Current Protocols in Immunology (in press)

Characterizing Protein-Protein Interactions by Sedimentation Velocity Analytical Ultracentrifugation

Patrick H. Brown, Andrea Balbo, and Peter Schuck[#]

Protein Biophysics Resource, DBEPS, National Institute for Biomedical Imaging and Bioengineering**,** National Institutes of Health, Bethesda, MD

[#]to whom correspondence should be addressed

keywords: sedimentation equilibrium, sedimentation velocity, chemical equilibria, reversible interactions, multi-protein complex, analytical ultracentrifugation, size-distribution, Gilbert-Jenkins theory, Lamm equation, Bayesian analysis

Mini-abstract: This Unit introduces the basic principles and practice of sedimentation velocity analytical ultracentrifugation for the study of reversible protein interactions, such as the characterization of self-association, heterogeneous association, multi-protein complexes, binding stoichiometry, and the determination of association constants. The analytical tools described include sedimentation coefficient and molar mass distributions, multi-signal sedimentation coefficient distributions, Gilbert-Jenkins theory, different forms of isotherms, and global Lamm equation modeling. Concepts for the experimental design are discussed, and a detailed step-by-step protocol guiding through the experiment and the data analysis is available as an internet resource.

Address for Correspondence:

Dr. Peter Schuck
National Institutes of Health
13 South Drive, Bldg. 13, Rm. 3N17
Bethesda, MD 20892

Phone: (301) 435-1950
Fax: (301) 496-6608
Email: pschuck@helix.nih.gov

1. Introduction

Sedimentation velocity (SV) analytical ultracentrifugation is a highly versatile method that can shed light on many aspects of protein interactions, including the number of different classes of complexes, their stoichiometry, binding constants from the nanomolar to millimolar range, and structural changes associated with binding. It is frequently applied to determine gross shapes and conformational changes of proteins in solution (Calarese et al. 2003; Harding et al. 2003; West et al. 2004; Burgess et al. 2005; Connaghan-Jones et al. 2006), characterize protein self-association and oligomeric state (Schuck et al. 2000; Arthos et al. 2002; Ali et al. 2003; Guan et al. 2004; Solovyova et al. 2004; Burgess et al. 2005; Doun et al. 2005; Hsu et al. 2005; Lelj-Garolla and Mauk 2005; Li et al. 2005; Perugini et al. 2005; Breuer et al. 2006; Chan et al. 2006; Egan et al. 2006; Lelj-Garolla and Mauk 2006; Yang et al. 2006; Jomaa et al. 2007; Patel et al. 2007), as well as hetero-association (West et al. 2004; Gupta et al. 2005; Deng et al. 2007; Li et al. 2007) and mixed processes involving single- or multi-step binary systems (Dam et al. 2003; Minor et al. 2005; Dam et al. 2006; Deka et al. 2007), and three-component protein systems with ternary macromolecular complexes (Houtman et al. 2004; Greive et al. 2005; Houtman et al. 2006). Although not a very informative method for determining detailed binding kinetics, it can be readily detected whether the kinetics is fast or slow relative to the time-scale of sedimentation (10^3-10^4 sec). It is also a highly sensitive tool to characterize the quality of a protein preparation regarding purity and stability (Berkowitz 2006; Liu et al. 2006a; Gabrielson et al. 2007; Pekar and Sukumar 2007).

In analytical ultracentrifugation, the sedimentation takes place in free solution, in the absence of matrices or surfaces, and generally without the need for attaching labels. The data interpretation is based on first principles and can give absolute protein molecular weights and hydrodynamic translational frictional coefficients (leading to low resolution shape information of macromolecules in solution).

The principle of SV is very simple – the application of a high gravitational field to an initially uniform and well-mixed dilute solution of the protein sample, and the real-time imaging and mathematical analysis of the evolving concentration gradients arising from the centrifugal force. The resulting sedimentation boundaries show strongly size-dependent separation of protein species. In contrast, for example, to conventional size-exclusion chromatography, the faster migrating protein complexes will always remain in a bath of the slower sedimenting species, such that association and dissociation events from reversible complexes will continue to take place throughout the experiment, maintaining reversible complexes in a way that reflects the chemical equilibria and kinetic properties of the interaction. At the same time, different size complexes can be identified from the movement of their respective boundary. The latter aspects highlight why SV is uniquely suited for the study of non-covalent protein interactions.

As one of the key techniques in the development of physical biochemistry, analytical ultracentrifugation has a very rich history (Schachman 1959; Elzen 1988; Schachman 1989; Schachman 1992), that is far beyond the scope of the present introduction. The theoretical tools and the practice of SV have dramatically changed during the last decade with the introduction of modern computational approaches to the data analysis and the ability to routinely and rapidly solve the underlying equations. This has allowed the departure from previous strategies of applying data transforms to partial data sets of suitable properties, and the need to design experiments accommodating the analytical constraints, and in the process compromising resolution or other detailed information on the macromolecular sample under study. Instead, it is now possible to optimize the experiments to generate the most detailed information, and then to directly fit the observed data from the entire sedimentation process with a variety of models that explicitly embody different hypothesis about the macromolecular sample. This approach provides much more detailed and reliable information on the sedimenting sample and the protein interactions present. The latter approach of direct boundary modeling is now most widely used and discussed exclusively in this Unit. For current, more general reviews, see (Lebowitz et al. 2002; Balbo and Schuck 2005; Howlett et al. 2006; Scott and Schuck 2006; Schuck 2007b).

Due to the close relationship with SV it is also highly recommended for the reader to consult the Unit on Equilibrium Sedimentation (SE) [Equilibrium Sedimentation Unit in this Series]. SV and SE have many aspects in common besides the centrifuge and the capabilities of the detection system, such as sample preparation, prediction of protein buoyancy, buffer requirements for the different detection systems, range of binding constants that can be characterized, etc. Therefore, many cross-references will be made. At the same time, there are many differences arising from the observation of the equilibrium state *versus* processes far from it, respectively. The latter provides much higher resolution and sensitivity, but obviously poses a more complicated data analysis problem based on concepts different from those of SE. SV takes less time (typically several hours), generates higher centrifugal fields (typically a few 10^5 g) and higher pressures (up to 30 MPa), and requires slightly larger sample amounts (typically 0.1 – 1 milligram for a series of experiments). Insight can be gained from contrasting the different concepts from SV and SE, and it can be highly advantageous in the experimental study to exploit their complementarity. It is advantageous also to apply SV analytical ultracentrifugation in the context of other biophysical experiments, as it is an excellent complementary tool for all approaches that measure signals averaged over all species in the sample.

The present Unit is only a commentary that aims to serve as an introduction to selected topics of the state-of-the-art application of SV to the study of protein interactions, such as those frequently encountered in immunological systems. It is impossible to thoroughly discuss the multitude of possible variations of experiments and data analysis. Our goal is for the reader to gain conceptual understanding and experimental knowledge of the technique – together with the step-by-step protocol provided as an internet resource (Balbo et al. 2007) – sufficient for planning and conducting basic experiments and reliable data analysis for standard problems in homogeneous or heterogeneous protein interactions. A second goal is to provide a foundation for further reading on the methods presented, or on more specialized topics when faced with more intricate interacting protein systems. For the data analysis, we focus on our software SEDFIT and SEDPHAT, for which further tutorial material and a web-based, searchable help system are available at www.analyticalultracentrifugation.com. Workshops for the application of the practical and theoretical tools covered in this Unit are being held regularly in our laboratory at the National Institutes of Health, Bethesda, Maryland.

2. Basic Principles: Theory

In this section, the theoretical background of the most common analytical tools is briefly outlined. Although all the computational procedures of the advanced data analysis approaches are encapsulated in the software, eliminating the need to follow in detail the mathematical relationships, it is advisable for the researcher applying SV to have clarity about the concepts behind the approaches and their specific capabilities and strengths as well as limitations.

a) Single Ideally Sedimenting Protein Species

The sedimentation process is governed by the gravitational force $F_{sed} = m\omega^2 r$ (with m the protein mass, ω the rotor angular velocity, and r the distance from the center of rotation), the buoyancy force $F_b = -m\bar{v}\rho\omega^2 r$ (with $\bar{v}$ the protein partial-specific volume and ρ the solvent density) opposing the sedimentation, and the hydrodynamic friction $F_f = -fv = s(kT/D)\omega^2 r$ (with the frictional coefficient f, the linear migration velocity v, sedimentation coefficient s, the Boltzmann constant k, the absolute temperature T, and the diffusion coefficient D). The sedimentation coefficient s is defined as the ratio of absolute linear migration velocity v to the applied centrifugal field, $s = v/\omega^2 r$, and is a molecular constant.

The sedimentation coefficient is commonly measured in units of Svedberg, abbreviated S, which is related to the SI units by 1 S = 10^{-13} sec. From the balance of these three forces, one can derive the Svedberg equation

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

(with M denoting the protein molar mass, and R denoting the gas constant) (Svedberg and Pedersen 1940). The Svedberg equation is a fundamental relationship between the three directly measurable quantities for a single protein species: the sedimentation coefficient (in SV, governing the migration velocity of the sedimentation boundary with time), the diffusion coefficient (in SV, determining the evolution of the shape of the sedimentation boundary with time), and the molar mass (in SE, determining the steepness of the exponential equilibrium gradient after a long time [Equilibrium Sedimentation Unit]).

The sedimentation coefficient is directly related to the hydrodynamic translation frictional ratio f/f_0, and can be interpreted in terms of model shapes and compared to predictions from hydrodynamic theory for given structures (see below). For this purpose, it is frequently useful to transform the experimentally measured s-value, s_{exp}, to an equivalent s-value that would have been observed under the standard solvent conditions of water at 20°C, $s_{20,\mathrm{w}}$,

$$s_{20,w} = s_{\mathrm{exp}}\left(\frac{\eta_{\mathrm{exp}}}{\eta_{20,w}}\right)\left(\frac{1-\bar{v}\rho_{20,w}}{1-\bar{v}\rho_{\mathrm{exp}}}\right) \tag{2}$$

correcting for contributions arising just from buffer viscosity and density, η and ρ, respectively. The frictional ratio can then be obtained from the ratio of $s_{20,\mathrm{w}}$ to the s-value of a smooth, solid sphere $s_{sphere,20w}$ of the same molar mass and density. $s_{sphere,20w}$ reflects the highest theoretically possible sedimentation rate for any particle of the given mass and density:

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

and the hydrated frictional ratio f/f_0 follows as

$$f\,/\,f_0 = s_{sphere,20w}\,/\,s_{20w} \tag{4}$$

Typical values for the hydrated frictional ratio range between 1.3 for nearly globular hydrated proteins to 2.0 for very elongated or glycosylated proteins. This range of common f/f_0-values determines the s-values one would expect for proteins of given molar mass. *Vice versa*, if the molar mass is not known, it strictly cannot be determined from the s-value alone, but from the common ranges of f/f_0 a corresponding range of possible molar mass values can be deduced. Further, as will be outlined below, weight-average values of f/f_0 for the protein sample under study can be extracted from the diffusional spread of the experimental data using the $c(s)$ and $c(M)$ methods, providing molar mass estimates. The interpretation of the experimentally measured s-values and f/f_0-values in terms of low-resolution shape information will be discussed below.

In a microscopic picture, if it was possible to observe a sedimenting point particle that starts to sediment at the meniscus position r_m at time $t = 0$, it would assume a trajectory $r(t) = r_m \exp\left(s\omega^2 t\right)$. If we identify the 'position' of this particle with the mid-point of the sedimentation boundary, a crude analysis could proceed by determining s from the slope of a plot $\log(r(t)/r_m)$ versus t. However, while this approach is useful as a conceptual tool, modern approaches based on the complete, macroscopically measured concentration profile and its evolution with time are far superior with regard to the precision, detail, and flexibility.

In order to arrive at a model that directly describes the measured data, we require a theoretical description of the macroscopic concentration profiles. Sedimentation and diffusion fluxes determine the macromolecular concentration gradients that arise if an initially homogeneous, sector-shaped solution is placed in a centrifugal field. This leads to the Lamm equation for a single, ideally sedimenting species

$$\frac{\partial \chi}{\partial t} = \frac{1}{r}\frac{\partial}{\partial r}\left[rD\frac{\partial \chi}{\partial r} - s\omega^2 r^2 \chi \right] \qquad (5)$$

which predicts the time-course $\chi(r,t)$ of the macromolecular sedimentation (Lamm 1929). While historically the lack of a closed-form analytical solution of the Lamm equation in the radial geometry of Eq. 5 presented a major hurdle in sedimentation analysis, with modern numerical methods it can now be solved very efficiently (see, e.g., (Brown and Schuck in press) and the references cited therein). It predicts the experimentally observed concentration profiles for dilute macromolecular solutions with high accuracy within the limitations of the optical detection (see below). The sedimentation patterns strongly depend on the macromolecular size and on the rotor speed, and some examples are shown in Figure 1. In extension of Eq. 5, corrections can be made for density gradients arising from water compressibility, or sedimenting co-solutes that dynamically change the solution density and viscosity (Schuck 2004a; b). Importantly, it can also be extended to account for protein interactions, as will be discussed below.

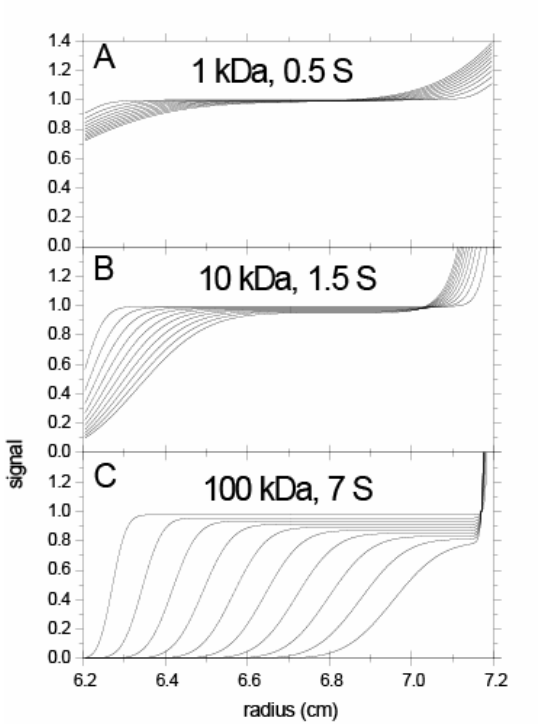

Figure 1: Characteristic shapes of sedimentation profiles for globular species of different size at a rotor speed of 50,000 rpm. Depicted are the concentration distributions in 10 min. intervals after start of centrifugation. It can be discerned that with increasing molar mass, the boundaries become steeper and the displacement with time increases. Larger size species than shown here can be easily accommodated, and typically a 1000-fold size range can be easily observed in a single experiment.

Due to practical limitations in the interference optical detection system, the experimentally measured data are the superposition of the signal arising from macromolecular redistribution plus a radial-dependent offset profile, termed TI ('time-invariant') noise as it is constant in time. This baseline profile arises from the radial dependence of imperfections in the optical elements. Further, a time-dependent, radially constant baseline offset, termed RI noise, arises for each scan due to the inability to assign absolute fringe shift numbers. Fortunately, it is possible to directly compute best-fit estimates of both the TI and RI signal components as part of the least-squares fit of the raw data (Schuck and Demeler 1999), thus fitting

$$a_{obs}(r,t) \cong \chi(r,t) + b_{TI}(r) + \beta_{RI}(t) \quad (6)$$

where $a_{obs}(r,t)$ is the actually measured signal, $\chi(r,t)$ the signal arising from macromolecular redistribution, and b_{TI} and β_{RI} the systematic noise offsets, respectively. Both types of offsets are routinely subtracted from the raw data (without changing statistical information content), in order to enable visual inspection of the experimental macromolecular sedimentation pattern $a_{obs}(r,t) - b_{TI} - \beta_{RI}$, which otherwise may be obscured (see Figure 12 below).

However, even with consideration of these unavoidable signal offsets, only in rare cases is it possible to fit the single-species Lamm equation Eq. 5 directly to experimental data. In theory, this would provide D, and together with the s-value define the molar mass. The pitfall is that any impurities, even at trace levels, will lead to excess boundary spreading. When misinterpreting the boundary as if arising from a single species, this will cause an overestimate of the diffusion coefficient and, consequently, and underestimate of the molar mass. Therefore, molar mass estimates from single-species Lamm equation fits can be considered only lower limits of the true value. Because it is usually very difficult to discern visually, and sometimes not trivial to unravel computationally, possible heterogeneity of the sample, the molar mass values should be trusted as good estimates of the true molar mass only if near perfect quality of fit is observed over the entire time-course of sedimentation (for fit criteria see below). Even if the only quantity of interest is an average s-value, the direct fitting of single-species Lamm equation is not the optimal approach, since it will result in an ill-defined average if impurities are unresolved.

b) Sedimentation coefficient and molar mass distributions

i) Sedimentation coefficient distribution c(s)

A much more detailed description of the sedimenting macromolecular mixture can be obtained with the model of sedimentation coefficient distributions. Here, many different size particles are considered with a continuous range of sedimentation coefficients. The concentration at each s-value, $c(s)$, is returned from the least-squares analysis of the experimental data:

$$a(r,t) \cong \int c(s)\chi_1(s, D(s), r, t)ds \quad (7)$$

where $a(r,t)$ denotes the experimental signal, $\chi_1(s,D,r,t)$ denotes the normalized Lamm equation solution for a single species (Schuck 2000). The units of $c(s)$ are such that integration over a peak gives the total signal of material sedimenting within the peak. In essence, the $c(s)$ method can be understood intuitively as matching the experimental sedimentation data with the best possible combination of sedimentation patterns from different-sized species (such as those shown in Figure 1). This is illustrated in Figure 2.

Integration of the $c(s)$ peaks leads to a well-defined and precise weight-average s-value (or signal-average s-value, respectively) of the species sedimenting within the integration limits (Schuck 2003). A few other important aspects of this approach are discussed in the following. For more detailed discussion of the general use and properties of this distribution, see (Schuck et al. 2002; Dam and Schuck 2004; Schuck 2006); software tutorials and example applications are available on the SEDFIT website (Schuck 2007c).

A naïve determination of $c(s)$ via Eq. 7 would be an ill-posed problem, i.e. many different solutions may fit the data similarly well, such that experimental noise may amplify and influence major features of the distribution (Provencher 1979). Therefore, it is essential that regularization, such as maximum entropy, be used in the computation. It provides as an output not the absolute best-fit solution, which may exhibit peaks not statistically reliably warranted by the data, but instead the simplest $c(s)$ distribution that that fits the data statistically indistinguishably well compared to the best fit. To this end, the user of SEDFIT is required to supply a pre-determined confidence level (P-value, usually set at 0.68) to determine the extent of regularization. Regularization is an implementation of the principle of Occam's

razor: it ensures that we interpret only those features of the distribution that are essential to explain the data, and are not misled by peaks arising from noise amplification. How to extend this method to enhance sensitivity and resolution is discussed further below.

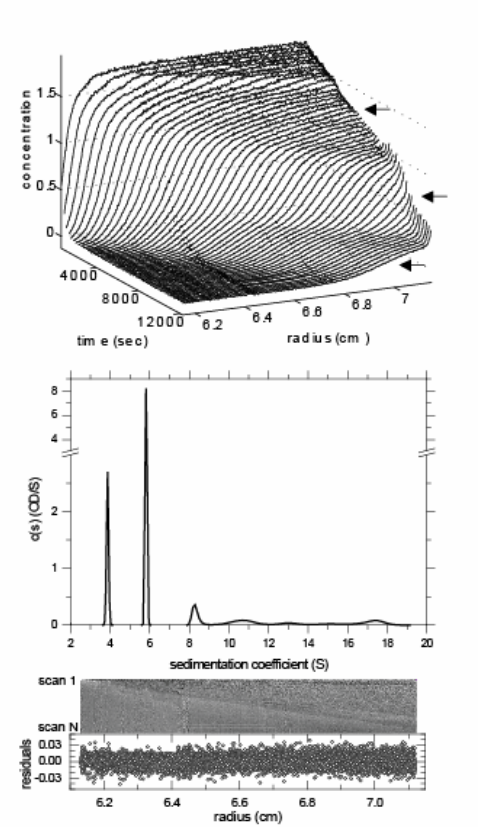

Figure 2: Experimental data and the corresponding c(s) sedimentation coefficient distribution. Top panel: Experimental data, showing the observed concentration profiles as a function of time. Systematic time-invariant and radial-invariant noise contributions calculated as part of the c(s) analysis were removed from the data for clarity. As indicated by the arrows, two main boundaries can be visually discerned, as well as relatively broadly distributed increasing signals at radii higher than the main boundary (the 'solution plateau'). The main boundaries correspond to two protein components that were mixed in this sample. Middle Panel: c(s) sedimentation coefficient distribution showing two distinct sharp peaks corresponding to the two main boundary components (from the individual monomeric protein species), and a relatively broad distribution of larger species formed by homo- and hetero-oligomers. Lower Panels: Residuals of the c(s) fit. The upper graphics is a bitmap representation, where the residuals of each scan at each radial point are depicted as lines of pixels with varying grayscale. The lower graph is an overlay of all data points as a function of radius. The bitmap representation reveals a slight diagonal feature arising from imperfections in the fit of the slower sedimenting species (in this case marginally tolerable; see the section 2.f of this Unit on Fit Criteria).

A second, conceptually very important aspect of the $c(s)$ distribution is that it reduces the two parameters s and D, which jointly determine the sedimentation behavior of an ideal species, to a distribution of only one parameter, s. This reduction can be achieved in different ways depending on the system under study, but most commonly it is realized by adopting a scaling law between s and D, which can be conveniently expressed via a weight-average frictional ratio $(f/f_0)_w$:

$$D(s) = \frac{\sqrt{2}}{18\pi} kT\ s^{-1/2} \left(\eta\left(f/f_0\right)_w\right)^{-3/2} \left(\left(1-\bar{v}\rho\right)/\bar{v}\right)^{1/2} \qquad (8)$$

(Schuck et al. 2002). $(f/f_0)_w$ is a scaling parameter that can be derived from the experimental data by non-linear regression, and is usually well-defined. Because $(f/f_0)_w$ only determines (and extracts information from) the diffusional spread, the precise choice of $(f/f_0)_w$ has little influence on the peak position and area in the $c(s)$ distribution, but mostly governs the resolution. Due to the relatively narrow range of f/f_0 values for folded proteins, it is usually a good choice for the study of protein mixtures.

ii) Variations of $c(s)$ tailored to specific systems or data

Other models for $D(s)$ are available that are more specifically tailored to certain systems. This includes a bimodal $c(s)$ model, where two separate $(f/f_0)_w$ values may be determined from data that exhibit visually separating sedimentation boundaries. This can be appropriate for mixtures of chemically dissimilar species, such as proteins and nucleic acids, or proteins and carbohydrates, as these species frequently

separate during sedimentation and exhibit different frictional ratios. (A related, but more general model is the hybrid discrete/continuous model in SEDPHAT, where the distribution can be designed as a combination of separate continuous segments and discrete species, see below.) Another specialized model is that for a protein of known molar mass undergoing conformational change. Further, for the study of filamenting proteins a model with the hydrodynamic scale-relationship of worm-like chains is available (Binger et al. submitted), as well as the possibility for an arbitrarily user-defined relationship between *s* and *M*. These variations and others become useful if particular knowledge on the interacting components and their behavior is available. But even in the absence of such refinements, for the study of protein mixtures and their interactions the standard *c*(*s*) method has proven to be very successful (Schuck 2007a).

iii) Molar mass distribution *c*(*M*)

Once the relationship *D*(*s*) has been established, for example via the quantity $(f/f_0)_w$, the Svedberg equation Eq. 1 can be used to scale *c*(*s*) into a molar mass distribution *c*(*M*). (For convenience, the keyboard shortcut control-M in SEDFIT will display the molar mass estimates for each peak.) However, care must be taken to interpret this value: It is susceptible to the precise value of f/f_0 for each species, and only if the distribution exhibits a single major peak, and when the correct values for the protein partial specific volume $\bar{v}$ and the solvent density ρ and viscosity η under the experimental conditions are used, it will be a good estimate. Typically, the precision of *c*(*M*) in this case is better than ±10%. This is usually sufficient to identify the protein oligomeric state in solution. This was illustrated, for example, in the study of the oligomeric state of a natural killer cell receptor fragment in (Dam and Schuck 2004). *c*(*M*) can be particularly useful if the protein sample exhibits microheterogeneity, such as from glycosylation or from conformational mixtures (Kornblatt and Schuck 2005). If, on the other hand, microheterogeneity is absent, a powerful approach to measure species molar masses is the hybrid discrete/continuous distribution in SEDPHAT, where the *c*(*s*) peaks are replaced with discrete single-species Lamm equation solutions with independently adjustable molar mass values (Boukari et al. 2004; Greive et al. 2005; Chou et al. 2006).

iv) Size-and-shape distributions

A generalization of the sedimentation coefficient distribution *c*(*s*) is the two-dimensional size-and-shape distribution c(*s*,*M*) (Brown and Schuck 2006). Here, no relationship *D*(*s*) is required, and both parameters can freely and independently vary (Figure 3):

$$a(r,t) \cong \int c(s,M)\chi_1(s,M,r,t)\,ds\,dM \qquad (9)$$

This distribution can be expressed in different coordinates, since the parameters *s* and *M* allow the transformation into a distribution of Stokes radii R_S, diffusion coefficients *D*, and frictional ratios f/f_0. Such a distribution is useful, for example, when dealing with chemically dissimilar species or mixtures of proteins that are known to exhibit a significant, possibly continuous variation in frictional ratio. Because the information content of the SV data may not be sufficient to define features in both dimensions well, the distribution can be condensed back to a sedimentation coefficient distribution free of any scale-relationship, termed 'general *c*(*s*,*)'

$$c(s,*) = \int c(s,M)\,dM \qquad (10)$$

which gives the population of species sedimenting at a given s-value irrespective of their frictional coefficients (or molar masses). Equivalent integrations along other coordinates can give similarly general molar mass distributions, Stokes radii distributions, etc. General *c(s,*)* should be used if none of the

available models for $D(s)$ seems suitable.

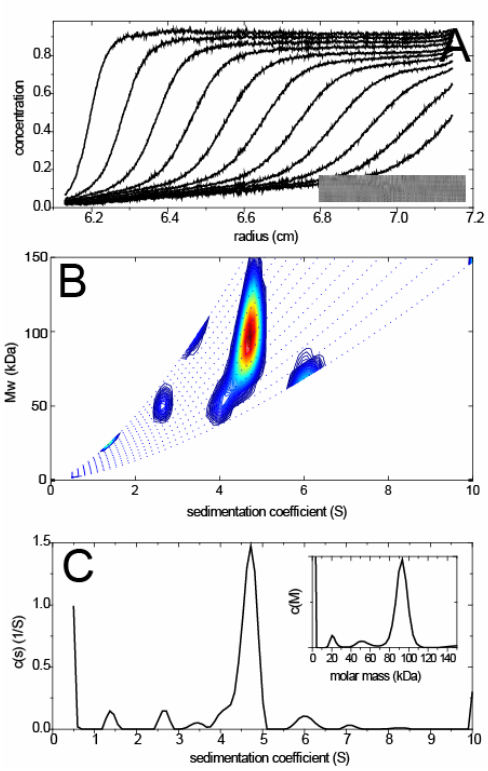

Figure 3: Analysis of experimental sedimentation velocity data from the study of the oligomeric state of a glycosylated NK receptor fragment (Dam and Schuck 2004; Brown and Schuck 2006). (A) Representative subset of the raw data after elimination of systematic noise contributions. The inset shows a residuals bitmap of the size-and-shape c(s,M) analysis. A naïve single species analysis (Figure 13) would lead to a molar mass close to the monomer ~ 60 kDa, an estimate qualitatively in error due to the presence of impurities. (B) 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 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. (C): 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,*).

v) Multi-signal sedimentation coefficient distribution $c_k(s)$

The multi-signal $c_k(s)$, can be highly useful in practice when studying interactions of proteins with significantly different spectral properties (Balbo et al. 2005). In this method, data are acquired from the same cell at multiple signals (e.g., wavelengths) $a_\lambda(r,t)$, and both diffusion and the spectral contributions are simultaneously deconvoluted as

$$a_\lambda(r,t) \cong \sum_{k=1}^{K} \varepsilon_{k\lambda} \int_{s_{min}}^{s_{max}} c_k(s)\chi_1(s, F_{k,w}, r, t)ds \qquad (11)$$

(with $\varepsilon_{k\lambda}$ the extinction coefficient of component k at wavelength λ, predetermined in a separate SV experiment, and $F_{k,w}$ the weight-average frictional coefficient). This results in a high-resolution sedimentation coefficient distribution for each component, which gives information how much of each protein is sedimenting in a certain s-value peak. This directly reports on the stoichiometry of the complexes formed. It has the virtues of excluding effects of impurities from contamination and degradation products from biasing the estimate of complex stoichiometry, and of not depending on absolute knowledge of the concentrations of the active species. As will be discussed in more detail below, for reversible complex formation it does require either the complex to be kinetically stable on the time-scale of sedimentation (k_{off} < 0.001/sec), or concentrations higher than K_D to be used for at least one of the components in order to populate and hydrodynamically separate the complex.

The spectral requirements can be met by attaching extrinsic chromophoric labels. However, in many cases intrinsic differences between proteins in their content of aromatic amino acids are generating sufficiently characteristic ratios of 280 nm absorbance *vs.* interference signal. Clearly, protein interactions with other polymers or compounds that do not absorb strongly in the UV (as carbohydrates

and many detergents), or that dominate the UV absorbance (such as nucleic acids) also provide natural applications for multi-signal analyses.

While this multi-signal $c_k(s)$ method is a convenient approach, for example, in cases where the s-value or molar mass of a complex alone would not provide unambiguous stoichiometry information, this approach is uniquely suited to unravel the number of different classes of complexes and their stoichiometry for extended, multi-step assembly processes of two or three different protein components (Figure 4). Examples for the application of $c_k(s)$ in the literature can be found in (Minor et al. 2005; Houtman et al. 2006; Deka et al. 2007).

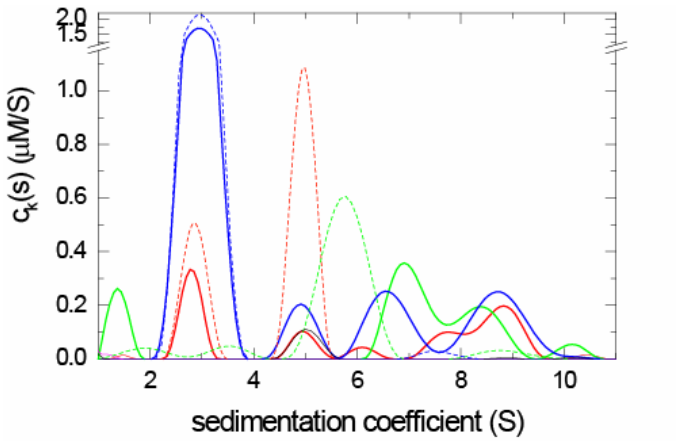

Figure 4: Example of the multi-signal $c_k(s)$ analysis of a triple protein mixture of a viral glycoprotein (green), its cognate receptor (blue), and a heterogeneous antigen-recognition receptor fragment (red). The content of each protein component in the different s-ranges is obtained from the global 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 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 dotted lines. The formation of coexisting binary complexes (~5 S, ~7 S) and a ternary complex with 1:1:1 stoichiometry (~ 8.5 S) can be discerned.

vi) Incorporation of prior knowledge in regularization

The sensitivity and resolution of the distributions described above can be enhanced significantly by using a Bayesian extension of the regularization principle described above to incorporate knowledge that is available *a priori* (Brown et al. 2007). The Bayesian approach is a very powerful tool to probe alternative explanations of the data, in the light of certain prior hypotheses (Sivia 1996).

As described above, in the standard regularization we extract the essential pieces of information of the experiment by calculating the 'simplest' distribution that can fit the data statistically indistinguishably well compared to the overall best fit (avoiding the susceptibility of the latter to artificial peaks from noise amplification in the ill-conditioned analysis). The conventional regularization in $c(s)$ embodies the notion that the ideal 'simplest' distribution (which would be returned in the absence of data) be flat and featureless, and only with increasing signal/noise ratio of the data will peaks emerge at specific locations with increasing sharpness. The Bayesian extension implemented in SEDFIT and SEDPHAT, termed $c^{(P)}(s)$, allows the researcher to redefine the 'simplest' distribution to be one that conforms completely to the prior expectation. With increasing experimental information will the distribution deviate from the prior expectation, but only to the extent necessary, again, to fit the data statistically indistinguishably well (compared to both the instable best-fit, and to the solution from conventional regularization).

It is important to note that the distribution $c^{(P)}(s)$ will generally not simply reflect the prior expectation, as would be the case for ordinary constraints, but show in more detail which additional unexpected features are necessary to explain the data. An specific example of prior knowledge, termed $c^{(P\delta)}(s)$, is that the sample may not be pure, but that all species are intrinsically mono-disperse (due to the discrete nature of protein masses) (Figure 5). A different example for interacting heterogeneous systems is the introduction of the s-values of the individual species (from a separate experiments in the same run) (Figure 6). For more details, see (Brown et al. 2007).

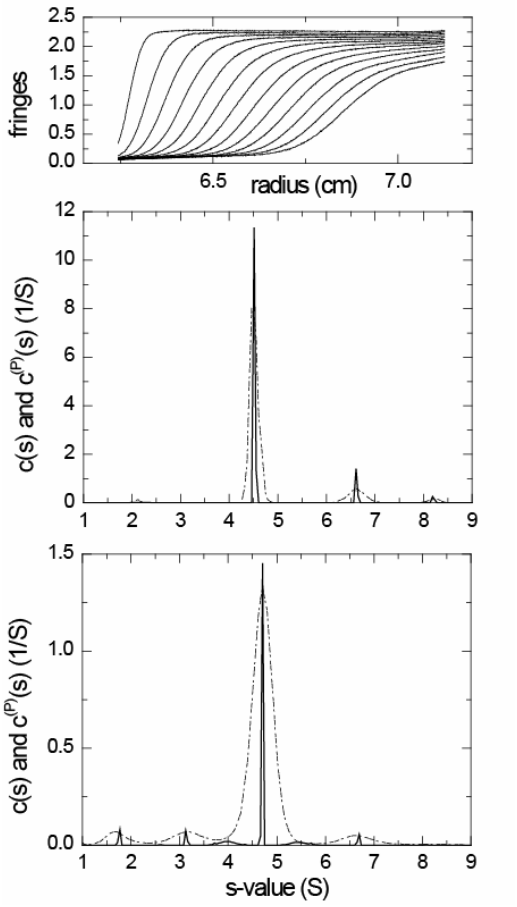

Figure 5. Prior knowledge in the analysis of non-interacting macromolecules. Top: SV data from a BSA sample. Middle: The BSA data are analyzed first with conventional c(s) (dotted line). Generally, the peak width in c(s) can be a result from either a true polydispersity of the protein (e.g., strong heterogeneity in glycosylation, in conformation, primary sequence, etc.), or from the standard regularization favoring broader peaks for data with low signal/noise ratio. As a second stage, the $c^{(P\delta)}(s)$ distribution is calculated (solid line; SEDFIT keyboard shortcut control-X). Here, the prior knowledge is used that each species that occurs would have a discrete molar mass and s-value. This knowledge may be derived from general expectation of the discrete nature of biological macromolecules, or be confirmed by mass spectroscopy. Thus, all peaks from the first c(s) distribution transformed into δ-peaks and used as Bayesian prior expectation for calculating the $c^{(P\delta)}(s)$ distribution. As expected, the same peaks occur in c(s) as in $c^{(P\delta)}(s)$, with the latter being significantly sharper. Bottom: The same procedure applied to the data from the glycosylated NK receptor fragment shown in Figure 3. Here, it should be noted that in $c^{(P\delta)}(s)$ two new smaller peaks at 4 S and 5.5 S emerge. Their presence means that they are essential part of a model containing a sharp main 4.7 S peak. The smaller peaks could be a compensation for an unaccounted true heterogeneity within the main peak, or reflect true sedimenting species previously unresolved. Both c(s) and $c^{(P\delta)}(s)$ have the same high quality of fit and are indistinguishable based on the experimental SV data alone. The possibility of their distinction resides only in considering the confidence in the prior knowledge. In the present case, strong heterogeneity of the glycosylated receptor, as detected by mass spectroscopy, would make the expectation of a sharp main peak not a safe assumption. This example highlights how the Bayesian analysis can be used to actively explore the flexibility of the data interpretation, not by force-fitting models with different constraints, but more subtly by allowing the data to compensate, if necessary, for the bias introduced via the prior expectation, such that highly detailed alternative pictu emerges, for which the same quality of fit is ensured. For a detai discussion, see (Brown et al. 2007).

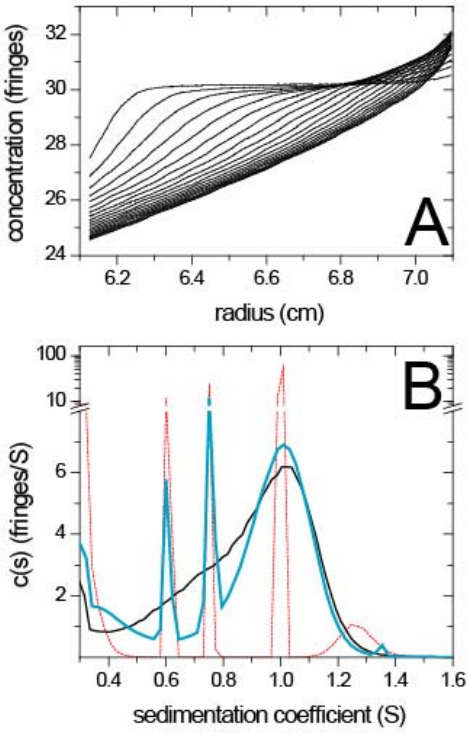

Figure 6. Prior knowledge in the analysis of interacting macromolecules forming dynamically stabilized complexes. (A) Interference optical sedimentation velocity data of a mixture of peptides derived from the adaptor protein SLP-76 (11.7 kDa) and PLC-γ (7.4 kDa) which form complexes with 1:1 stoichiometry (Balbo et al. 2005; Brown et al. 2007). (B) The conventional sedimentation coefficient distribution c(s) is shown as black solid line. Using the prior knowledge of the s-values of the individual species (~0.6 S and ~0.75 S, obtained from experiments conducted in different cells of the same run), the resulting distribution $c^{(P)}(s)$ is shown as blue line, exhibiting a well-defined peak for the complex formed. In a third step, it is hypothetically assumed that the complex peak arises from one single species, expressed as the prior expectation that the complex peak should have an unchanged signal-average s-value, but be sharp. The resulting distribution $c^{(P\delta)}(s)$ is shown as red dotted line. It can be discerned that with this prior knowledge, an additional peak at ~ 1.25 S would be necessary to explain the data. As a consequence, the data suggest that there is not a single mono-disperse, stably sedimenting complex species, contrary to expectations. Since both the blue and the red distributions are of indistinguishable quality of fit, it is not possible to decide from the data alone whether the blue peak is a better reflection of the true situation of a sedimentation coefficient distribution arising, for example, from dynamic instability of the complex on the time-scale of sedimentation, or from conformational heterogeneity, or if the red distribution is reflecting a truly existing stable, mono-disperse complex population but in the presence of slight further aggregation. By using the hypothetical prior expectation one can probe the set of possible, alternative explanations of the data.

b) Interacting systems

Mixtures of interacting proteins sediment characteristically different from non-interacting mixtures, in a way that depends on the life-time of the complexes relative to the time-scale of the SV experiment. Even though the data set from a single sample will comprise information from a range of sample concentrations between zero and the loading concentration, the features characteristic for interactions are best recognized and best quantified when running mixtures at several different loading concentrations and/or molar ratios in order to explore the entire binding isotherm by shifting the relative population of free and complexes.

i) Lamm equation modeling and *c*(*s*) analysis

From a strictly theoretical perspective the most attractive approach to characterize protein interactions in SV would be to globally fit the complete sedimentation profiles obtained at all concentrations directly with numerical solutions of the coupled Lamm equations of the interacting system

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

where q_i denote s the time-dependent local reaction flux. The precise form of Eq. 12 and, in particular, its reaction fluxes requires specification of the number and stoichiometry of species. It can be implemented either for instantaneous equilibria following mass action law, or considering finite reaction kinetics (Urbanke et al. 1980; Schuck 2003; Stafford and Sherwood 2004; Dam et al. 2005). In the kinetic case, for example, for a simple bimolecular reaction to form a 1:1 complex, q_i becomes $q_1 = q_2 = -q_3 = -q$ with $q = k_{on}\chi_1\chi_2 - k_{off}\chi_3$. Eq. 12 is available in SEDPHAT for global fitting of experimental SV profiles for a number of binary and ternary interaction models.

However, several practical reasons make this approach more difficult than it may at first appear: (i) It requires prior knowledge on the number and stoichiometry of the complexes formed, which is frequently far from trivial, and should be assessed prior to attempting Lamm equation modeling, for example, using *c*(*s*) or multi-signal $c_k(s)$ (see below). (ii) Different chemical off-rate constants influence the sedimentation patterns of reacting systems significantly only in a relatively narrow range between ~ 10^{-4}/sec – 10^{-3}/sec, making this a parameter poorly determined by the data. This problem is exacerbated for small proteins that exhibit larger diffusional spreading of the boundaries. (iii) Similar to the drawbacks of fitting single-species Lamm equations to putatively pure protein samples, the results of modeling with Eq. 12 will be very susceptible to bias from unrecognized contaminating species. (iv) It has been pointed out that microheterogeneity either of the binding properties (Cann 1982), or arising from conformational mixtures (Werner and Schachman 1989), or from heterogeneous glycosylation will lead to excess boundary spread, which is difficult to incorporate into models of the type Eq. 12 (Dam et al. 2005).

Modeling data directly with explicit Lamm equations for reacting systems (Eq. 12) was successfully applied in many studies (Lewis et al. 2002; Ali et al. 2003; Dam et al. 2005; Hsu et al. 2005; Chan et al. 2006; Connaghan-Jones et al. 2006; Schmeisser et al. submitted). However, this approach is typically applied after a more robust quantitative analysis step that precedes it and helps to initialize its parameters.

As a first step in the characterization of protein interactions, we recommend the application of the *c*(*s*) analysis to the data obtained at a range of loading concentrations. This allows one to diagnose the behavior of the interacting system. Complex formation is indicated by the appearance of new peaks and/or the concentration-dependent shift of the peak positions, as will be described in more detail below. From the comparison of the resulting distributions with the distribution from the samples studied separately, and from comparison among the different mixtures, a number of conclusions can be drawn.

Compared to the Lamm equation modeling with reaction terms, the quantitative analysis of sedimentation coefficient distributions for deriving the binding constants is much less susceptible to sample impurities and aggregates. The c(s) approach to interactions has the drawback of not delivering more than qualitative information on kinetic rate constants, but the advantage of being robust against microheterogeneity, impurities and degradation products from proteolytic activity or aggregation.

ii) Sedimentation coefficient distribution analysis of interacting systems with slow kinetics

Identifying interactions with slow chemical conversion ($k_{off} < 10^{-3} - 10^{-4}$/sec) is straightforward (Figure 7). In this case, all sedimenting species are sufficiently stable to hydrodynamically separate, and the $c(s)$ peaks directly reflect the sedimenting species. An important consequence of slow chemical conversion is that the peak positions do not change with loading concentration. In diagnosing this case, allowance must be made for the fact that the data acquired at lower concentrations will have lower signal-to-noise ratio, which can cause peaks to merge in $c(s)$. What changes with concentration are the relative peak areas, which represent the shifting populations as predicted by mass action law.

Two pieces of information can be extracted from the $c(s)$ curves: 1) The area under each peak should be determined by integration, and the isotherms of their concentration-dependent shift in population be modeled directly with mass action law. For example, for a bimolecular reaction A+B ←→ AB, the isotherms of the signals from the individual species (a_A, a_B, and a_{AB}, respectively) take the form

$$\begin{aligned} a_A(c_{Atot}, c_{Btot}) &= \varepsilon_A \left(c_{Atot} - c_{AB}\right) \\ a_B(c_{Atot}, c_{Btot}) &= \varepsilon_B \left(c_{Btot} - c_{AB}\right) \\ a_{AB}(c_{Atot}, c_{Btot}) &= \left(\varepsilon_A + \varepsilon_B\right) c_{AB} \\ c_{AB} &= \frac{1}{2K}\left[1 + K\left(c_{Atot} + c_{Btot}\right) - \sqrt{1 + 2K\left(c_{Atot} + c_{Btot}\right) + K^2\left(c_{Atot} - c_{Btot}\right)^2}\right] \end{aligned} \qquad (13)$$

(with extinction coefficients ε_A and ε_B and the total loading concentrations c_{Atot} and c_{Btot} for components A and B, respectively, and the association constant K)

2) The shift in relative population causes an increase in the weight-average s-value as a function of loading concentration, $s_w(c)$. The quantity s_w can be quantified for each distribution by integration of the $c(s)$ distribution over all peaks (but excluding peaks from contaminating species). There is a deep connection between s_w and mass balance considerations (Schachman 1959), as judged by the change in area under the sedimentation boundaries

$$s_w = -\frac{1}{\omega^2 r_p^2 c_p} \int_{r_m}^{r_p} c(r,t) r dr = \int_{s_{min}}^{s_{max}} c(s) s ds \Big/ \int_{s_{min}}^{s_{max}} c(s) ds \qquad (14)$$

(with r_p representing an arbitrarily chosen radius in the plateau region, with the plateau concentration $c_p(r_p)$, and with the meniscus position r_m). s_w can be calculated by integration of $c(s)$ or any other boundary model that faithfully represents the boundary area, irrespective of the details of the model itself (Schuck 2003). In the above example of Eq. 13, the isotherm of weight-average (or, more precisely, signal-average) s-value versus sample composition is

$$s_w(c_{Atot}, c_{Btot}) = \frac{a_A s_A + a_B s_B + a_{AB} s_{AB}}{a_A + a_B + a_{AB}} \qquad (15)$$

Generally, an increase of $s_w(c)$ at higher concentrations is a highly sensitive and powerful proof for the existence of an interaction – irrespective of the reaction scheme, affinity, or kinetics – and permits quantitation of even very weak interactions (Patel et al. 2007).

For convenience, the types of isotherms Eq. 13 and Eq. 15 are implemented in SEDPHAT for a variety of different interaction models, and for global fitting of multiple data sets.

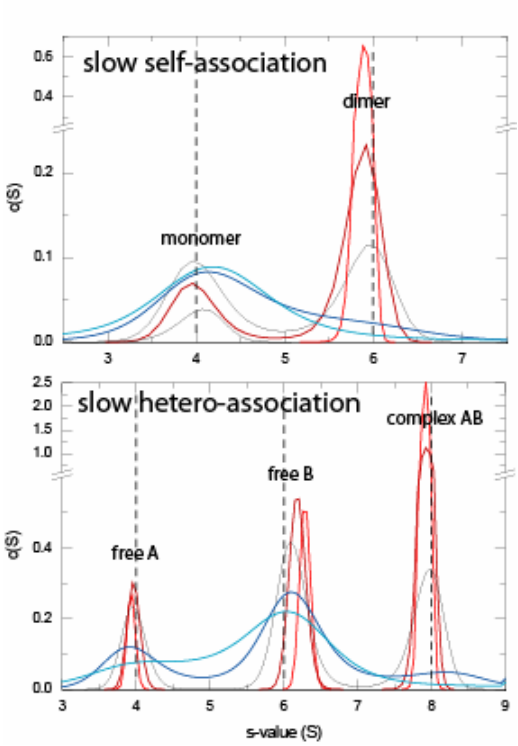

Figure 7: Examples of c(s) distributions for slowly interacting systems. Shown are (scaled) distributions from a simulated self-associating system of a 50 kDa protein sedimenting at 4S forming a dimer with 6 S (top) and a heterogeneous association of a 50 kDa (4 S) and a 100 kDa protein (6 S) forming a 1:1 complex with 8 S (bottom), both sedimenting at 50,000 rpm. For both systems, the binding constant K_D was assumed to be 1 μM, $k_{off} = 5\times10^{-5}$/sec, and protein concentrations were 0.1 (light blue), 0.3 (dark blue), 1 (black), 3 (dark red), and 10 μM (red), with equimolar concentrations for the hetero-association. For slow reactions, the relative peak areas change but the peak positions remain virtually constant. The peaks directly represent the number of sedimenting species. However, the finite reaction rates will make the peak s-values slightly different form those of the underlying species. At low protein concentration, the lower signal-to-noise ratio of the data will cause the c(s) peaks to be of low resolution and even merge. (This may be improved in a second stage using $c^{(P\delta)}(s)$.)

ii) Sedimentation coefficient distribution analysis of interacting systems with fast kinetics

For systems with rapidly interconverting free and complex species ($k_{off} > 10^{-3}$/sec), a different pattern appears with peaks shifting in a concentration-dependent fashion: The interacting mixture exhibits a coupled sedimentation of free species and complex, with an s-value that reflects the fractional time in which, on average, the molecules spend in the complex state (sedimenting fast), as opposed to the dissociated state (sedimenting slow). This is termed the 'reaction boundary'. The precise boundary shape and position is complicated by the fact that within the sedimentation boundary the concentrations of all components change, such that a certain dispersion of the s-values naturally arises. As shown elsewhere (Dam et al. 2005), there is a deeper connection between the sedimentation coefficient distribution $c(s)$ and reaction boundaries, which can be understood on the basis of the constant bath approximation of the Lamm equation for reacting systems, which predicts the reacting boundary to diffuse, in a first order approximation, with a single intermediate diffusion coefficient (Krauss et al. 1975; Urbanke et al. 2005). Reaction boundaries can be described in good approximation by $c(s)$ (Dam and Schuck 2005; Brown et al. 2007), with the limitation that residually unaccounted coupling of the diffusion and reaction fluxes tends to produce underestimates for the weight-average frictional ratio f/f_0 (particularly at concentrations ~ K_D). Thus, the quality of $c(M)$ estimates from reaction boundaries is reduced (with M generally being underestimates), but, for concentrations >> K_D may still allow the gross estimate of the average molar masses, frequently revealing the oligomeric state. However, the strength of the application of $c(s)$ to reacting systems is that it can still deconvolute diffusional broadening and it represents well the

underlying sedimentation coefficient distribution. This is highly useful for the quantitative thermodynamic analysis of interactions.

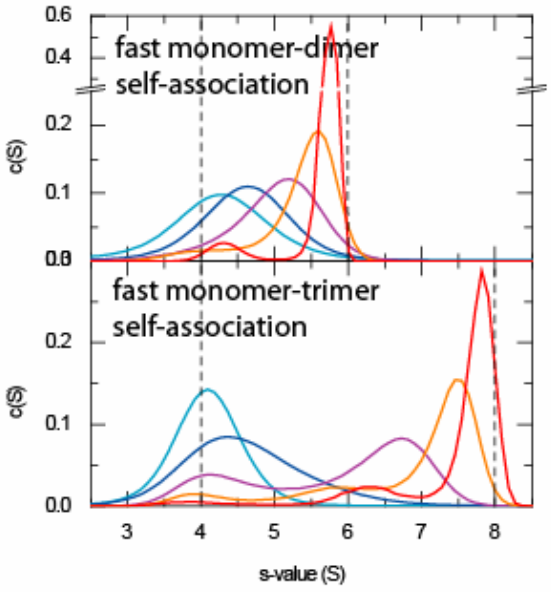

Figure 8: Examples of c(s) distributions for rapidly interacting self-association. Shown are (scaled) distributions from a simulated self-associating system of a 50 kDa protein sedimenting at 4S forming a dimer with 6 S (top), and the same monomer forming a trimer at 8 S, respectively. c(s) curves are calculated for protein concentrations of 0.1 (light blue), 0.3 (dark blue), 1 (violet), 3 (orange), and 10 μM (red), at association constants of 10^6/M for the monomer-dimer and 10^{12}/M for the monomer-trimer system, respectively (both having an effective K_D of 1 μM). For both systems, the peak position changes strongly with concentration in the vicinity of K_D. Characteristic for the higher-order self-associations (n > 2) is that the s-range is more drawn out, and that a distinct monomeric peak can be discerned at all concentrations. This is due to the higher difference of the s-values for oligomers with n > 2, causing stronger separation and a distinctly slower sedimenting trailing part of the sedimentation boundary at concentrations < ~0.3 K_D, which is dominated by the monomeric form. It should be noted that even at concentrations 10fold K_D the peak position of the main peak does poorly reflect the s-value of the highest oligomer.

For rapidly interacting systems, it is indispensable that they be studied at a range of concentrations. The major characteristics of $c(s)$ curves from interacting systems is the change of peak positions with sample concentration and composition. Examples for rapid self-associations are shown in Figure 8. For the quantitative analysis of self-associations, we extract only the $s_w(c)$, and fit them to expressions analogous to Eq. 15. In the analysis, care should be taken to examine samples at a wide concentration range (desirable are 2 orders of magnitude), in order to allow the separate characterization of the monomer s-value and oligomer s-value, as well as the binding constant. It should be noted that very low concentrations giving signals on the order of the noise of data acquisition can be sufficient in $c(s)$ to determine s_w. If only a narrow concentration range is available, these values are highly correlated (Schuck 2003). Alternatively, in some cases it is possible to constrain the s-values of the monomer or the oligomer to those measured independently with protein variants that show much stronger or much weaker self-association, respectively, or which show slower self-association kinetics under some conditions (Buisson et al. 2001). Further possibilities are the theoretical prediction (see below) of the monomer or oligomer s-value from known crystal structures (Dam et al. 2003), if available, or the assumption of hydrodynamic scaling laws (Frigon and Timasheff 1975).

Heterogeneous interactions with rapid interconversion of species can show more complex patterns. They are easier to study due to the fact that two components A and B (assuming bimolecular reactions) can be varied separately, such that characteristic patterns of sedimentation boundaries appear (Figure 9). Besides the reaction boundary, for bi-molecular heterogeneous associations a second peak is present at lower s-values, coincident with the s-value of either one of the free components. It is caused by the trailing part of the boundary where only one single component exists. Usually, but not always, this is the slower-sedimenting of the two components. Sometimes, when the molar ratio of the larger to the smaller species is increased, the species that constitutes this 'undisturbed' boundary switches from A to B within a concentration series. Also, while for self-associations the s-value of the reaction boundary monotonically increases with concentration, this is not necessarily the case for hetero-associations,

dependent on the relative concentrations of A and B in a concentration series. (However, s_w always increases with concentrations for interacting systems.) See Figure 9 for examples.

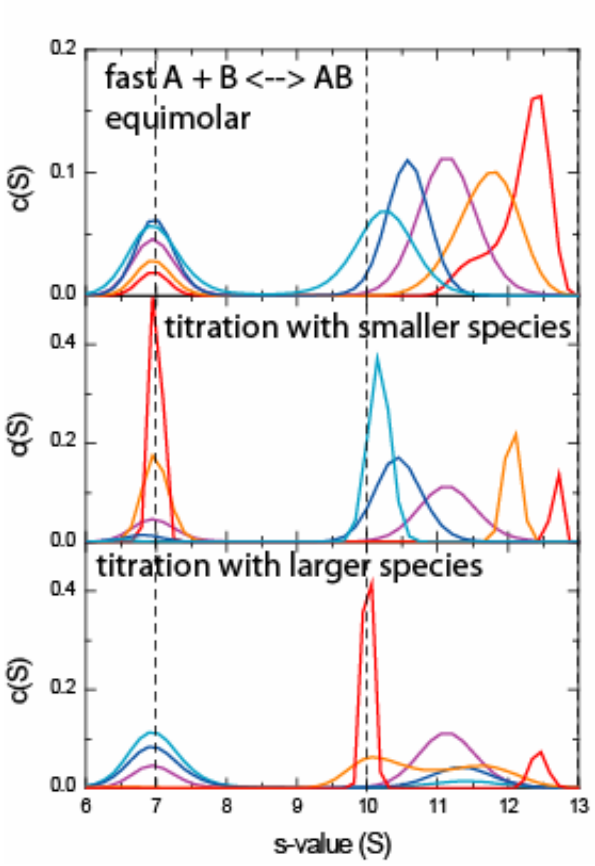

Figure 9: Examples of c(s) distributions for rapidly interacting hetero-associating systems A + B ←→AB. Shown are (scaled) distributions from a simulated interaction of a 100 kDa protein 'A' sedimenting at 7 S and a 200 kDa protein 'B' sedimenting at 10 S forming a complex with 13 S. c(s) curves are calculated for different experimental configurations, mimicking either equimolar concentrations (top), the titration of constant B with increasing A (middle), and the titration of constant A with increasing B (bottom), with protein concentrations of 0.1-fold K_D (light blue), 0.3-fold K_D (dark blue), K_D (violet), 3-fold K_D (orange), and 10 -fold K_D (red). In any configuration, the peak position of the reaction boundary changes strongly with concentration. For the equimolar case and for the case of variable A at constant B, the undisturbed boundary is at the s-value of the slower sedimenting A, and the s-value of the reaction boundary is continuously increasing with concentration. This is not the case, however, for the titration of constant concentration of the smaller species A with increasing concentrations of the larger species B (bottom): here, although initially all of B is in the reaction boundary, at a molar excess of B greater than ~3-fold, the larger species B constitutes the undisturbed boundary and the peak at the s-value of the smaller A disappears. During this switchover, the concentration dependence of the s-value of the reaction boundary changes direction: being initially halfway between the s-value of B and the complex, increasing concentrations of B first draw it towards that of B, and then closer to the s-value of the complex. The precise s-values and the areas under the peaks can be modeled with predictions from Gilbert-Jenkins theory, as implemented in SEDPHAT. An example is shown in Figure 10.

The precise s-value and relative areas of the different peaks of the resulting sedimentation coefficient distribution may not appear intuitive at first, but one can rely on the detailed description of the reacting system by Gilbert-Jenkins theory of coupled migration in the presence of rapid chemical reactions (Gilbert and Jenkins 1956). The latter is beyond the scope of the current introduction, but in essence predicts the velocity distribution of the reaction boundary by solving the problem of how much of each species is required for a co-sedimenting reacting system to assume a given velocity v (in the approximation of a linear geometry and no diffusion)

$$(v - v_a)\frac{\partial c_A}{\partial v} = (v - v_b)\frac{\partial c_B}{\partial v} = -(v - v_c)\frac{\partial c_C}{\partial v} \qquad (16)$$

with the constraints of mass conservation and mass action law for the given loading concentrations (Gilbert and Gilbert 1978). Assuming sedimentation of an infinitesimal amount of such a system, the remaining total concentrations are slightly depleted in the trailing boundary portion, leading to a different composition of the system sedimenting with a slightly lower velocity, and so on, ultimately giving rise to a sedimentation velocity distribution for each species. From this, the shape and amplitude of the reaction boundary as well as the amplitude and s-value of the undisturbed boundary follows. For a more detailed discussion, see (Gilbert and Jenkins 1959; Fujita 1975; Dam and Schuck 2005).

For the data analysis of fast heterogeneous associations of two proteins, three pieces of information can be extracted: 1) The area under each peak as a function of loading concentrations constitutes an isotherm that can be fit with expressions from Gilbert-Jenkins theory for the relative magnitudes of the 'undisturbed' and the 'reaction' boundary. 2) The concentration-dependent s-value of the 'reaction' boundary forms an isotherm that can also be fit with predictions of Gilbert-Jenkins theory. This is highly useful to predict the s-value of the complex species. 3) The concentration-dependence of the weight-average s-value, as determined by integration over all peaks (except known impurities or aggregates outside the range of s-values of the interacting systems). Analogously to the treatment of slowly interacting systems, these isotherms are subjected to global fitting in SEDPHAT. An example is shown in Figure 10.

It is worthwhile recapitulating where in these analyses the information is taken from. As described above, the analysis of the weight-average sedimentation coefficient $s_w(c)$ is based on mass balance considerations – essentially using the information on the depletion of total material due to sedimentation with time. The reason for the robustness of this information is that it resides in the area under the concentration profiles and is little affected by reaction kinetics and not at all by diffusion (Schuck 2003). Going further, the analyses of the individual peak areas, as well as the analysis of the s-value of the reaction boundary, takes advantage of the underlying multi-modal nature of the boundaries and their associated sedimentation coefficients. In many cases, the bimodal nature of sedimentation patterns can even be clearly visibly discerned in the raw data, although in some cases it becomes apparent only after diffusional deconvolution in $c(s)$. The s-values and the relative amplitudes of these boundary components also are robust features of the data. What is left out of this analysis is the detailed interpretation of the boundary spread, which is most influenced by microheterogeneity and impurities. However, qualitative information about the reaction rate constants is obtained by the distinction between the patterns of 'slow' and 'fast' reactions.

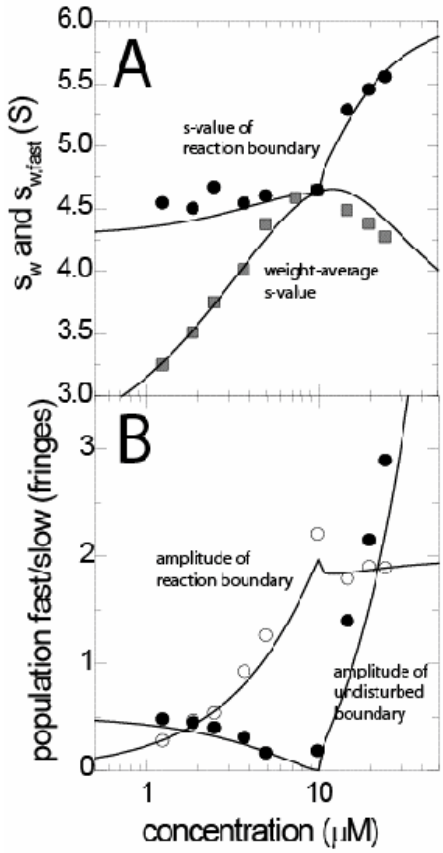

Figure 10: Analysis of experimental data from the sedimentation of a natural killer cell receptor Ly49C (31 kDa, 2.66 S) interacting with a MHC molecules H-2Kb (45 kDa, 3.56 S) sedimenting at 50,000 rpm (Dam et al. 2003). From the c(s) analysis, a bimodal boundary shape was detected characteristic for rapidly reacting two-component heter-associations. Extracted from the data were the concentration-dependence of the overall weight-average s-value $s_w(c)$ (Panel A, gray squares), the s-value of the reaction boundary $s_{fast}(c)$ (Panel A, black circles), as well as the amplitudes of the undisturbed and the reaction boundaries (Panel B, solid and empty circles, respectively). The isotherms were fitted globally with SEDPHAT, based on theoretical expressions of a rapid interaction with two equivalent non-interacting sites with K_D = 1.7 μM, indicated by the solid lines, with s-values of 4.96 S and 6.11 S for the 1:1 and 2:1 complexes. For a detailed description of this system and data analysis, see (Dam et al. 2003; Dam et al. 2005; Dam et al. 2006).

As introduced above, the multi-signal sedimentation coefficient distribution $c_k(s)$ can be applied to the study of heterogeneous interactions when the interacting components have different optical properties that can be detected by combined data acquisition using the interference and absorbance system, and/or the absorbance system at multiple wavelengths. The global $c_k(s)$ of the different signals displays information on the composition of different peaks. For slowly interacting systems, and for rapidly interacting systems either at concentrations >> K_D or with one component in > 5-fold molar excess (Balbo et al. 2005), the composition of the reaction boundary directly reports on the complex stoichiometries. This can be highly useful or indispensable for determining the binding model, in particular for multi-step, extended associations. However, once the model has been identified, for the quantitative isotherm analysis for the determination of binding constants, $c_k(s)$ does not provide an additional advantage over the global analysis of all isotherm data derived from the conventional $c(s)$ analyses at all signals. In the latter, the different signal contributions can be naturally accounted for by considering each component's extinction coefficient. The availability of different signals that arise to different extent from the interacting components will be highly useful.

iii) Repulsive interactions

Repulsive interactions occur at high protein concentrations. Repulsive interactions are ubiquitous in cellular environment, and sometimes may cause significant differences in binding constants *in vitro* compared to *in vivo* (Minton 2001). In SE, effects of crowding on protein interactions have been experimentally studied using a fractionation technique, and were interpreted in a firm theoretical framework (Rivas et al. 1999). In contrast, studies under crowded conditions have not been established in SV, and we are not aware of a similar theoretical framework for data interpretation of SV of multi-component solutions under crowded conditions.

Nevertheless, in SV repulsive effects are encountered at high protein concentrations, and the experimenter needs to be familiar with them and know how they can be – to some extent – taken into account. A first-order approximation for single-component solutions of the effect of decreasing s-value with increasing protein concentration is the description

$$s_{i,meas}(c) = s_{i,ideal} \big/ \left(1 + k_S c\right) \qquad (17)$$

, where $s_{i,meas}$ and $s_{i,ideal}$ are the s-values experimentally observed at a finite concentration c and ideally in infinite dilution, and the non-ideality constant k_S. k_S depends on the partial-specific volume and the shape of the macromolecule. Typical magnitudes for k_S are on the order of 10 ml/g or less, but larger values can be expected for asymmetric molecules (Schachman 1959; Creeth and Knight 1965; Rowe 1992). For example, a value of 0.009 ml/mg for compact, approximately spherical proteins, leads at 0.1 mg/ml to ~ 0.1% lower s-value as compared to that in ideal infinite dilution (i.e. within typical experimental error), but ~ 1% lower s-value at 1 mg/ml, and ~ 8 % lower at 10 mg/ml. Therefore, at concentrations < 1 mg/ml the effects are usually close to negligible. However, k_S is much larger for highly elongated macromolecules, such that non-ideality effects set in at much lower concentrations. For non-interacting species, the experimental s-values can be fit with Eq. 17 to derive k_S, and generally more importantly, the s-value of that species extrapolated to infinite dilution.

Alternatively, first-order corrections for non-ideality can be made in the Lamm equation Eq. 5 (Dishon et al. 1967; Solovyova et al. 2001). This also requires consideration of the concentration dependence of the diffusion coefficient, $D_{meas}(c) = D_{ideal}\left(1 + k_D c\right)$, expressed through the coefficient k_D. Both k_S and k_D are can be determined by Lamm equation modeling. They are related to the second virial coefficient via $k_S + k_D \cong 2MA_2$, a relationship that has been used for the evaluation of weak interparticle interactions in complex solvents, with application to the study of crystallization conditions (Solovyova et al. 2001).

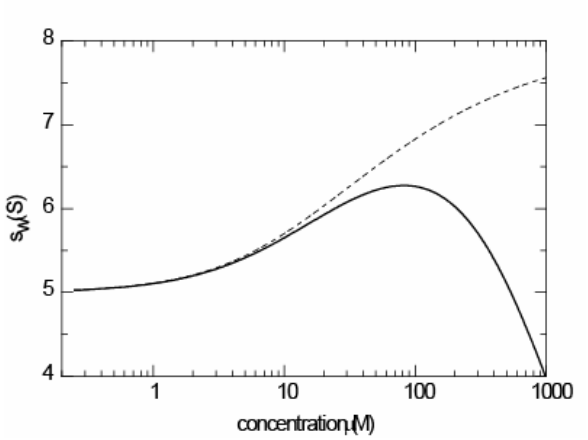

Figure 11: Self-association in the presence of hydrodynamic non-ideal sedimentation. The weight-average s-values $s_w(c)$ for a model monomer-dimer system with $s_1 = 5$ S, $s_2 = 8$ S, $K_A = 2\times10^4$ M^{-1} and with a non-ideality coefficient k_s of 0.009 ml/mg approximating that expected for spherical particles. Shown is the s_w isotherms that would be measured in the presence (solid line) and absence (dashed line) of hydrodynamic non-ideality.

Frequently when studying interacting systems, concentrations greater than 1 mg/ml have to be used, such that non-ideality effects are superimposed to the concentration-dependence of the s-values arising from (attractive) protein interactions. Figure 11 shows an example for the deviations in the isotherm of weight-average sedimentation coefficient $s_w(c)$ to be expected for a weak monomer-dimer interaction. Although the mutual repulsive interactions may be different for monomer and dimer (and their cross-terms), the data do not have sufficient information to extract several non-ideality coefficients. A pragmatic and probably in most cases safe assumption is that a first-order approximation is sufficient, in which all repulsive terms are described by one average value for k_S. Similar is true for other interaction schemes. However, the analyses that are based on the amplitudes of the different peaks (representing the species populations for slowly interacting systems, or the magnitude of the undisturbed and reaction boundary, respectively) are much less susceptible to non-ideality-induced shifts in s-values. Similarly, the analysis of the peak composition with multi-signal $c_k(s)$ does not depend on the species s-values. An example of the analysis of very weak interactions in the presence of non-ideality is the application to carbohydrate interactions (Patel et al. 2007).

The presented approaches can only be a considered first-order approximation for the treatment of slight or moderate non-ideality, and it is preferable if experiments can be conducted at concentrations less than a few mg/ml (or even < 1 mg/ml where non-ideality is usually negligible). At very high protein concentrations the repulsive interactions can completely dominate the sedimentation and show more complex patterns of concentration profiles. This includes the classic Johnston-Ogston effect (Johnston and Ogston 1946), which is a boundary inversion in the slowly sedimenting component due to the differential concentration dependence in the presence and absence (below and above the boundary) of a faster sedimenting component. In addition, optical aberrations frequently occur under such conditions.

c) Hydrodynamic Analysis of Gross Solution Shape and Conformational Changes

For a macromolecule of known molar mass and density (partial-specific volume), from the measured s-value the hydrodynamic translational frictional ratio can be determined. Because this approach is first-principle based and does not require any size standards, and due to the high precision of the experimentally measured sedimentation coefficients, it is possible to interpret the frictional ratio quantitatively in terms of different macromolecular shapes in solution. It is obvious, though, that a single number, even though a precisely determined one, cannot reveal unambiguously a three-dimensional structure. The latter consideration is crucial for the interpretation of the f/f_0 value.

Traditionally, f/f_0 is interpreted in terms of alternate hydrodynamically equivalent model shapes. For example, we can obtain the dimensions of the hydrodynamically equivalent smooth and compact prolate ellipsoids, oblate ellipsoids, and rods of the same mass. Thus, if additional knowledge on the type

of shape is available or hypothesized, the gross dimensions of the macromolecule in solution can be obtained from the experimental f/f_0. However, if translational frictional coefficients for the protein in different oligomeric states are available, the ambiguity between a flat, roughly disk-shaped (oblate) and a cigar-shaped (prolate) molecule may be removed and geometric aspects of complex formation be deduced (Schuck et al. 2000; Burgess et al. 2005). For example, the ratio of frictional coefficients from monomer and dimer of a linearly elongated molecule with end-to-end association or a disk that associates by stacking are very different, in the former case leading to a higher frictional ratio of the dimer, in the latter case to a lower frictional ratio of the dimer.

A much stronger use of the experimental frictional ratio can be made in comparative mode, either comparing different protein variants, proteins with/without small ligand, or under different solution conditions that may induce conformational changes. Frequently, crystal structures are available from which a theoretical frictional ratio (or directly a theoretical s-value) may be calculated and compared to the experimental value (Errington and Rowe 2003). Sometimes structures of subunits are available and different hypotheses for their spatial arrangement can be tested against the experimental data. Theoretical predictions of hydrodynamic properties can be made, for example, with the software HYDROPRO (Garcia De La Torre et al. 2000) and/or SOMO (Rai et al. 2005) on the basis of pdb files. Care must be taken, however, to use the entire molecule in the structure subjected to the calculation, including tags, or flexible regions potentially missing from the crystal structure. For the latter, frequently several hypothetical conformations are assumed spanning very extended and very compact geometric arrangements, translating into a range of possible s-values. Similarly, the hydrodynamic information can complement data from other biophysical solution techniques (see below).

2. Basic Principles: Experimental

It is assumed in the following that the reader is familiar with the description of the experimental aspects of SE analytical ultracentrifugation described in Unit **[Copy-Editor, please insert ref to Equilibrium Sedimentation Unit]** of this series. Many of the experimental considerations will be very similar for SV, as it is based on the same experimental device. This includes the choice of the detection system and its consequences for the possible buffers, the range of protein concentrations, and the factors relating to protein partial-specific volume. However, slight differences arise in the SV configuration and additional aspects appear, for example, due to the use of higher rotor speeds. These topics are important for the design of a SV experiment and discussed in the following. For more detailed instructions, the reader is referred to the step-by-step protocol provided as internet resource.

a) Choice of Detection System

Of the two commonly available detection systems, the absorbance (ABS) and the interferometric (IF) refractive index system, in the absence of special considerations outlined below, the latter is usually the system of choice for SV experiments. This is despite the systematic signal offsets arising from jitter, integral fringe shifts, and radial-dependent baseline, because in SV experiments, the data have sufficient information for these signal contributions to be readily calculated (Figure 12). Further, due to the shorter experimental time and the use of a single rotor speed throughout, there is no need for ‘aging’ of the centerpieces (in contrast to SE experiments

when IF optics is used). This allows the excellent sensitivity of the IF system, the fast rate of scanning, and the high radial resolution to be fully exploited.

However, when using the IF system, due to the higher rotor speed, even greater care must be taken in SV ultracentrifugation to establish that the buffer in the reference sector is precisely matched in chemical composition. Sedimenting salt contributes a very significant signal, which takes the form of a baseline slope increasing with time (or decreasing with time if the excess salt is in the reference buffer). If unmatched, these signals have to be modeled explicitly and can cause some correlation with a broad distribution of very large s-values. (For the same reason, it is also very important in SV to precisely match the volumes inserted in the sample and reference sector of the ultracentrifugal cell, see below.) Matching routinely works very well using dialysate or running buffer from gel filtration. However, in cases where there is a large amount of co-solvent giving very high IF signals (e.g., high salt buffers > 1 M, and buffers containing sucrose, glycerol, or detergents), it can be extremely difficult to precisely match the signals in sample and reference sector, and the use of the ABS system is advantageous. If the use of the IF system is critical, it may be best to omit the co-solvent altogether from the reference solution and explicitly model their signal contribution. (In this case, it might be advantageous to sediment the co-solvent in a separate cell).

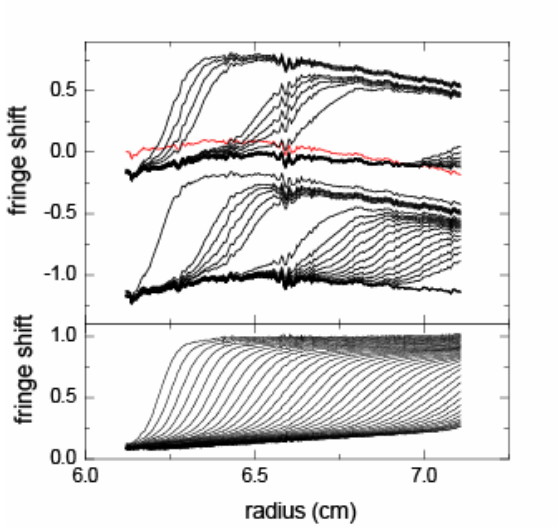

Figure 12: Example of raw interference optical data (top). Both the time-dependent (radial-invariant, RI) signal offsets arising from jitter and integral fringe shifts, as well as the radial-dependent (time-invariant, TI) signal offset arising from imperfections in the smoothness of the optical elements can be clearly discerned. After fitting the data with a model including terms for TI and RI noise (see Eq. 6), 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.

Another case where the ABS system obviously is advantageous is the study of heterogeneous interactions of proteins with different extinction properties. As mentioned above in the introduction of the multi-signal sedimentation coefficient distribution $c_k(s)$, it may not be necessary to introduce extrinsic chromophoric labels if the mg/ml based extinction coefficient of two proteins is different, and the IF and the ABS system (e.g., at 280 nm) are used simultaneously. If a fluorescent label is present on the molecules, this is accompanied by a strong absorbance band that usually allows convenient selective detection with the ABS system. If the ABS system is applied acquiring data at different wavelengths, it is desirable that extinction maxima or minima are used such as to minimize the effect of small inaccuracies in the wavelength control. (However, usually a sufficient number of scans are available to selectively include only those scans where the wavelengths are at the desired value.)

Generally, it should be noted that the slower scanning speed of the ABS system is not a very significant drawback and one should not be discouraged from its use. Even the ABS system does invariably display small systematic, radial-dependent signal offsets resulting from imperfections in the windows, such that the inclusion of TI noise in the data analysis is recommended, except for very small species < 5 kDa.

b) Buffers

As discussed in the Unit on SE ultracentrifugation, buffers need to be compatible with the optical detection system. There are several special considerations due to the sedimentation of the buffer components in the high centrifugal fields employed in SV ultracentrifugation. Besides the refractive index signals arising from the buffer components, as indicated above, the density gradient associated with the sedimenting buffer components at high concentrations can be a concern, particularly when using co-solvents like sucrose and glycerol. In this case, regions of different solvent densities are dynamically created during the sedimentation experiment, which noticeably alters the macromolecular sedimentation pattern. This can be taken into account computationally in SEDFIT if the sedimentation and diffusion coefficient of the co-solvent are measured separately (Schuck 2004a). Whether or not significant density gradients occur can be determined from a separate experiment where the sedimentation profiles of the co-solvent alone (but at the same concentration and other buffer conditions as in the macromolecular experiment) are measured with the IF system. This will report on the final concentration difference between meniscus and bottom relative to the loading concentration, and density tables or the software SEDNTERP can be consulted to determine whether the maximal shift in sedimentation coefficients induced by the solvent buoyancy change is significant and needs to be accounted for.

c) Size, Partial-Specific Volume, and Prior Characterization of the Proteins

The size of molecules that can be conveniently studied by SV ranges from ~ 1 kDa or less to > 1 GDa. Over most of this range, protein interactions can be studied as outlined above. However, for species < 10 kDa and > 10 MDa special considerations arise from the presence of very strong and very little diffusion, respectively.

For molecules < 10 kDa diffusion is usually quite strong and the associated boundary broadening leads to higher correlation of fitting parameters and ambiguity in the correct analysis model. However, diffusional deconvolution of the sedimentation coefficient distribution in the $c(s)$ approach works very well also in this size range. An additional opportunity for study of small molecules is to exploit the data available on back-diffusion from the bottom of the cell to give more direct molar mass information. This can be accomplished both by adjusting the fitting limits in the SV analysis to include this range, and by letting the experiment continue until sedimentation equilibrium is established, i.e., appending a SE experiment at the end of the SV run. In fact, for very small molecules, SE experiments and SV experiments have both to be conducted at the highest possible rotor speed, and the approach to equilibrium of SE can be regarded as the SV experiment. (In contrast, for larger proteins, e.g. greater than ~ 30 kDa, the back-diffusion information is ordinarily not useable due to the high local concentration at the bottom of the cell, the high signal gradients, and associated optical aberrations. Further, at the high local concentrations at the bottom of the cell, many molecules adsorb to the centerpiece

which will alter the back-diffusion in an unpredictable extent. However, in this size range the information on molar mass can be usually extracted conveniently from the sedimentation boundaries.)

For very large species (> 10 MDa), the sedimentation boundaries become very sharp and over the short experimental time little diffusional boundary spreading can be observed. This makes the analysis of sedimentation coefficients easier, but the determination of molar masses more difficult. Because it can be problematic to conduct SE experiments on such large macromolecules, independent information on the diffusion coefficient may be desirable, for example, from dynamic light scattering.

Regarding the protein partial-specific volume, the same issues exist in SV as in SE, and the reader is referred to the discussion in the SE Unit [Copy-Editor, please insert ref to Equilibrium Sedimentation Unit] for how to predict $\bar{v}$ values. For glycosylated proteins, it is important not only to be aware of the carbohydrate contributions to the protein buoyancy, but also to its friction. Frequently, glycosylated proteins exhibit very high frictional ratios due to carbohydrates extending out into the solution. It is useful to know about the heterogeneity of the glycosylation (for example from mass spectroscopy), since heterogeneous samples cannot be expected to exhibit sedimentation patterns like that of a single species (see above). Their heterogeneity can be taken into account better with distribution models, such as $c(s)$, $c(M)$, or the two-dimensional distributions. Similarly, when studying membrane proteins, the contribution of detergent to a protein/detergent complex can be eliminated with regard to the partial-specific volume and buoyant mass under density matching conditions (where $\rho_0 = 1/\bar{v}_D$), but the detergent will continue to have a very large effect on the hydrodynamic friction and the s-value (Gohon et al. 2004).

SV offers a much better resolution than SE ultracentrifugation, and as a consequence, impurities, in particular small molecular weight impurities, can be much more tolerable as long as they sediment outside of the range of s-values expected for the interacting proteins. Nevertheless, it is highly desirable that the proteins be purified by size-exclusion chromatography as the last step of purification, and that the purity is greater than 95%. Purity should be assessed also by SDS-PAGE, and the molar mass should be known, preferably from mass spectroscopy. Extinction coefficients should be estimated from the amino acid sequence for setting up the experiment, but if they are critical (such as in the multi-signal $c_k(s)$ analysis) should be determined in separate experiments in the ultracentrifuge. (This will eliminate discrepancies from imperfections in the monochromator.)

d) Concentrations and Sample Volumes

The best sample concentrations depend on the question asked with the experiment, the type of interacting system and its binding constants, as well as on the detection system that is used. Generally the same considerations are valid as outlined in the Unit on SE ultracentrifugation. Briefly, to study binding constants it is important to achieve concentrations where both free and complex species are detected and their relative concentration can be shifted. (The calculator function in SEDPHAT can be used predicting each species concentrations given the total loading concentrations for each component and an estimate of the equilibrium constants.) Unlike in SE ultracentrifugation, however, the loading concentrations are an upper limit of the concentration range observed in a single experiment, because in SV we usually cannot take advantage of the

concentrating effect of the centrifugal field close to the bottom of the cell. It is the loading concentrations that should span ideally from 0.1-fold to 10-fold K_D. Due to the information gained from the concentration-dependence of the *c*(*s*) peaks (Figures 7 – 9), it is crucial to study the proteins at a range of concentrations, and we recommend that even for studying proteins that are not expected to self- or hetero-associate, since in our experience, the presence of interactions frequently may not have been apparent on the basis of other common prior characterization techniques. For heterogeneous interactions, it is crucial to study each component alone in separate cells, in addition to the mixtures. Whereas in SE it was highly desirable to conduct sets of experiments as either dilution or titration series so as to invoke additional mass conservation constraints in the global analysis, there is no advantage in the SV analysis of using such particular configurations.

One non-trivial problem related to the sample concentrations of slowly interacting systems can be to establish that the mixture is in chemical equilibrium before the start of the experiment. For example, if the result of an initial SV run indicates complex life-times of > 10^4 sec (by displaying peaks for the complexes with concentration-independent s-values), one needs to verify that, before the start of the SV run, the sample had a chance to chemically equilibrate for several hours or overnight if the mixture was prepared by dilution of a higher concentrated stock. Ordinarily this will require repeating the run with the appropriate incubation times for the sample mixtures (Yikilmaz et al. 2005).

When the hydrodynamic analysis of the shape of a rapidly reversible complex is desired, one should raise the concentrations as much as possible to saturate the complex, but not yet as to encounter hydrodynamic repulsive non-ideality. To determine the complex s-value, the analysis of the binding isotherm is necessary. In this regard, the application of Gilbert-Jenkins theory to the analysis of the concentration-dependent s-value of the fast boundary component is particularly useful, as the measured values are closer to the s-value of the complex.

For an optimal SV analysis, long solution columns are required. In standard 12 mm centerpieces, we routinely use 400 µl samples (for both sample and reference sector, precisely matched in volume and composition). In order to save material for the samples at high concentrations, it is possible to use 3 mm centerpieces reducing the required volume to 100 µl per cell. (In our experience, we have not observed detrimental effects on the data quality from the change in focal plane due to the different cell geometry.) This can have the added advantage of avoiding too high optical densities > 1.5 OD to be recorded with the ABS system (too high signals are not a concern in the IF system). For the lower concentration samples, the 12 mm centerpieces are preferred due to their better signal-to-noise ratio.

e) Rotor Speed, Rotor Temperature, and Experimental Time

The goal of the SV experiment is hydrodynamic separation, which can be achieved with best resolution in high centrifugal fields. The rotor speed should be 50,000 rpm for most medium sized proteins, or 60,000 for species < 10 – 20 kDa. If the IF system is employed, the scans are acquired in sufficiently high rate to study even large molecules (e.g., 500 kDa) at 50,000 rpm. Only above that range is it necessary to reduce the rotor speed to 40,000 rpm or below. In exceptional cases (e.g., when the hydrodynamic resolution of species and the characterization of the binding constants are not important but the determination of molar mass is the most

important aspect of the experiment), one may decrease the rotor speed to extend the diffusional time and increase the boundary spread.

The long solution columns and high rotor speeds lead to the generation of a moderate pressure (~ 20 – 30 MPa) at the bottom of the cell, which in rare cases affects the assembly of protein complexes when significant changes of the partial-specific volume occur during complex formation (Harrington and Kegeles 1973). A well-known example is the depolymerization of tubulin at high rotor speeds (Marcum and Borisy 1978). Such pressure effects would be observed in a rotor-speed dependence of the equilibrium constants, and are not typically considered. A different pressure effect occurs from the compressibility of water, which is typically neglected, but can be significant in particular in the study of small molecules (Schuck 2004b). It can be accounted by corrections in Eq. 5. that describe the local change in buoyancy of the sedimenting proteins, as implemented in SEDFIT.

In contrast to SE ultracentrifugation, it is mandatory for SV that the rotor and the sample is thoroughly temperature equilibrated. In practice, convection induced by local temperature fluctuations is among the chief limiting factors in the precision of the sedimentation coefficients. A rigorous temperature equilibration procedure is described in the step-by-step protocol. The run temperature itself is not critical and can be adjusted easily between 10 and 30 °C (and with more care between 4 and 40 °C) as dictated by the requirements of the study. Usually, it is most convenient to conduct the SV experiments at close to 20 °C, the standard temperature for hydrodynamic calculations, so as to minimize the buffer correction factors.

The experimental time is typically several hours, depending on the s-value of the proteins under study. It is very important to let the sedimentation proceed until the trailing part of the boundary has disappeared. With samples not previously studied in SV, it is frequently advantageous to scan for a long time after this, as this can be helpful later in the data analysis stage to determine the nature of unexpected small molecular weight contaminants. Also, the baseline parameters will be better defined if very late scans are available with little or virtually no further sedimentation.

f) Fit Criteria and The Precision of the Derived Parameters

It is important to establish that the data analysis model actually fits the raw data. This issue has not always been part of the tradition of SV analysis, in particular when using previous data transformation methods that lack back-transformation of the model into the raw data space (Schuck et al. 2002; Schuck 2003). However, it is obviously essential for a reliable and rational data interpretation. This is particularly true when using more detailed analysis models, and it is a great strength of the direct data analysis approach followed in the present Unit that it can provide rigorous comparison of raw data and best-fit model. If the best-fit model does not match the data well, it is obviously questionable whether the parameters described by it reflect the true molecular properties.

As in all biophysical disciplines, in addition to the statistical errors from noise in the data acquisition, there are sources of systematic errors that are poorly controlled in experimental practice. These systematic errors limit the level of detail beyond which the data are not commonly interpreted. In SV ultracentrifugation, the fit quality should be such that the ordinary overlay of residuals of all scans does not allow discerning any systematic deviations. However,

this is usually not a sufficient criterion. On the other hand, a strictly mathematical runs-test for systematics of residuals (such as the quantity Z calculated in SEDFIT, which ideally should be close to unity (Straume and Johnson 1992)) is much too stringent for practical use in most SV analyses. In order to permit a well-balanced visual discrimination of systematic versus statistical residuals better suitable to the signal/noise ratio in SV, we have developed a graphical representation of the residuals of all scans, by displaying the residuals in a 2d grayscale bitmap, where each radius at each scan time is represented by a pixel (see Figure 13) (Schuck et al. 2002). Here, residuals originating from instabilities in the TI and RI signal contributions typically display as horizontal or vertical lines, whereas systematic misfits of the signal from the macromolecular sedimentation boundary produces diagonal features. An ideal fit only containing statistical errors will have a bitmap that is entirely featureless to the human eye. In practice, the fit to experimental SV data should be free of diagonal features. An example for the difference between an excellent and a poor fit in the bitmap representation is shown in Figure 13.

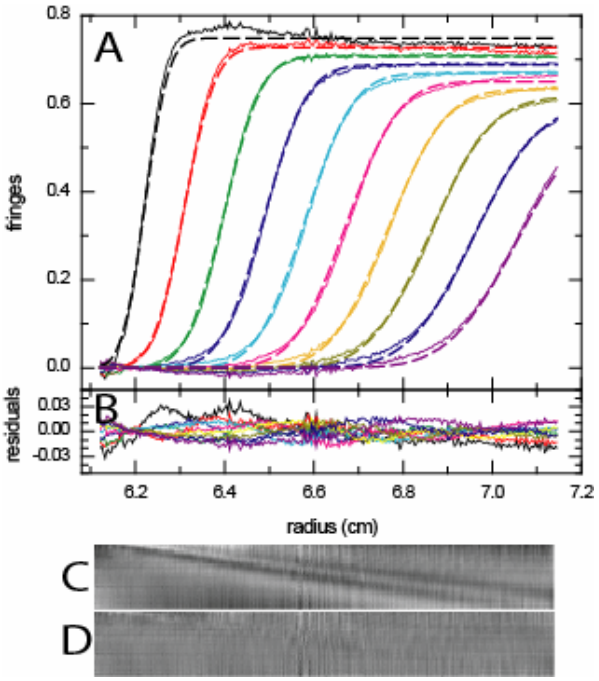

Figure 13: Demonstration of a fit with insufficient quality. The data form Figure 3 (Panel A, solid lines, showing only a representative subset of the scans) are modeled with an impostor single-species fit (Panel A, dashed lines). Although the theoretical boundaries roughly outline the shapes of the sedimentation boundary, and the root-mean-square deviation is only 0.0084 fringes – only slightly above the typically encountered noise of data acquisition – the residuals overlay (Panel B) shows systematic deviations. The bitmap representation of the residuals (Panel C) highlights the misfit by showing a significant diagonal feature arising from systematic misfits in the sedimenting boundary. In contrast, fitting the data with a $c(s)$ model (data not shown) results in a residuals bitmap (Panel D) with very little diagonal features, but some vertical and horizontal lines indicative of the remaining residuals being due to technical imperfections in the data acquisition process (e.g., vibrations of optical elements). The root-mean square deviation for the $c(s)$ model is 0.0033 fringes. The importance of critical inspection of the fit quality can be seen from comparing the molar mass estimate derived from the inadequate single-species fit of 60.1 kDa with the molar mass distributions shown in Figure 3: the single-species fit would lead to qualitatively wrong conclusions about the oligomerization state of the molecule. The s-value of the major species is not quite as sensitive to imperfections in the fit.

One important strategy to achieve good fits is to include the meniscus of the solution column as an adjustable fitting parameter. For models that include small Mw species that show some back-diffusion within the radial range of data analysis, the bottom of the solution column should be an adjustable fitting parameter, too. Although a rough visual localization of the meniscus is possible from the radial range of its optical artifact, the minimum and maximum radius values of this signal feature should be used only to define upper and lower limits of the true meniscus location. Since the data analysis results in s-values that are far more precise than the possible visual localization of the meniscus, any fixed constraint will likely lead to a bias resulting in significant misfit. (It is a misconception that the true meniscus position would have to coincide with the maximum signal from the optical artifact. Among other factors, the location will depend on the nature of the macromolecule and solvent.)

The most common sources of error in the SV experiment are signal offsets from optically unmatched buffer salts, and convection. The optically unmatched buffer salts contribute signal in the interference optics only, and can frequently be accounted for by including an extra discrete low Mw species into the data analysis, such as the '*c*(*s*) with 1 discrete component' model in SEDFIT, optimizing the apparent Mw and s of this species. For this to work best, it is beneficial to include very late scans into the analysis, even beyond the point where the macromolecular sedimentation has completed (i.e. the trailing part of the boundary has reached the bottom of the cell). In order to better identify contaminations from small Mw impurities, it is prudent to conduct the experimental data acquisition for a long time (e.g. overnight) beyond the expected macromolecular sedimentation time.

Convection is the turbulent, macroscopic movement of volume elements of the solution. This can be driven by temperature gradients, and/or misalignment or scratches on the centerpiece. Even though convection may occur only at the beginning of the experiment, its results do affect all the later scans by delaying the elapsed time at which the boundary arrives at a certain radius. Most analysis models depend on the entire history of the experiment from its start to the last scan modeled (this also includes all of the superseded data transformation and/or differentiating approaches, with the only exceptions being Lamm equation models that explicitly take an initial scan as starting point (Cox 1966; Schuck et al. 1998), and a particular form of a differential second moment mass conservation approach to derive s_w (Schachman 1959; Schuck 2003)). One indicator of convection is that the best-fit meniscus position is at the minimum or maximum allowed value (when these limits have been set close to the bounds of the optical artifact), within what can be identified as the air-to-air region above the solution column, or within the solution column, respectively.

If it is not possible to arrive at a virtually perfect fit within the statistical noise of the data, it is important to know how susceptible different analysis parameters can be to misfits. The most robust number is usually the weight-average s-value s_w of the entire distribution (excluding signals at the lowest s-values that may be affected by possible correlation with the baseline (Schuck 2006)) and the associated loading concentration. Similarly robust is the s_w and concentration value integrating across the major peaks. The typical precision of a well-executed SV analysis is better than 0.01 S (for species < 10 S). For minor peaks in *c*(*s*), the concentration over a wider s-range is more robust than the precise location of a peak. For a detailed practical study on the precision of trace components, see (Pekar and Sukumar 2007). Finally, the most susceptible parameters are those that quantify the boundary spreading, such as the molar masses, frictional ratios, diffusion coefficients, and kinetic parameters in Lamm equation models with reaction terms.

Due to the high signal/noise ratio of typical SV experiments, the systematic errors can have a bigger impact than statistical parameter uncertainty. However, it is important to be aware of the statistical precision of the different parameters, as well, in particular when studying interacting systems and interpreting details of the distributions. It is not unusual that features can be discerned with confidence from signal amplitudes lower than the noise of data acquisition, due to the large number of data points typically entering the analysis (10^4-10^5). In SEDFIT, a tool is available for the Monte-Carlo analysis of the statistical errors of s-values and concentration values from integrating *c*(*s*) over certain ranges. It should be noted that the use of Bayesian prior knowledge is a powerful tool to explore statistically equivalent, alternative data interpretations. In SEDPHAT, for the analysis of interacting systems, Monte-Carlo routines are

available, as well. However, it should be noted that for complex error surfaces, which may arise in global fits of interacting systems, the Monte-Carlo approach is not rigorous due to the automatic optimization required for each iteration. In this case, manually exploring the error contour using F-statistics (Straume and Johnson 1992) is the best approach (for more details see (Schuck 2007c)). Finally, error limits can be greatly reduced by globally modeling data from different concentrations, both for interacting and non-interacting species models, in SEDPHAT (Hsu et al. 2005; Chou et al. 2006; Lelj-Garolla and Mauk 2006; Liu et al. 2006b).

3. Complementarity to Other Biophyiscal Techniques

SV ultracentrifugation can be particularly strong method to complement other biophysical techniques in the study of protein interactions. This is due to the fact that in SV, it is frequently possible to hydrodynamically resolve different co-existing species, which allows to determine the association scheme of an interaction (number of species and their stoichiometry) frequently better than on the basis of an isotherm analysis of an average property in solution alone, be it heats in isothermal titration calorimetry (ITC) or spectroscopic signals in circular dichroism (CD), fluorescence spectroscopy, or other spectroscopic or scattering approaches. For example, for reactants of sufficient size, it can be convenient to recollect the material after an ITC titration for study by SV centrifugation, independently (and separately) confirming the complex stoichiometry, active concentrations, and aggregation state. Further, SV can usually give unequivocal information on protein oligomeric state under solution conditions employed in other techniques. The latter is very important, for example, for the mobile analyte used in surface plasmon resonance (SPR) or other biosensing techniques (Schuck 1997).

As indicated in the Unit on SE ultracentrifugation [Copy-Editor, please insert ref to Equilibrium Sedimentation Unit], SV is routinely used in our laboratory to study samples prior to SE ultracentrifugation. This is because of the ability of SV to detect impurities and degradation products contributing to the signal, the possibility to better identify the interaction scheme, and the faster time-scale of SV. *Vice versa*, it can be very useful to confirm the stoichiometry of complexes by independent study of the molar mass in SE. Global analysis of SV and SE data in SEDPHAT can be a stringent test for sample purity and the interaction model (Davis et al. 2004).

Another technique that is traditionally used side-by-side with SV is dynamic light scattering (DLS), as both can measure protein translational diffusion coefficients in free solution (molecular weight numbers supplied by some DLS manufacturer software should be ignored). The agreement between the diffusion coefficients measured by DLS and SV is excellent for mono-disperse samples (Schuck et al. 2000). Similarly, for a monomer-oligomer assembly process, excellent agreement between the ratios of diffusion coefficients measured by fluorescence correlation spectroscopy and the ratios of s-values was reported (Boukari et al. 2004). A second area of very productive complementarity between DLS and SV is the study of conformational changes upon ligand binding. If the ligand is small, conformational changes leading to an increase of the frictional coefficient after binding are associated with a lower s- and higher D-value, while those leading to a decrease of the frictional coefficient are associated with a higher s- and lower D-value. Considering the slightly increased mass of the protein/ligand

complex, the former can be unambiguously identified directly by SV but not easily with DLS, and the latter case can be unambiguously identified directly by DLS and not easily with SV.

When studying conformational changes upon ligand binding, SV is complementary also to CD, as conformational changes in the secondary structure detectable by CD may leave the tertiary structure (and translational friction coefficient) unchanged, while changes in the subunit arrangement may be detected by SV but not be accompanied with a CD spectroscopic signature. It should be noted that sample amounts and concentrations are well compatible between CD and SV.

It can be very interesting to compare the solution structure of proteins or protein complexes with available crystal structures of the protein (complex) or of subunits. As outlined above, this can be done on the basis of hydrodynamic data from SV ultracentrifugation. Clearly, small angle x-ray and neutron scattering provide much more detail about the solution structure (Svergun and Vachette 2007), although sometimes information on the translational frictional ratio from SV may aid in further improving the discrimination between alternate models (Furtado et al. 2004; Nollmann et al. 2004). *Vice versa*, structural information can be highly useful to predict sedimentation coefficients of complex species in interacting systems, which can allow for one to fix these parameters and reduce parameter correlations in the modeling of SV isotherms.

In summary, the application of SV can frequently contribute synergistically to the specific unique aspects of macromolecular interactions probed by other techniques. Isotherms derived from different techniques can be modeled globally in SEDPHAT. When samples are studied in sequence with other methods, it should be noted that, although the sample can be recovered from SV and studied, for example, by SDS-PAGE or mass spectroscopy, the proteins are usually not suitable for quantitative biophysical studies after SV, due to their exposure at extremely high concentration with the bottom of the centerpiece, where frequently poorly reversible aggregation and surface film formation takes place.

4. Summary

In this Unit, we have introduced basic practical and theoretical aspects for the design of standard SV ultracentrifugation experiments for the study of protein interactions. It is thought to be used in conjunction with the internet-based practical step-by-step protocol (Balbo et al. 2007) that describes in detail how to conduct an SV experiment. More tutorial material and a complete detailed reference manual for the SEDFIT and SEDPHAT software can be found at (Schuck 2007c). Finally, hands-on workshops are conducted regularly in our laboratory at the National Institutes of Health.

Analytical ultracentrifugation is a very powerful method with a long history with virtually unlimited number of experimental strategies and variations. For obvious reasons, we have confined the present commentary to the currently most commonly used experimental and analysis approaches, which usually give the most reliable and detailed results for a wide class of problems. Topics not treated include analytical zone centrifugation, isopycnic density gradient sedimentation, sedimentation with density contrast, and specific aspects of the study of detergent-solubilized membrane proteins. For these and for more advanced aspects on the methods described, the reader is referred to the literature cited.

Selected Key References:

General Reviews

Schuck, P. 2007. Sedimentation velocity in the study of reversible multiprotein complexes. In *Biophysical Approaches for the study of complex reversible systems.* (ed. P. Schuck), pp. 469-518. Springer, New York.

Howlett, G.J., Minton, A.P., and G. Rivas. 2006. Analytical ultracentrifugation for the study of protein association and assembly. Curr. Opin. Chem Biol. 10:430-436

Scott, D.J., and P. Schuck. 2005. A brief introduction to the analytical ultracentrifugation of proteins for beginners. In *Modern Analytical Ultracentrifugation: Techniques and Methods* (eds D.J. Scott, S.E. Harding, and A.J. Rowe) pp. 1-25. The Royal Society of Chemistry, Cambridge

Balbo, A., and P. Schuck. 2005. Analytical ultracentrifugation in the study of protein self-association and heterogeneous protein-protein interactions. In *Protein-Protein Interactions* (eds E. Golemis, and P.D. Adams) pp 253-277. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York

Lebowitz, J., Lewis M.S., and P. Schuck. 2002. Modern analytical ultracentrifugation in protein science: a tutorial review. Protein Sci. 11:2067-2079

Laue, T. 2001. Biophysical Studies by ultracentrifugation. *Current Opinion in Structural Biology* 11: 579-583.

Schachman, H. K. 1959. Ultracentrifugation in Biochemistry. Academic Press, New York.

Svedberg, T., and K. O. Pedersen. 1940. The Ultracentrifuge. Oxford University Press, London.

References on Special Topics

Berkowitz, S.A., and Philo, J.S. 2006. Monitoring the homogeneity of adenovirus preparations (a gene therapy delivery system) using analytical ultracentrifugation. *Anal Biochem* 362: 16-37.

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

Gohon, Y., Pavlov, G., Timmins, P., Tribet, C., Popot, J.L., and Ebel, C. 2004. Partial specific volume and solvent interactions of amphiphol A8-35. *Anal Biochem* 334: 318-334.

Howlett, G.J. 1992. Sedimentation analysis of membrane proteins. In *Analytical Ultracentrifugation in Biochemistry and Polymer Science*. (eds. S.E. Harding, A.J. Rowe, and J.C. Horton), pp. 470-483. The Royal Society of Chemistry, Cambridge, U.K.

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

Pekar, A., and Sukumar, M. 2007. Quantitation of aggregates in therapeutic proteins using sedimentation velocity analytical ultracentrifugation: practical considerations that affect precision and accuracy. *Anal Biochem* 367: 225-237.

Schuck, P. 2004. A model for sedimentation in inhomogeneous media. I. Dynamic density gradients from sedimenting co-solutes. *Biophys. Chem.* 108: 187-200.

Schuck, P. 2004. A model for sedimentation in inhomogeneous media. II. Compressibility of aqueous and organic solvents. *Biophys. Chem.* 187: 201-214.

Modern SV Data Analysis for Interacting Systems

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

Dam, J., Velikovsky, C.A., Mariuzza, R., Urbanke, C., and Schuck, P. 2005. Sedimentation velocity analysis of protein-protein interactions: Lamm equation modeling and sedimentation coefficient distributions c(s). *Biophys J* 89: 619-634.

Dam, J., and Schuck, P. 2005. Sedimentation velocity analysis of protein-protein interactions: Sedimentation coefficient distributions c(s) and asymptotic boundary profiles from Gilbert-Jenkins theory. *Biophys J* 89: 651-666.

Balbo, A., Minor, K.H., Velikovsky, C.A., Mariuzza, R., Peterson, C.B., and Schuck, P. 2005. Studying multi-protein complexes by multi-signal sedimentation velocity analytical ultracentrifugation. *Proc Natl Acad Sci U S A* 102: 81-86.

Literature Cited

Ali, S.A., Iwabuchi, N., Matsui, T., Hirota, K., Kidokoro, S., Arai, M., Kuwajima, K., Schuck, P., and Arisaka, F. 2003. Rapid and dynamic association equilibrium of a molecular chaperone, gp57A, of bacteriophage T4. *Biophys J* **85:** 2606-2618.

Arthos, J., Cicala, C., Steenbeke, T.D., VanRyk, D., Dela Cruz, C., Khazanie, P., Selig, S.M., Hanback, D.B., Nam, D., Schuck, P., et al. 2002. Efficient inhibition of HIV-1 viral replication by a novel modification of sCD4. *J.Biol.Chem.* **277:** 11456-11464.

Balbo, A., Brown, P., and Schuck, P. 2007. www.analyticalultracentrifugation.com/SVProtocol.htm.

Balbo, A., Minor, K.H., Velikovsky, C.A., Mariuzza, R., Peterson, C.B., and Schuck, P. 2005. Studying multi-protein complexes by multi-signal sedimentation velocity analytical ultracentrifugation. *Proc Natl Acad Sci U S A* **102:** 81-86.

Balbo, A., and Schuck, P. 2005. Analytical ultracentrifugation in the study of protein self-association and heterogeneous protein-protein interactions. In *Protein-Protein Interactions*. (eds. E. Golemis, and P.D. Adams), pp. 253-277. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Berkowitz, S.A. 2006. Role of analytical ultracentrifugation in assessing the aggregation of protein biopharmaceuticals. *Aaps J* **8:** E590-605.

Binger, K.J., Pham, C.L.L., Wilson, L.M., Bailey, M.F., Lawrence, L.J., Schuck, P., and Howlett, G.J. submitted. Apolipoprotein C-II amyloid fibrils assemble via a reversible pathway that includes fibril breaking and re-joining.

Boukari, H., Nossal, R., Sackett, D.L., and Schuck, P. 2004. Hydrodynamics of nanoscopic tubulin rings in dilute solution. *Physical Review Letters* **93:** 098106.

Breuer, S., Gerlach, H., Kolaric, B., Urbanke, C., Opitz, N., and Geyer, M. 2006. Biochemical indication for myristoylation-dependent conformational changes in HIV-1 Nef. *Biochemistry* **45:** 2339-2349.

Brown, P., Balbo, A., and Schuck, P. 2007. Using prior knowledge in the determination of macromolecular size-distributions by analytical ultracentrifugation. *Biomacromolecules* **8:** 2011-2024.

Brown, P., and Schuck, P. in press. A new adaptive grid-size algorithm for the simulation of sedimentation velocity profiles in analytical ultracentrifugation. *Comp. Phys. Comm.*

Brown, P.H., and Schuck, P. 2006. Macromolecular Size-And-Shape Distributions by Sedimentation Velocity Analytical Ultracentrifugation. *Biophys. J.* **90:** 4651-4661.

Buisson, M., Valette, E., Hernandez, J.F., Baudin, F., Ebel, C., Morand, P., Seigneurin, J.M., Arlaud, G.J., and Ruigrok, R.W. 2001. Functional determinants of the Epstein-Barr virus protease. *J Mol Biol* **311:** 217-228.

Burgess, B.R., Schuck, P., and Garboczi, D.N. 2005. Dissection of merozoite surface protein 3, a representative of a family of Plasmodium falciparum surface proteins, reveals an oligomeric and highly elongated molecule. *J Biol Chem* **280:** 37236-37245.

Calarese, D.A., Scanlan, C.N., Zwick, M.B., Deechongkit, S., Mimura, Y., Kunert, R., Zhu, P., Wormald, M.R., Stanfield, R.L., Roux, K.H., et al. 2003. Antibody domain exchange is an immunological solution to carbohydrate cluster recognition. *Science* **300:** 2065-2071.

Cann, J.R. 1982. Theory of sedimentation for antigen-antibody reactions: effect of antibody heterogeneity on the shape of the pattern. *Molecular Immunity* **19:** 505-514.

Chan, A.C., Lelj-Garolla, B., F, I.R., Pedersen, K.A., Mauk, A.G., and Murphy, M.E. 2006. Cofacial heme binding is linked to dimerization by a bacterial heme transport protein. *J Mol Biol* **362:** 1108-1119.

Chou, C.Y., Jen, W.P., Hsieh, Y.H., Shiao, M.S., and Chang, G.G. 2006. Structural and functional variations in human apolipoprotein E3 and E4. *J Biol Chem* **281:** 13333-13344.

Connaghan-Jones, K.D., Heneghan, A.F., Miura, M.T., and Bain, D.L. 2006. Hydrodynamic analysis of the human progesterone receptor A-isoform reveals that self-association occurs in the micromolar range. *Biochemistry* **45:** 12090-12099.

Cox, D.J. 1966. Sedimentation of an initially skewed boundary. *Science* **152:** 359-361.

Creeth, J.M., and Knight, C.G. 1965. On the estimation of the shape of macromolecules from sedimentation and viscosity measurements. *Biochim Biophys Acta* **102:** 549-558.

Dam, J., Baber, J., Grishaev, A., Malchiodi, E.L., Schuck, P., Bax, A., and Mariuzza, R.A. 2006. Variable dimerization of the Ly49A natural killer cell receptor results in differential engagement of its MHC class I ligand. *J Mol Biol* **362:** 102-113.

Dam, J., Guan, R., Natarajan, K., Dimasi, N., Chlewicki, L.K., Kranz, D.M., Schuck, P., Margulies, D.H., and Mariuzza, R.A. 2003. Variable MHC class I engagement by Ly49 NK

cell receptors revealed by the crystal structure of Ly49C bound to H-2Kb. *Nature Immunology* **4:** 1213-1222.

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

Dam, J., and Schuck, P. 2005. Sedimentation velocity analysis of protein-protein interactions: Sedimentation coefficient distributions c(s) and asymptotic boundary profiles from Gilbert-Jenkins theory. *Biophys J* **89:** 651-666.

Dam, J., Velikovsky, C.A., Mariuzza, R., Urbanke, C., and Schuck, P. 2005. Sedimentation velocity analysis of protein-protein interactions: Lamm equation modeling and sedimentation coefficient distributions c(s). *Biophys J* **89:** 619-634.

Davis, A.J., Perugini, M.A., Smith, B.J., Stewart, J.D., Ilg, T., Hodder, A.N., and Handman, E. 2004. Properties of GDP-mannose pyrophosphorylase, a critical enzyme and drug target in Leishmania mexicana. *J Biol Chem* **279:** 12462-12468.

Deka, R.K., Brautigam, C.A., Tomson, F.L., Lumpkins, S.B., Tomchick, D.R., Machius, M., and Norgard, M.V. 2007. Crystal structure of the Tp34 (TP0971) lipoprotein of treponema palladium: Implications of its metal-bound state and affinity for human lactoferrin. *J Biol Chem* **282:** 5844-5958.

Deng, L., Langley, R.J., Brown, P.H., Xu, G., Teng, L., Wang, Q., Gonzales, M.I., Callender, G.G., Nishimura, M.I., Topalian, S.L., et al. 2007. Structural basis for the recognition of mutant self by a tumor-specific, MHC class II-restricted T cell receptor. *Nat Immunol* **8:** 398-408.

Dishon, M., Weiss, G.H., and Yphantis, D.A. 1967. Numerical simulations of the Lamm equation: III. Velocity centrifugation. *Biopolymers* **5:** 697-713.

Doun, S.S., Burgner, J.W., 2nd, Briggs, S.D., and Rodwell, V.W. 2005. Enterococcus faecalis phosphomevalonate kinase. *Protein Sci* **14:** 1134-1139.

Egan, C.A., Houston, K.M., Alcocer, M.J., Solovyova, A., Tate, R., Lochnit, G., McInnes, I.B., Harnett, M.M., Geyer, R., Byron, O., et al. 2006. Lack of immunological cross-reactivity between parasite-derived and recombinant forms of ES-62, a secreted protein of Acanthocheilonema viteae. *Parasitology* **132:** 263-274.

Elzen, B. 1988. *Scientists and rotors. The development of biochemical ultracentrifuges*. Dissertation, University Twente, Enschede.

Errington, N., and Rowe, A.J. 2003. Probing conformation and conformational change in proteins is optimally undertaken in relative mode. *Eur Biophys J* **32:** 511-517.

Frigon, R.P., and Timasheff, S.N. 1975. Magnesium-induced self-association of calf brain tubulin. I. Stoichiometry. *Biochemistry* **14:** 4559-4566.

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

Furtado, P.B., Whitty, P.W., Robertson, A., Eaton, J.T., Almogren, A., Kerr, M.A., Woof, J.M., and Perkins, S.J. 2004. Solution structure determination of monomeric human IgA2 by X-ray and neutron scattering, analytical ultracentrifugation and constrained modelling: a comparison with monomeric human IgA1. *J Mol Biol* **338:** 921-941.

Gabrielson, J.P., Brader, M.L., Pekar, A.H., Mathis, K.B., Winter, G., Carpenter, J.F., and Randolph, T.W. 2007. 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* **96:** 268-279.

Garcia De La Torre, J., Huertas, M.L., and Carrasco, B. 2000. Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. *Biophys J* **78:** 719-730.

Gilbert, G.A., and Jenkins, R.C. 1956. Boundary problems in the sedimentation and electrophoresis of complex systems in rapid reversible equilibrium. *Nature* **177:** 853-854.

Gilbert, G.A., and Jenkins, R.C. 1959. Sedimentation and electrophoresis of interacting substances. II. Asymptotic boundary shape for two substances interacting reversibly. *Proc Royal Soc A* **253:** 420-437.

Gilbert, L.M., and Gilbert, G.A. 1978. Molecular transport of reversibly reacting systems: Asymptotic boundary profiles in sedimentation, electrophoresis, and chromatography. *Methods Enzymol* **48:** 195-211.

Greive, S.J., Lins, A.F., and von Hippel, P.H. 2005. Assembly of an RNA-protein complex. Binding of NusB and NusE (S10) proteins to boxA RNA nucleates the formation of the antitermination complex involved in controlling rRNA transcription in Escherichia coli. *J Biol Chem* **280:** 36397-36408.

Gohon, Y., Pavlov, G., Timmins, P., Tribet, C., Popot, J.L., and Ebel, C. 2004. Partial specific volume and solvent interactions of amphiphol A8-35. *Anal Biochem* **334:** 318-334.

Guan, R., Malchiodi, E.L., Wang, Q., Schuck, P., and Mariuzza, R.A. 2004. Crystal structure of the C-terminal peptidoglycan-binding domain of human peptidoglycan recognition protein Ialpha. *J Biol Chem* **279:** 31873-31882.

Gupta, N., Arthos, J., Khazanie, P., Steenbeke, T.D., Censoplano, N.M., Chung, E.A., Cruz, C.C., Chaikin, M.A., Daucher, M., Kottilil, S., et al. 2005. Targeted lysis of HIV-infected cells by natural killer cells armed and triggered by a recombinant immunoglobulin fusion protein: implications for immunotherapy. *Virology* **332:** 491-497.

Harding, S.E., Longman, E., Carrasco, B., Ortega, A., and Garcia De La Torre, J. 2003. Studying antibody conformations by ultracentrifugation and hydrodynamic modeling. *Methods in Molecular Biology* **248:** 93-113.

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

Houtman, J.C., Higashimoto, Y., Dimasi, N., Cho, S., Yamaguchi, H., Bowden, B., Regan, C., Malchiodi, E.L., Mariuzza, R., Schuck, P., et al. 2004. Binding specificity of multiprotein signaling complexes is determined by both cooperative interactions and affinity preferences. *Biochemistry* **43:** 4170-4178.

Houtman, J.C., Yamaguchi, H., Barda-Saad, M., Braiman, A., Bowden, B., Appella, E., Schuck, P., and Samelson, L.E. 2006. Oligomerization of signaling complexes by the multipoint binding of GRB2 to both LAT and SOS1. *Nat Struct Mol Biol* **13:** 798-805.

Howlett, G.J., Minton, A.P., and Rivas, G. 2006. Analytical ultracentrifugation for the study of protein association and assembly. *Curr Opin Chem Biol* **10:** 430-436.

Hsu, W.C., Chang, H.C., Chou, C.Y., Tsai, P.J., Lin, P.I., and Chang, G.G. 2005. Critical assessment of important regions in the subunit association and catalytic action of the severe acute respiratory syndrome coronavirus main protease. *J Biol Chem* **280:** 22741-22748.

Johnston, J.P., and Ogston, A.G. 1946. A boundary anomaly found in the ultracentrifugal sedimentation of mixtures. *Trans. Faraday Soc.* **42:** 789-799.

Jomaa, A., Damjanovic, D., Leong, V., Ghirlando, R., Iwanczyk, J., and Ortega, J. 2007. The inner cavity of Escherichia coli DegP protein is not essential for molecular chaperone and proteolytic activity. *J Bacteriol* **189:** 706-716.

Kornblatt, J.A., and Schuck, P. 2005. Influence of temperature on the conformation of canine plasminogen: an analytical ultracentrifugation and dynamic light scattering study. *Biochemistry* **44:** 13122-13131.

Krauss, G., Pingoud, A., Boehme, D., Riesner, D., Peters, F., and Maass, G. 1975. Equivalent and non-equivalent binding sites for tRNA on aminoacyl-tRNA synthetases. *Eur. J. Biochem.* **55:** 517-529.

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

Lebowitz, J., Lewis, M.S., and Schuck, P. 2002. Modern analytical ultracentrifugation in protein science: a tutorial review. *Protein Sci* **11:** 2067-2079.

Lelj-Garolla, B., and Mauk, A.G. 2005. Self-association of a small heat shock protein. *J Mol Biol* **345:** 631-642.

Lelj-Garolla, B., and Mauk, A.G. 2006. Self-association and chaperone activity of Hsp27 are thermally activated. *J Biol Chem* **281:** 8169-8174.

Lewis, R.J., Scott, D.J., Brannigan, J.A., Ladds, J.C., Cervin, M.A., Spiegelman, G.B., Hoggett, J.G., Barak, I., and Wilkinson, A.J. 2002. Dimer formation and transcription activation in the sporulation response regulator Spo0A. *J Mol Biol* **316:** 235-245.

Li, H., Van Vranken, S., Zhao, Y., Li, Z., Guo, Y., Eisele, L., and Li, Y. 2005. Crystal structures of T cell receptor (beta) chains related to rheumatoid arthritis. *Protein Sci* **14:** 3025-3038.

Li, H., Zhao, Y., Guo, Y., Li, Z., Eisele, L., and Mourad, W. 2007. Zinc induces dimerization of the class II major histocompatibility complex molecule that leads to cooperative binding to a superantigen. *J Biol Chem* **282:** 5991-6000.

Liu, J., Andya, J.D., and Shire, S.J. 2006a. A critical review of analytical ultracentrifugation and field flow fractionation methods for measuring protein aggregation. *Aaps J* **8:** E580-589.

Liu, Y., Cheney, M.D., Gaudet, J.J., Chruszcz, M., Lukasik, S.M., Sugiyama, D., Lary, J., Cole, J., Dauter, Z., Minor, W., et al. 2006b. The tetramer structure of the Nervy homology two domain, NHR2, is critical for AML1/ETO's activity. *Cancer Cell* **9:** 249-260.

Marcum, J.M., and Borisy, G.G. 1978. Sedimentation velocity analyses of the effect of hydrostatic pressure on the 30 S microtubule protein oligomer. *J Biol Chem* **253:** 2852-2857.

Minor, K.H., Schar, C.R., Blouse, G.E., Shore, J.D., Lawrence, D.A., Schuck, P., and Peterson, C.B. 2005. A mechanism for assembly of complexes of vitronectin and plasminogen activator inhibitor-1 from sedimentation velocity analysis. *J. Biol. Chem.* **31:** 28711-28720.

Minton, A.P. 2001. The influence of macromolecular crowding and macromolecular confinement on biochemical reactions in physiological media. *J Biol Chem* **276:** 10577-10580.

Nollmann, M., Stark, W.M., and Byron, O. 2004. Low-resolution reconstruction of a synthetic DNA holliday junction. *Biophys J* **86:** 3060-3069.

Patel, T.R., Harding, S.E., Ebringerova, A., Deszczynski, M., Hromadkova, Z., Togola, A., Paulsen, B.S., Morris, G.A., and Rowe, A.J. 2007. Weak self-association in a carbohydrate system. *Biophys J* **93:** 741-749.

Pekar, A., and Sukumar, M. 2007. Quantitation of aggregates in therapeutic proteins using sedimentation velocity analytical ultracentrifugation: practical considerations that affect precision and accuracy. *Anal Biochem* **367:** 225-237.

Perugini, M.A., Griffin, M.D., Smith, B.J., Webb, L.E., Davis, A.J., Handman, E., and Gerrard, J.A. 2005. Insight into the self-association of key enzymes from pathogenic species. *Eur Biophys J* **34:** 469-476.

Provencher, S.W. 1979. Inverse problems in polymer characterization: direct analysis of polydispersity with photon correlation spectroscopy. *Makromol. Chem.* **180:** 201-209.

Rai, N., Nollmann, M., Spotorno, B., Tassara, G., Byron, O., and Rocco, M. 2005. SOMO (SOlution MOdeler) differences between X-Ray- and NMR-derived bead models suggest a role for side chain flexibility in protein hydrodynamics. *Structure (Camb)* **13:** 723-734.

Rivas, G., Fernandez, J.A., and Minton, A.P. 1999. Direct observation of the self-association of dilute proteins in the presence of inert macromolecules at high concentration via tracer sedimentation equilibrium: theory, experiment, and biological significance. *Biochemistry* **38:** 9379-9388.

Rowe, A.J. 1992. The Concentration Dependence of Sedimentation. In *Analytical Ultracentrifugation in Biochemistry and Polymer Science*. (eds. S.E. Harding, A.J. Rowe, and J.C. Horton), pp. 394-406. Royal Society of Chemistry, Cambridge.

Schachman, H.K. 1959. *Ultracentrifugation in Biochemistry*. Academic Press, New York.

Schachman, H.K. 1989. Analytical ultracentrifugation reborn. *Nature* **341:** 259-260.

Schachman, H.K. 1992. Is There a Future for the Ultracentrifuge? In *Analytical Ultracentrifugation in Biochemistry and Polymer Science*. (eds. S.E. Harding, A.J. Rowe, and J.C. Horton), pp. 3-15. Royal Society of Chemistry, Cambridge.

Schmeisser, H., Gorshkova, I., Brown, P., Kontsek, P., Schuck, P., and Zoon, K. submitted. Two interferons alpha influence each other during their interaction with the extracellular domain of human type I interferon receptor subunit 2.

Schuck, P. 1997. Use of Surface Plasmon Resonance to Probe the Equilibrium and Dynamic Aspects of Interactions Between Biological Macromolecules. *Ann. Rev. Biophys.Biomol. Struct.* **26:** 541-566.

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

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

Schuck, P. 2004a. A model for sedimentation in inhomogeneous media. I. Dynamic density gradients from sedimenting co-solutes. *Biophys. Chem.* **108:** 187-200.

Schuck, P. 2004b. A model for sedimentation in inhomogeneous media. II. Compressibility of aqueous and organic solvens. *Biophys. Chem.* **187:** 201-214.

Schuck, P. 2006. Diffusion-deconvoluted sedimentation coefficient distributions for the analysis of interacting and non-interacting protein mixtures. In *Modern Analytical Ultracentrifugation: Techniques and Methods*. (eds. D.J. Scott, S.E. Harding, and A.J. Rowe), pp. 26-50. The Royal Society of Chemistry, Cambridge.

Schuck, P. 2007a. *http://www.analyticalultracentrifugation.com/references.htm*.

Schuck, P. 2007b. Sedimentation velocity in the study of reversible multiprotein complexes. In *Biophysical Approaches for the study of complex reversible systems.* (ed. P. Schuck), pp. 469-518. Springer, New York.

Schuck, P. 2007c. *www.analyticalultracentrifugation.com*.

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

Schuck, P., MacPhee, C.E., and Howlett, G.J. 1998. Determination of sedimentation coefficients for small peptides. *Biophys. J.* **74:** 466-474.

Schuck, P., Perugini, M.A., Gonzales, N.R., Howlett, G.J., and Schubert, D. 2002. Size-distribution analysis of proteins by analytical ultracentrifugation: strategies and application to model systems. *Biophys J* **82:** 1096-1111.

Schuck, P., Taraporewala, Z., McPhie, P., and Patton, J.T. 2000. Rotavirus nonstructural protein NSP2 self-assembles into octamers that undergo ligand-induced conformational changes. *J Biol Chem* **276:** 9679-9687.

Scott, D.J., and Schuck, P. 2006. A brief introduction to the analytical ultracentrifugation of proteins for beginners. In *Modern Analytical Ultracentrifugation: Techniques and Methods*.

(eds. D.J. Scott, S.E. Harding, and A.J. Rowe), pp. p.1-25. The Royal Society of Chemistry, Cambridge.

Sivia, D.S. 1996. *Data Analysis. A Bayesian Tutorial.* Oxford University Press, Oxford.

Solovyova, A., Schuck, P., Costenaro, L., and Ebel, C. 2001. Non-ideality by sedimentation velocity of halophilic malate dehydrogenase in complex solvents. *Biophysical Journal* **81:** 1868-1880.

Solovyova, A.S., Nollmann, M., Mitchell, T.J., and Byron, O. 2004. The solution structure and oligomerization behavior of two bacterial toxins: pneumolysin and perfringolysin O. *Biophys J* **87:** 540-552.

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

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

Svedberg, T., and Pedersen, K.O. 1940. *The ultracentrifuge*. Oxford University Press, London.

Svergun, D.I., and Vachette, P. 2007. Structure analysis of macromolecular complexes. In *Protein Interactions. Biophysical Approaches for the Study of Complex Reversible Systems.* (ed. P. Schuck), pp. 317-366. Springer, New York.

Urbanke, C., Witte, G., and Curth, U. 2005. A sedimentation velocity method in the analytical ultracentrifuge for the study of protein-protein interactions. In *Protein-Ligand Interactions: Methods and Applications*. (ed. G.U. Nienhaus), pp. 101-113. Humana Press, Totowa.

Urbanke, C., Ziegler, B., and Stieglitz, K. 1980. Complete evaluation of sedimentation velocity experiments in the analytical ultracentrifuge. *Fresenius Z. Anal. Chem.* **301:** 139-140

Werner, W.E., and Schachman, H.K. 1989. Analysis of the ligand-promoted global conformational change in aspartate transcarbamoylase. Evidence for a two-state transition from boundary spreading in sedimentation velocity experiments. *J Mol Biol* **206:** 221-230.

West, A.P., Jr., Herr, A.B., and Bjorkman, P.J. 2004. The chicken yolk sac IgY receptor, a functional equivalent of the mammalian MHC-related Fc receptor, is a phospholipase A2 receptor homolog. *Immunity* **20:** 601-610.

Yang, M., Horii, K., Herr, A.B., and Kirley, T.L. 2006. Calcium-dependent dimerization of human soluble calcium activated nucleotidase: characterization of the dimer interface. *J Biol Chem* **281:** 28307-28317.

Yikilmaz, E., Rouault, T.A., and Schuck, P. 2005. Self-association and ligand-induced conformational changes of iron regulatory proteins 1 and 2. *Biochemistry* **44:** 8470-8478.