:9379–9388. [PubMed: 10413513]

Rivas G, Minton AP. Beyond the second virial coefficient: Sedimentation equilibrium in highly non-ideal solutions. Methods. 2011; 54:167–174. [PubMed: 21112402]

Roark DE. Sedimentation equilibrium techniques: multiple speed analyses and an overspeed procedure. Biophys Chem. 1976; 5:185–196. [PubMed: 963214]

Rowe AJ. Ultra-weak reversible protein-protein interactions. Methods. 2011; 54:157–166. [PubMed: 21338686]

Rowe, AJ. Analytical Ultracentrifugation: Sedimentation Equilibrium. In: Roberts, GCK., editor. Encyclopedia of Biophysics. Springer; Berlin: 2010.

Salvay AG, Santamaria M, LeMaire M, Ebel C. Analytical ultracentrifugation sedimentation velocity for the characterization of detergent-solubilized membrane proteins Ca++-ATPase and ExbB. J Biol Phys. 2007; 33:399–419. [PubMed: 19669527]

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

Schuck P. Size distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. Biophys J. 2000; 78:1606–1619. [PubMed: 10692345]

Schuck P. On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation. Anal Biochem. 2003; 320:104–124. [PubMed: 12895474]

Schuck P. A model for sedimentation in inhomogeneous media. I Dynamic density gradients from sedimenting co-solutes. Biophys Chem. 2004; 108:187–200. [PubMed: 15043929]

Schuck P. On computational approaches for size-and-shape distributions from sedimentation velocity analytical ultracentrifugation. Eur Biophys J. 2009; 39:1261–1275. [PubMed: 19806353]

Schuck P. Diffusion of the reaction boundary of rapidly interacting macromolecules in sedimentation velocity. Biophys J. 2010a; 98:2741–2751. [PubMed: 20513419]

Schuck P. Sedimentation patterns of rapidly reversible protein interactions. Biophys J. 2010b; 98:2005–2013. [PubMed: 20441765]

Schuck P. Some statistical properties of differencing schemes for baseline correction of sedimentation velocity data. Anal Biochem. 2010c; 401:280–287. [PubMed: 20206114]

Schuck, P. Sedimentation velocity analytical ultracentrifugation. In: Roberts, GCK., editor. Encyclopedia of Biophysics. Springer; Berlin: 2012.

Schuck, P.; Balbo, A.; Brown, PH.; Zhao, H. Analytical Ultracentrifugation. In: Meyers, RA., editor. Encyclopedia of Analytical Chemistry. John Wiley; Chichester: 2010.

Schuck P, Demeler B. Direct sedimentation analysis of interference optical data in analytical ultracentrifugation. Biophys J. 1999; 76:2288–2296. [PubMed: 10096923]

Schuck P, MacPhee CE, Howlett GJ. Determination of sedimentation coefficients for small peptides. Biophys J. 1998; 74:466–474. [PubMed: 9449347]

Schuck P, Perugini MA, Gonzales NR, Howlett GJ, Schubert D. Size-distribution analysis of proteins by analytical ultracentrifugation: strategies and application to model systems. Biophys J. 2002; 82:1096–1111. [PubMed: 11806949]

Schuck P, Rossmanith P. Determination of the sedimentation coefficient distribution by least-squares boundary modeling. Biopolymers. 2000; 54:328–341. [PubMed: 10935973]

Schuck P, Taraporewala Z, McPhie P, Patton JT. Rotavirus nonstructural protein NSP2 self-assembles into octamers that undergo ligand-induced conformational changes. J Biol Chem. 2000; 276:9679–9687. [PubMed: 11121414]

Schwieters CD, Suh JY, Grishaev A, Ghirlando R, Takayama Y, Clore GM. Solution structure of the 128 kDa enzyme I dimer from Escherichia coli and its 146 kDa complex with HPr using residual dipolar couplings and small- and wide-angle X-ray scattering. J Am Chem Soc. 2010; 132:13026–13045. [PubMed: 20731394]

Solovyova A, Schuck P, Costenaro L, Ebel C. Non-ideality by sedimentation velocity of halophilic malate dehydrogenase in complex solvents. Biophys J. 2001; 81:1868–1880. [PubMed: 11566761]

Stafford WF. Boundary analysis in sedimentation transport experiments: a procedure for obtaining sedimentation coefficient distributions using the time derivative of the concentration profile. Anal Biochem. 1992; 203:295–301. [PubMed: 1416025]

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

van Holde KE, Baldwin RL. Rapid attainment of sedimentation equilibrium. J Phys Chem. 1958; 62:734–749.

Vistica J, Dam J, Balbo A, Yikilmaz E, Mariuzza RA, Rouault TA, Schuck P. Sedimentation equilibrium analysis of protein interactions with global implicit mass conservation constraints and systematic noise decomposition. Anal Biochem. 2004; 326:234–256. [PubMed: 15003564]

Yphantis DA. Equilibrium ultracentrifugation of dilute solutions. Biochemistry. 1964; 3:297–317. [PubMed: 14155091]

Zhao H, Balbo A, Brown PH, Schuck P. The boundary structure in the analysis of reversibly interacting systems by sedimentation velocity. Methods. 2011a; 54:16–30. [PubMed: 21315155]

Zhao H, Berger AJ, Brown PH, Kumar J, Balbo A, May CA, Casillas E, Laue TM, Patterson GH, Mayer ML, Schuck P. Analysis of high-affinity assembly for AMPA receptor amino-terminal domains. J Gen Physiol. 2012; 139:371–388. [PubMed: 22508847]

Zhao H, Brown PH, Balbo A, Fernandez Alonso MC, Polishchuck N, Chaudhry C, Mayer ML, Ghirlando R, Schuck P. Accounting for solvent signal offsets in the analysis of interferometric sedimentation velocity data. Macromol Biosci. 2010; 10:736–745. [PubMed: 20480511]

Zhao H, Brown PH, Schuck P. On the distribution of protein refractive index increments. Biophys J. 2011b; 100:2309–2317. [PubMed: 21539801]

Zhao H, Schuck P. Global multi-method analysis of affinities and cooperativity in complex systems of macromolecular interactions. Anal Chem. 2012; 84:9513–9519. [PubMed: 23020071]

Zhi L, Mans J, Paskow MJ, Brown PH, Schuck P, Jonji S, Natarajan K, Margulies DH. Direct interaction of the mouse cytomegalovirus m152/gp40 immunoevasin with RAE-1 isoforms. Biochemistry. 2010; 49:2443–2453. [PubMed: 20166740]

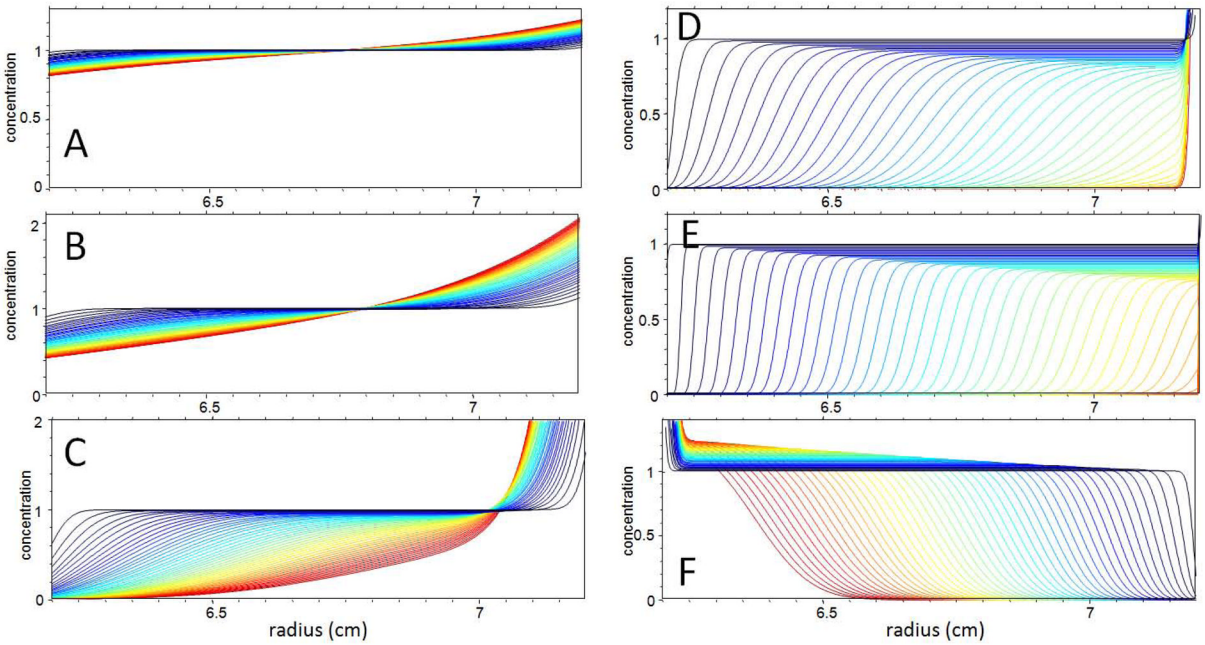

Figure 1.

Examples for the shapes of Lamm equation solutions. The profiles are calculated for particles of different size and sedimentation properties. All conditions are calculated for a rotor speed of 50,000 rpm, and 50 concentration profiles are shown at different (in each case equally spaced) time intervals. Later scans are indicated by a higher color temperature. (A) Small molecules with a sedimentation coefficient of 0.2 S and a diffusion coefficient of 6×10^{-6} cm^2/sec, in Δt = 300 sec intervals. Similar values are frequently observed for sedimenting buffer salts. (B) Sedimentation of a peptide of 1 kg/mol and 0.3 S, Δt = 1000 sec. (C) A small protein of 10 kg/mol and 1.5 S, Δt = 500 sec. (D) A protein of 100 kg/mol and 6 S, Δt = 300 sec. (E) Particle of 1 Mg/mol and 30 S, Δt = 50 sec. (F) A floating particle with a sedimentation coefficient of −3.0 S and a diffusion coefficient of 2.71×10^{-7} cm^2/sec. Such data patterns may be obtained, for example, with large emulsion or lipid particles. In flotation, the radial dilution is replaced with radial increase in concentration in the plateau region of successive profiles. (Figure reproduced from (Schuck, 2012).)

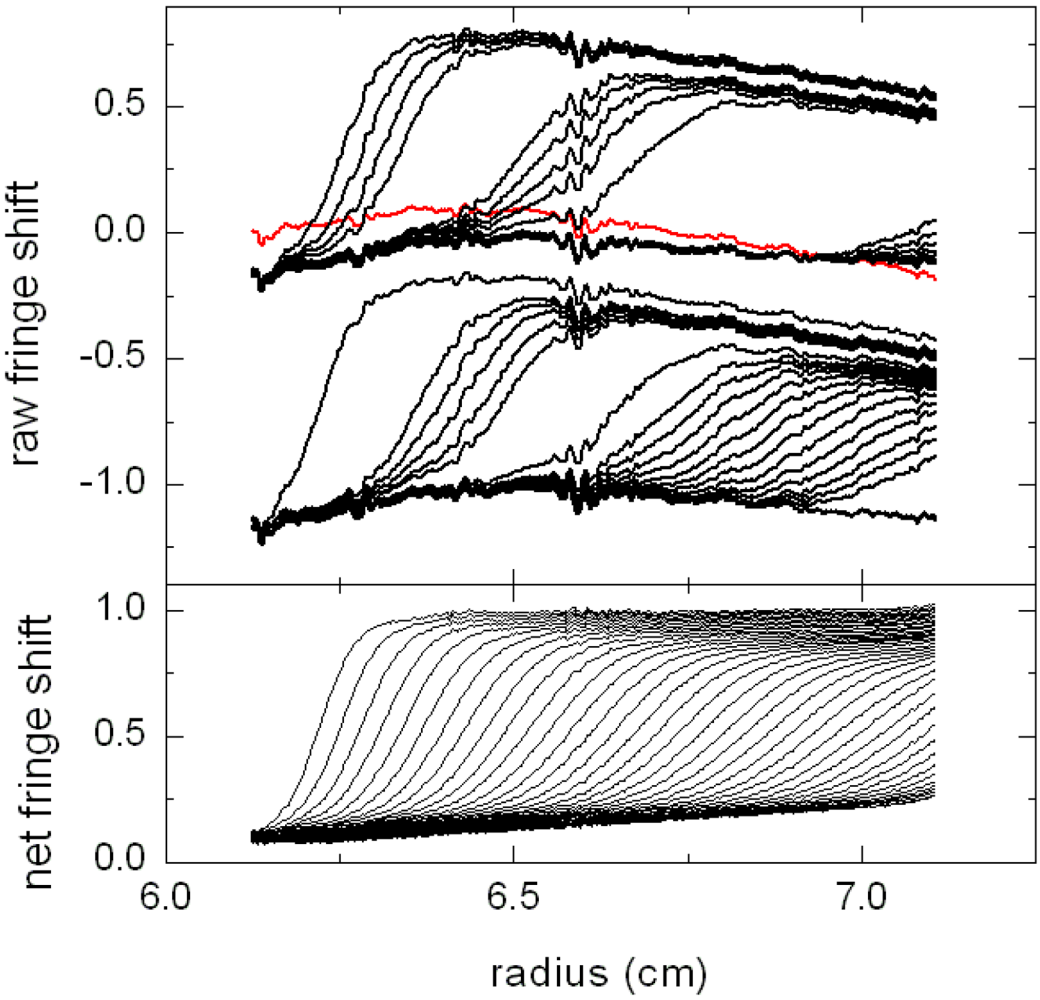

Figure 2.

Demonstration of interference optical data before (top) and after (bottom) subtracting the estimated TI and RI noise contributions. Both the time-dependent (radius-invariant, RI) signal offsets arising from jitter and integral fringe shifts, as well as the radius-dependent (time-invariant, TI) signal offset arising from imperfections in the smoothness of the optical elements, can be clearly discerned. After fitting the data with a model including terms for TI and RI noise, the best-fit values for the TI noise (red line in top panel) and for the RI noise (not shown) can be subtracted from the raw data, as shown in the lower panel. As long as the degrees of freedom for unknown TI and RI noise are maintained in the further analysis, this subtraction does not alter the information content of the data, but allows better visual inspection of the signal from the macromolecular redistribution. (Figure reproduced from Curr. Protoc. Immunol., Chapter 18, Unit 18 15 (2008)).

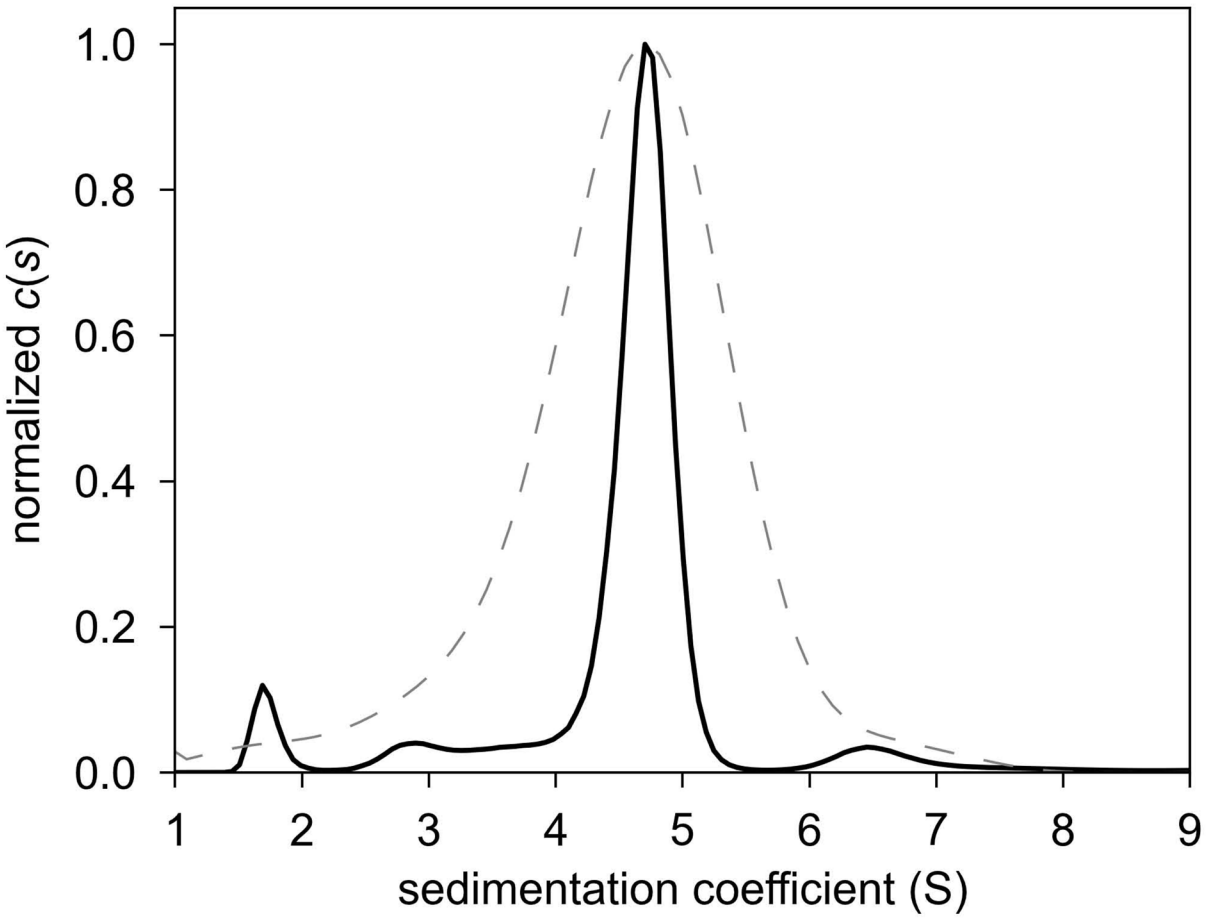

Figure 3.
c(s) distribution of the data shown in Figure 2. For comparison, the dashed line shows the *ls-g*(s)* distribution derived from a subset of the scans.

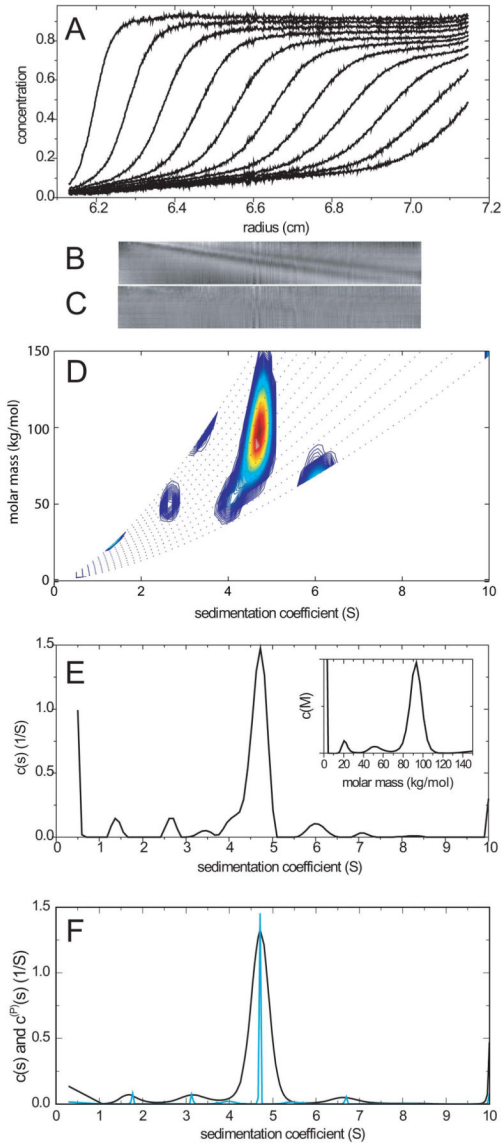

Figure 4.

Example of SV analyses of a system with non-interacting species. (A) Representative subset of the raw data after elimination of systematic noise contributions. (B) Residuals bitmap from a fit with insufficient quality: the data in (A) are modeled with an impostor single-species fit, resulting in clearly systematic deviations that can be discerned from the strong diagonal feature in the bitmap. (C) The quality of the fit with a c(s) model results in a residuals bitmap with very few diagonal features. There are some vertical and horizontal lines indicative of the remaining residuals due to technical imperfections in the data acquisition process, such as higher-order vibrations of optical components. (D) Size and-shape distribution, transformed into coordinates of sedimentation coefficient and molar mass. The color temperature of the contour lines indicates the population of species. Like in one-dimensional c(s), the peak-width in c(s,M) contains contributions both from regularization (reflecting limited resolution given the signal-to-noise ratio of the data) and from true heterogeneity. (E) Reduction of the c(s,M) distribution to a pure sedimentation coefficient distribution, general c(s,*). This is equivalent to a conventional c(s) analysis but

without any constraints to a common average frictional ratio of all species. The inset shows a pure molar-mass distribution, c(M,*), also derived by integration of c(s,M) in a direction orthogonal to c(s,*). (F) Size distribution c(s) using a hydrodynamic scaling law (black line with broad peaks). Also shown is the result of a Bayesian analysis using prior knowledge in the analysis of this non-interacting system, here in the form of $c^{(P\delta)}(s)$ (blue line with sharp peaks) using the hypothesis that the sample consists of discrete species. Generally, the peak width in c(s) can result from either a true polydispersity of the protein (e.g., strong heterogeneity in glycosylation, in conformation, primary sequence, etc.), or from the standard regularization favoring broader peaks for data with low signal/noise ratio. (Figure reproduced from (Schuck et al., 2010).)

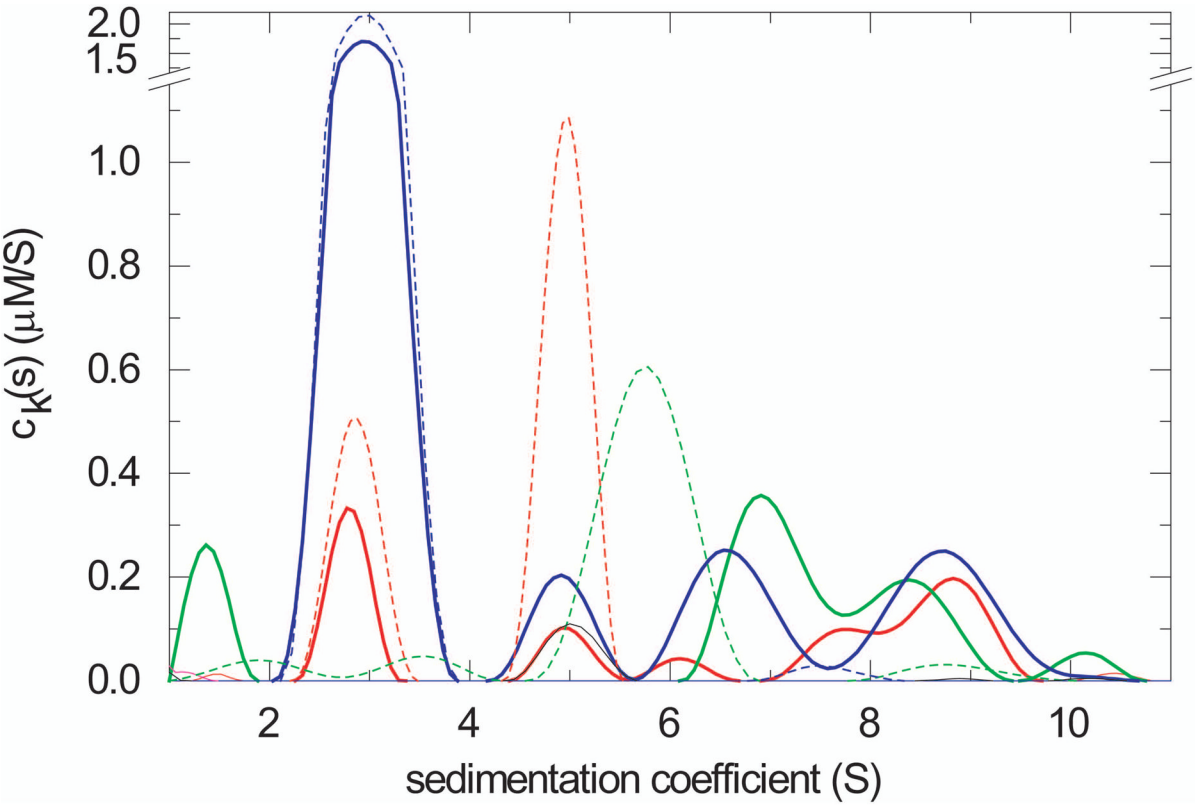

Figure 5.

Example of the multi-signal $c_k(s)$ analysis of a triple protein mixture of a viral glycoprotein (green), its cognate receptor (blue), and a heterogeneous antigen-recognition receptor fragment (red). The content of each protein component in the different *s*-ranges is obtained from the global analysis of sedimentation data acquired with the interference optics and with the absorbance system at two different wavelengths (data not shown), using two chromophorically labeled proteins and one unlabeled protein. Solid lines show the $c_k(s)$ analysis of the triple mixture. The analogous distributions of each protein alone are shown as dashed lines. The formation of two coexisting binary complexes at ~5 S and ~7 S and a ternary complex with 1:1:1 stoichiometry at ~8.5 S can be discerned. Figure reproduced from (Schuck et al., 2010).

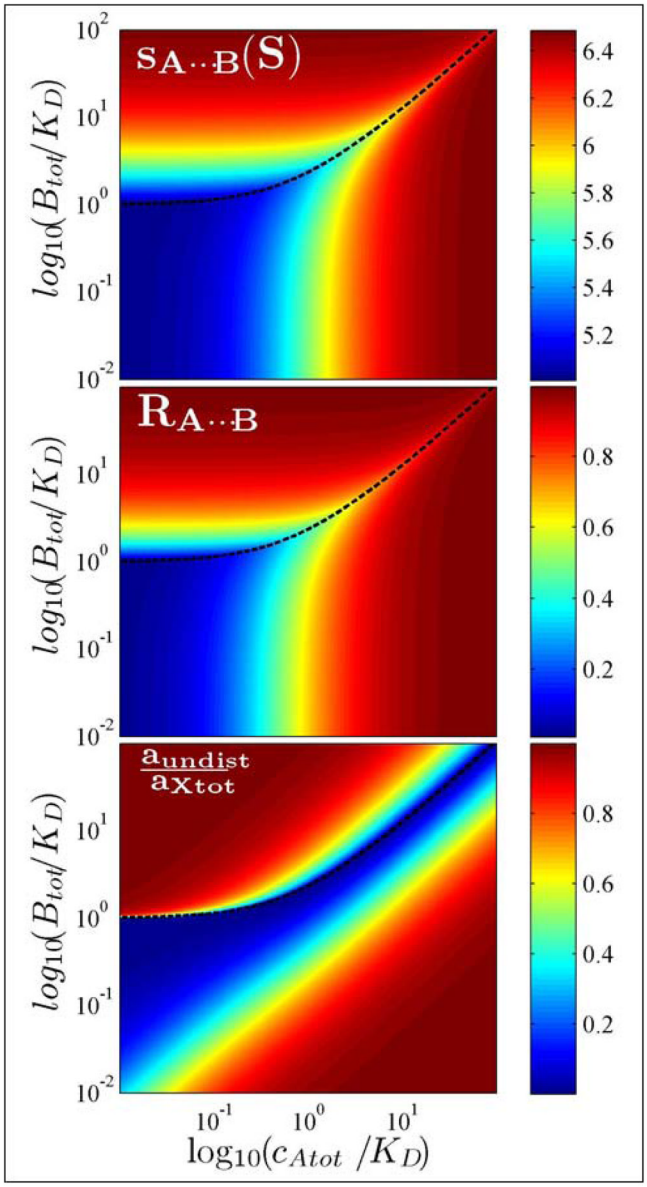

Figure 6.

Properties of the reaction boundary A···B as a function of the total loading concentration of A and B, calculated by EPT for the system of Figure 1. (Top) Velocity of the reaction boundary $s_{A\cdots B}$ following Eq. 6. (Middle) Composition $R_{A\cdots B}$ of the reaction boundary following Eq. 7. (Bottom) Fractional signal of the undisturbed boundary, assuming that both components are globular with equal weight-based extinction coefficients. In all plots the line for the phase transition $c_{Btot}^{*}(c_{Atot})$ is shown as black dotted line, separating the region of A···(B) in the upper left quadrant from B (A) elsewhere. Figure reproduced from (Schuck, 2010b).